

小麦‘Fielder’ ω -醇溶蛋白基因及启动子的克隆与序列分析

郭丹丹 李保云*

(中国农业大学 农学院/北京市作物遗传改良重点实验室/教育部作物杂种优势研究与利用重点实验室,北京 100193)

摘要 为研究六倍体小麦中 ω -醇溶蛋白基因簇的启动子差异,以普通小麦品种‘Fielder’为试验材料,利用同源克隆的方法,克隆 ω -醇溶蛋白基因的编码区和启动子序列,并进行生物信息学相关分析。结果表明:1)共克隆得到11条不同的基因序列(MN441496~MN441506),其中6条(MN441496~MN441497和MN441503~MN441506)编码ARE/Q型 ω -醇溶蛋白,5条(MN441498~MN441502)编码SRL型 ω -醇溶蛋白。2)ARE/Q型 ω -醇溶蛋白基因编码区长度范围在972~1 158 bp,SRL型 ω -醇溶蛋白基因编码区长度范围在1 303~1 419 bp,这种差异主要由中间重复区的Indel类型和数量不同造成。3)ARE/Q型 ω -醇溶蛋白基因启动子在-300 bp含有1个典型的endosperm box,而SRL型 ω -醇溶蛋白基因启动子在-300和-600 bp处各含有1个endosperm box。这些 ω -醇溶蛋白基因启动子上的差异,可能引起表达水平的差异。

关键词 小麦; A-PAGE; 醇溶蛋白; 启动子; endosperm box

中图分类号 Q946.2

文章编号 1007-4333(2020)07-0001-09

文献标志码 A

Cloning and sequence analysis of the coding region and promoter of ω -gliadin genes from wheat ‘Fielder’ (*Triticum aestivum* L.)

GUO Dandan, LI Baoyun*

(College of Agronomy and Biotechnology/Beijing Key Laboratory of Crop Genetic Improvement/Key Laboratory of Crop Heterosis and Utilization of Ministry of Education, China Agricultural University, Beijing 100193, China)

Abstract To understand the promoter differences of ω -gliadin gene cluster in hexaploid wheat, the coding region along with the promoter sequences of ω -gliadin in hexaploid common wheat cultivar ‘Fielder’ were obtained by homologous cloning. Bioinformatics analysis was conducted. The results showed that: 1) A total of 11 different ω -gliadin gene sequences were acquired, and six of which belonged to the ARE/Q type genes, the rest five were SRL type genes. 2) The coding region lengths were varied from 972 to 1 158 bp among the six ARE/Q type genes and from 1 303 to 1 419 bp among the five SRL type genes, which were the results of differences in types and quantities of Indels in the center repeat region. 3) There located a conserved typical endosperm box at -300 bp upstream of ARE/Q type ω -gliadin genes. Nevertheless, the endosperm box appeared at -300 and -600 bp upstream of SRL type ω -gliadin genes, respectively. The differences in the promoter of ω -gliadin gene cluster might influence the expression levels of different ω -gliadin genes.

Keywords wheat; A-PAGE; gliadin; promoter; endosperm box

小麦是世界上重要的三大粮食作物之一。小麦具有面筋蛋白,所以能够被加工成面条、面包、馒头和饼干等各种面食品。面筋蛋白主要由单体形式的醇溶蛋白和多聚体形式的谷蛋白组成。根据在酸性

聚丙烯酰胺凝胶电泳(Acid polyacrylamide-gel electrophoresis, A-PAGE)中迁移率的不同,醇溶蛋白可分为 α -、 β -、 γ -和 ω -醇溶蛋白4种类型,其中 α -、 β -、 γ -醇溶蛋白中脯氨酸、苯丙氨酸和谷氨酰胺的含

收稿日期:2019-09-16

基金项目:转基因生物新品种培育重大专项(2016ZX08002003);国家自然科学基金(31371607)

第一作者:郭丹丹,博士研究生,E-mail:gdd1990@cau.edu.cn

通讯作者:李保云,教授,主要从事小麦品质遗传改良研究,E-mail:baoyunli@cau.edu.cn

量较少,但是含硫氨基酸较多,因此被称为富硫醇溶蛋白,其中的半胱氨酸是维持单肽及肽链间二硫键的重要因素;而 ω -醇溶蛋白通常不含半胱氨酸,无法参与链内或链间二硫键的形成,多数只含1个甲硫氨酸残基,因此被称为贫硫醇溶蛋白,有少数 ω -醇溶蛋白含有奇数个半胱氨酸残基,能够参与谷蛋白聚合体的形成,导致面粉的品质变劣^[1-3]。醇溶蛋白含有能够诱发麸质不耐受人群乳糜泻(Celiac disease, CD)疾病的肽段,主要类型是 α -、 β -和 γ -醇溶蛋白,在 ω -醇溶蛋白中也含有部分肽段^[4]。通过RNAi介导使醇溶蛋白含量降低或沉默,能够在不影响总蛋白和淀粉含量的前提下达到降低CD毒性的目的^[5]。因此克隆 ω -醇溶蛋白基因对小麦品质的遗传改良具有重要意义。

ω -醇溶蛋白基因的基本结构,包括信号肽、N末端非重复区、占肽段90%~96%的中间重复区及C末端非重复区4部分^[6]。六倍体小麦可能含有15~18个 ω -醇溶蛋白基因^[7],编码基因位于第1部分同源染色体短臂上^[8],由1A、1D染色体基因编码的蛋白在A-PAGE上的迁移率较慢,称为 ω 1,2-醇溶蛋白(46~58 ku),推导的氨基酸序列N末端前3个通常为ARE/Q或KEL;由1B染色体基因编码的蛋白在A-PAGE上的迁移率较快,称为 ω 5-醇溶蛋白(55~65 ku),N末端前3个氨基酸为SRL^[9-10]。 ω 1,2-醇溶蛋白在中间重复区的重复单元类型与 ω 5-醇溶蛋白有明显区别,这种差异也表现在氨基酸组成上, ω 1,2-醇溶蛋白的谷氨酰胺与脯氨酸残基数比为4:3,而 ω 5-醇溶蛋白的比值为5:2^[11]。

ω -醇溶蛋白基因最早被克隆,是由于与 γ -醇溶蛋白具有交叉杂交的性质,才会在对 γ -醇溶蛋白基因进行克隆时被鉴定到^[12]。利用电泳或RP-UPLC等方法对醇溶蛋白分离后回收,再进行N末端测序或质谱鉴定也鉴定到一些醇溶蛋白^[13-14]。通过免疫化学方法制备特定醇溶蛋白的单克隆抗体,对 ω -醇溶蛋白的分离和鉴定也发挥了重要作用^[14]。通过基因组水平、转录水平和蛋白水平相结合的方法在二倍体祖先种乌拉尔图小麦中鉴定到2个 ω -醇溶蛋白基因及其蛋白^[4]。同样,利用三代全长转录组测序技术结合质谱技术,也鉴定到六倍体普通小麦品种‘小偃81’中有5个能够表达的 ω -醇溶蛋白^[15];通过将‘中国春’基因组进行De Novo BioNano组装,结合RNA-seq数据同样获得19条

ω -醇溶蛋白基因的序列信息^[16]。目前对小麦高分子量谷蛋白亚基(High molecular weight glutenin subunit, HMW-GS)和低分子量谷蛋白亚基(Low molecular weight glutenin subunit, LMW-GS)基因的调控研究已经较为深入^[17-20]。但是, ω -醇溶蛋白基因家族庞大,相似性较高,假基因比例高,基因中间重复区含有较多的重复单元,GC含量较高,这些因素导致相关研究更加困难。已有研究多数集中在区分不同 ω -醇溶蛋白及其基因的差异,对不同 ω -醇溶蛋白基因启动子的差异比较研究报道较少,本研究以春小麦‘Fielder’为材料,通过PCR和克隆测序法分离 ω -醇溶蛋白基因,分析编码区和启动子序列差异,旨在探明小麦中不同 ω -醇溶蛋白基因的启动子序列差异,以期对 ω -醇溶蛋白基因的转录调控机理研究奠定基础。

1 材料与方 法

1.1 试验材料

本试验所用六倍体普通小麦(*Triticum aestivum* L.)品种‘Fielder’和‘中国春’(CS)均由本实验室保存。

1.2 试验方法

1.2.1 蛋白质提取

取1粒小麦种子充分碾碎,置于1.5 mL离心管中,加入800 μ L 75%乙醇,室温震荡过夜,提取醇溶蛋白。

1.2.2 A-PAGE的制备及电泳

采用 $\text{FeSO}_4\text{-Vc-H}_2\text{O}_2$ 催化系统制备A-PAGE^[21],500 V稳压电泳3.5 h,对醇溶蛋白进行分离。

1.2.3 DNA提取

将小麦种子播种在培养皿上,暗培养1周左右,取4叶1心期的叶片,用CTAB法提取基因组DNA。

1.2.4 引物设计及扩增

根据已经公布的‘CS’ ω -醇溶蛋白基因上下游序列^[16],分别设计能够扩增不同基因组上的 ω -醇溶蛋白基因及启动子的引物,详见表1。

其中上游引物位于 ω -醇溶蛋白基因起始密码子上游1 kb左右;下游引物位于终止密码子下游100 bp左右。

采用大连宝生物(Takara)公司的Tks GflexTM DNA Polymerase高保真酶,按说明书进行PCR扩增。

表 1 ω -醇溶蛋白基因编码区和启动子
克隆引物及测序引物

Table 1 The ω -gliadin genes cloning primers and sequencing primers

引物 Primer	序列(5'-3') Sequence (5'-3')
$\omega A-F$	CCTACTGGTATGAAGGACAAC
$\omega A-R$	GGTTACACATTATGGTGCACA
$\omega B-F$	TTGCTCTATCTTTCCTCTTT
$\omega B-R$	CACATCACTTTATTTATCATCGT
$\omega D-F$	AAGACAAAAGTTGGAATGCT
$\omega D-R$	TGAACGCTCATTGACTAAACA
$\omega 1$	CATCTTTGTCCTCCTTGCCAT
$\omega 2$	AAGATTACAAACTTAGTTCC
$\omega 3$	CAGTGCCAGTAGGCTGCTAA
$\omega 4$	TTAGCAGCCTACTGGCACTG
$\omega 5$	TTCCACTCAAGTTATGCCAA
$\omega 6$	GAGGACAAAGATGAGGAAGG

注: ωA 、 ωB 和 ωD 分别为克隆六倍体小麦 A、B 和 D 基因组上 ω -醇溶蛋白基因启动子和编码区的引物, $\omega 1$ 和 $\omega 2$ 是根据 A 基因组序列设计的测序引物, $\omega 2$ 、 $\omega 3$ 和 $\omega 4$ 是根据 B 基因组序列设计的测序引物, $\omega 5$ 和 $\omega 6$ 是根据 D 基因组序列设计的测序引物。

Note: ωA , ωB and ωD are primers used to clone the promoter and coding sequence of ω -gliadin genes on the A, B and D subgenome of hexaploid common wheat, respectively. $\omega 1$ and $\omega 2$ are primers for sequencing designed according to the sequences of A subgenome. $\omega 2$, $\omega 3$ and $\omega 4$ are primers for sequencing designed according to the sequences of B subgenome. $\omega 5$ and $\omega 6$ are primers for sequencing designed according to the sequences of D subgenome.

1.2.5 目的片段回收与连接转化

将扩增产物在 1% 琼脂糖凝胶上进行分离, 用琼脂糖凝胶回收试剂盒(北京天根生化科技有限公司)回收目的条带, 与 pEasy-Blunt Zero 载体(北京全式金生物公司)连接, 将连接产物转化大肠杆菌 Trans-T1 感受态细胞(北京全式金生物公司), 涂布于含卡那霉素(Kan)的 LB 培养基上, 37 °C 倒置培养过夜。

1.2.6 重组克隆筛选

挑取一定数量的单克隆, 用 M13 引物进行菌落 PCR 检测, 将阳性克隆送北京擎科生物科技有限公

司测序。首先用载体上的 M13F 和 M13R 引物进行测序, 再根据 ω -醇溶蛋白基因非重复区差异设计引物, 对不同基因组的 ω -醇溶蛋白基因进行测序, 引物信息详见表 1。

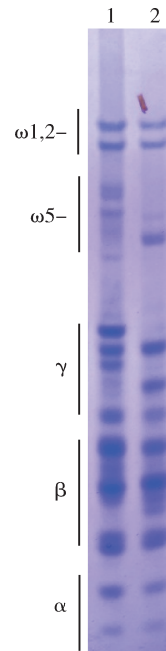
1.2.7 序列分析

采用 NCBI (<https://www.ncbi.nlm.nih.gov/>) 和 DNAMAN 软件对测序结果进行比对、拼接及翻译。利用 MegaX 软件 (<https://www.megasoftware.net/>) 进行系统发生树的构建, 启动子序列提交 PlantCARE 网站 (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) 进行 motif 分析。

2 结果与分析

2.1 小麦种子醇溶蛋白的 A-PAGE

提取 'CS' 和 'Fielder' 种子的醇溶蛋白进行 A-PAGE。从图 1 看出, 2 个材料的 ω -醇溶蛋白谱带差异较大。'CS' 在 $\omega 1, 2$ -醇溶蛋白区有 2 条带; 在 $\omega 5$ -醇溶蛋白区有 3 条带, 其中有 2 条带非常接近, 共 5 条 ω -醇溶蛋白条带。'Fielder' 在 $\omega 1, 2$ -醇



α -、 β -、 γ - 和 ω - 分别表示按迁移率不同划分的 4 组醇溶蛋白类型; $\omega 1, 2$ - 和 $\omega 5$ - 分别为 ω -醇溶蛋白的 2 种类型。1 和 2 分别为 'Fielder' 和 'CS'。

α -, β -, γ - and ω - are the four types of wheat gliadin; $\omega 1, 2$ - and $\omega 5$ - are the two types of ω -gliadin. Lane 1 and 2 are 'Fielder' and 'CS'.

图 1 小麦种子醇溶蛋白的 A-PAGE

Fig. 1 The A-PAGE of wheat grain gliadin

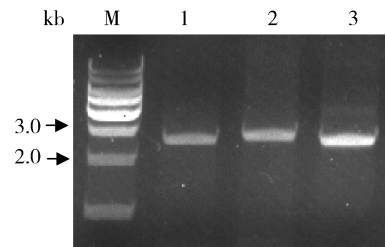
溶蛋白区有2条带,并且这2条带的迁移率与‘CS’非常接近;在 ω 5-醇溶蛋白区有5条带,其中有2条带非常接近,共7条 ω -醇溶蛋白条带。

2.2 小麦种子 ω -醇溶蛋白基因及启动子克隆

利用3组 ω -醇溶蛋白基因特异引物(表1),以‘Fielder’的基因组DNA为模板进行PCR扩增,得到约2 kb左右的目的条带(图2)。将目的条带回收后,连接T载体并转化至大肠杆菌中,挑选阳性单克隆进行序列测定,共获得11种序列。将这些序列在NCBI网站进行Blast,相似性较高的是 ω -醇溶蛋白基因,进一步分析发现这些序列含有 ω -醇溶蛋白基因的基本结构,即均为 ω -醇溶蛋白基因及其启动子序列(GenBank登录号:MN441496~MN441506)。

这些基因序列可以分为2组,ARE/Q和SRL型 ω -醇溶蛋白基因。编码区长度、推导的氨基酸数目以及重复单元种类和数量各不相同(表2)。

ARE/Q型 ω -醇溶蛋白基因编码区长度范围在972~1 158 bp,推导的氨基酸主要含有PQQPFP和PFPQQPQQ这2种类型的重复单元;SRL型 ω -醇溶蛋白基因编码区的长度变异范围为1 303~1 419 bp,主要重复单元为FPQQQ和PQQQFP。



M, 1 kb ladder, 1, 2 和 3 分别为 ω A、 ω B和 ω D引物的扩增产物。
M is 1 kb ladder. Lane 1, 2 and 3 show the amplifications of primer pairs of ω A, ω B and ω D, respectively.

图2 ω -醇溶蛋白基因扩增产物

Fig. 2 The amplification of ω -gliadin genes

表2 ω -醇溶蛋白基因及推导的氨基酸序列比较

Table 2 The comparison of gene and deduced amino acid sequences of ω -gliadin

GenBank 登录号 GenBank accession number	引物 Primer	启动子 长度/bp Length of promoter	编码区 长度/bp Length of coding sequence	类型 Type	氨基酸 数目 Number of amino acid	重复单元 类型 Repetitive unit
MN441496	ω A	1 024	1 140	ARE	379	PQQPFP(18)PFPQQPQQ(9)
MN441497	ω A	1 024	1 158	ARE	385	PQQPFP(18)PFPQQPQQ(9)
MN441498	ω B	1 130	1 320	SRL	419	FPQQQ(29)PQQQFP(10)
MN441499	ω B	1 130	1 303	SRL	376	FPQQQ(1)PQQQFP(0)
MN441500	ω B	1 130	1 419	SRL	472	FPQQQ(31)PQQQFP(13)
MN441501	ω B	1 130	1 320	SRL	91	FPQQQ(27)PQQQFP(9)
MN441502	ω B	1 130	1 335	SRL	444	FPQQQ(28)PQQQFP(11)
MN441503	ω D	1 133	972	ARQ	323	PQQPFP(12)PFPQQPQQ(7)
MN441504	ω D	1 133	1 077	ARE	359	PQQPFP(17)PFPQQPQQ(8)
MN441505	ω D	1 135	1 158	ARQ	385	PQQPFP(18)PFPQQPQQ(9)
MN441506	ω D	1 135	1 134	ARQ	377	PQQPFP(15)PFPQQPQQ(6)

注:括号内数字表示重复单元的数目。

Note: The number in the brackets indicates the amount of repetitive unit.

2.2.1 ω -醇溶蛋白基因编码区分析

A和D基因组上的 ω -醇溶蛋白基因相似性较高,从‘Fielder’中利用基因组特异性引物共扩增得

到的ARE和ARQ型 ω -醇溶蛋白基因序列分别有3种(表2),不同类型和数量的Indel是造成不同基因编码区长度变异的主要因素。其中MN441497

和 MN441505 的编码区长度最长,为 1 158 bp。在其余 4 种 ARE/Q 型基因序列中,共有 5 种类型的 Indel,长度分别为 15、18、24、48 和 123 bp(图 3)。MN441503 中含有 123、48 和 15 bp 3 种类型 Indel; MN441504 中含有 18、48 和 15 bp 3 种类型 Indel; MN441496 中只含有 18 bp 的 Indel; MN441506 中只含有 24 bp 的 Indel。共有 70 处 SNP 位点分布于

这 6 种 ω -醇溶蛋白基因序列之间,其中 22 个 SNP 为同义突变,48 个 SNP 为非同义突变。这些 Indel 碱基数均为 3 的倍数,SNP 变异也没有引入额外的终止密码子,因此 6 种 ARE/Q 型 ω -醇溶蛋白基因序列均含有完整编码框。不同长度和数量的 Indel 组合以及不同的 SNP 变异,使基因呈现多态性。其中 MN441496 和 MN441497 只有 18 bp Indel 的差异。

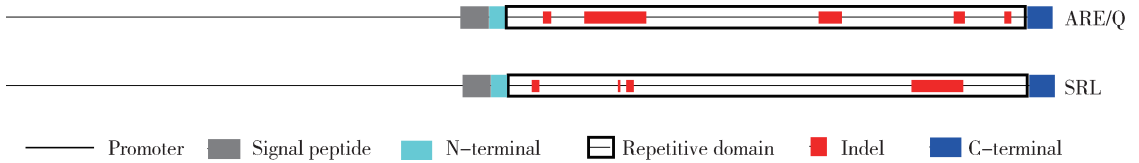


图 3 ARE/Q 型和 SRL 型 ω -醇溶蛋白基因模式图

Fig. 3 The schematic diagram of ARE/Q and SRL type ω -gliadin genes

MN441499 在 135 bp 处有 1 个 18 bp 的 Indel,在 151 bp 处插入了 1 个碱基 A,导致推导的氨基酸序列与醇溶蛋白相似性较低,推测其可能为移码突变的假基因。在另外 4 种 SRL 型 ω -醇溶蛋白基因的中间重复区共发现了 3 种类型的 Indel(图 3)。MN441500 和 MN441502 在 303 bp 处均有 1 个 3 bp 的 Indel, MN441498、MN441499、MN441500 和 MN441501 在相距 12 bp 后的位置有 1 个 18 bp 的 Indel, MN441498、MN441499、MN441501 和 MN441502 在相距 552 bp 后的位置有 1 个 102 bp 的 Indel,这些 Indel 碱基数均是 3 的倍数,没有引起移码突变。在 4 种 SRL 型 ω -醇溶蛋白基因中共发现 27 个 SNP 位点,其中有 7 个 SNP 为同义突变,17 个为非同义突变。还有 3 个位点的 SNP 引入了提前终止密码子,造成推导的氨基酸序列变短,其中 1 个 SNP 位点位于 MN441498 序列 ATG 下游 1 260 bp 处,三联体密码子 TAC 突变为 TAA;另外 2 个位点位于 MN441501 序列编码区 274 和 1 279 bp 处,编码脯氨酸的密码子 CAG 和 CAA 突变为终止密码子 TAG 和 TAA。

2.2.2 ω -醇溶蛋白基因的进化树分析

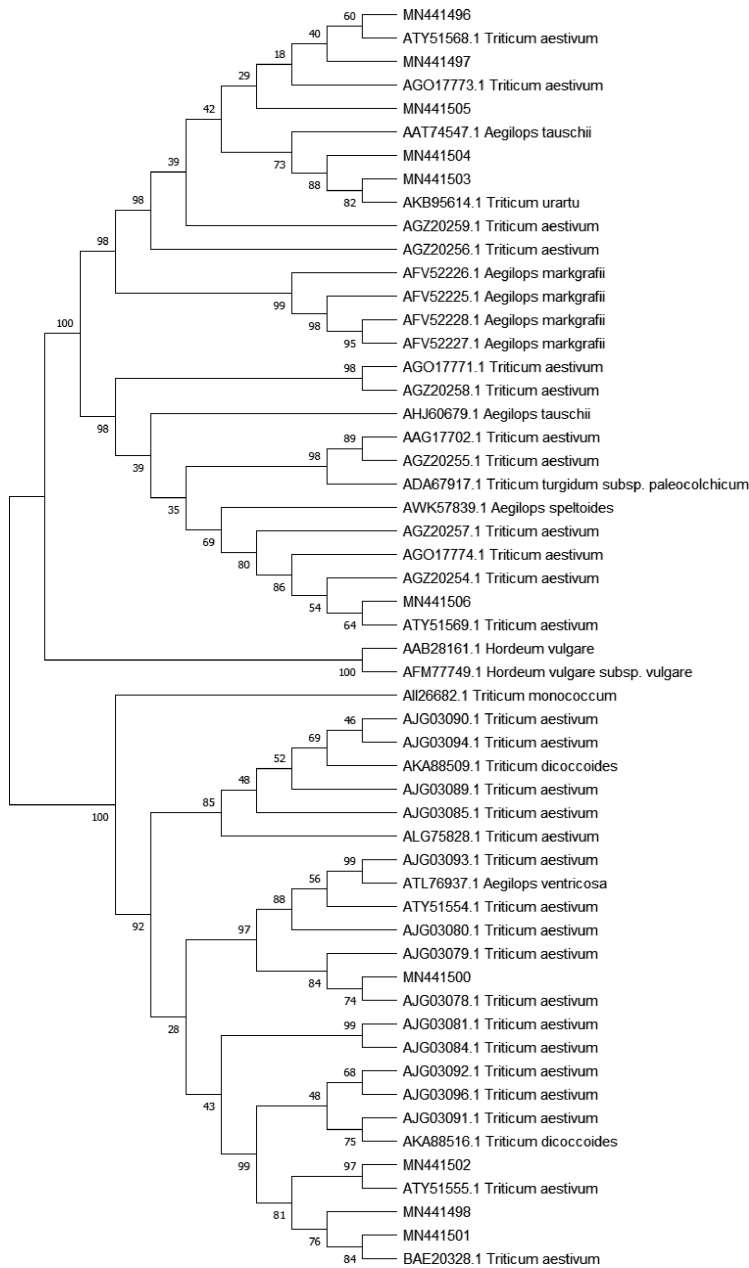
将克隆的 ω -醇溶蛋白基因序列在 NCBI 的 NR 数据库中进行 Blast,并与 44 条来源于不同小麦属具有完整编码框的 ω -醇溶蛋白基因构建 Neighbor-joining 系统发生树。从图 4 可知,系统发生树被分成 2 个分支,6 种 ARE/Q 型 ω -醇溶蛋白基因主要与普通小麦 (*Triticum aestivum* L.)、山羊草属 (*Aegilops tauschii* L.)、乌拉尔图小麦 (*Triticum urartu* L.) 及圆锥小麦 (*Triticum turgidum* L.) 中

的 ω -醇溶蛋白基因聚类在 1 个大分支上,而 4 种 SRL 型 ω -醇溶蛋白基因主要与普通小麦 (*Triticum aestivum* L.) 和二粒小麦 (*Triticum dicoccoides* L.) 中的 ω -醇溶蛋白基因聚在 1 个分支上。通过进化树分析可以看到 ARE/Q 型 ω -醇溶蛋白基因在进化上与 SRL 型 ω -醇溶蛋白基因相对独立。

2.2.3 ω -醇溶蛋白基因启动子分析

ω -醇溶蛋白基因与小麦中其他储藏蛋白基因启动子区类似^[17-18,22],ATG 上游序列包含有许多保守的 motif,其中 ARE/Q 型 ω -醇溶蛋白基因启动子含有 29 种 motif;SRL 型 ω -醇溶蛋白基因启动子含有 27 种 motif(表 3),相同的 motif 有 19 种,但是分布的位置不同,这些 motif 是 SPA、MYB、DOF 和 B3 等重要转录因子的识别序列。

在 ARE/Q 型 ω -醇溶蛋白基因中,MN441496 和 MN441497、MN441503 和 MN441504 以及 MN441505 和 MN441506 的启动子序列分别相同,SRL 型 ω -醇溶蛋白基因的 5 种启动子序列均不相同。ARE/Q 和 SRL 型基因启动子序列差异较大(表 3),但是存在保守的 motif 组合,比如 P-box 和 N-motif 组成的 endosperm box 只出现在 ARE/Q 型醇溶蛋白基因启动子序列的 -300 bp 处,但出现在 SRL 型醇溶蛋白基因启动子序列的 -300 bp 和 -600 bp 处。有些 motif 只特定出现在 1 种序列中,比如 ARE/Q 型 ω -醇溶蛋白基因 -137、-172、-389、-498、-499、-675、-946 和 -968 bp 处的 motif,SRL 型 -187、-367、-626 和 -933 bp 处的 motif。这些 motif 出现在不同基因启动子的不同位置,可能会对基因的表达产生不同的影响。



数字表示可信度。

The numbers in the figure indicate the credibility.

图4 克隆的 ω -醇溶蛋白基因与非冗余 ω -醇溶蛋白基因进化树分析

Fig. 4 Phylogenetic tree analysis based on the cloned genes and non-redundant ω -gliadins

表3 ARE/Q和SRL型醇溶蛋白基因启动子区保守motif分布

Table 3 The distribution of conserved motif in ARE/Q and SRL type ω -gliadin genes promoter region

Motif 名称 Name of motif	Motif 序列 Sequence of motif	转录因子 Transcriptional activator	ARE/Q 型 ARE/Q type	SRL 型 SRL type
RAV1	CAACA	AP2	-196(+), -371(+), -192(+), -252(+), -278(+), -376(+)	-192(+), -367(+), -379(+), -491(+), -679(+), -1 095(+)

表 3(续)

Motif 名称 Name of motif	Motif 序列 Sequence of motif	转录因子 Transcriptional activator	ARE/Q 型 ARE/Q type	SRL 型 SRL type
RAV1	TGTTG	AP2	-1 112(-)	
RY	CATGCA	B3	-94(+), -641(+)	-376(+)
skn-1	GTCAT	B3	-287(+), -294(+), -549(+)	-297(+), -597(+), -848(+)
skn-1	ATGAC	B3	-438(-), -846(-)	-666(-), -739(-), -889(-)
E-box	CAAATG	bHLH	-184(+)	-121(+)
E-box	CATTTG	bHLH		-278(+), -595(+)
E-box	CAATTG	bHLH	-683(+)	
E-box	CATATG	bHLH		-871(+), -892(+)
E-box	CACCTG	bHLH		-1 065(-)
ABRE	ACGTG	bZIP	-968(-)	-952(+)
N-motif	TGAGTCA	bZIP	-297(+), -552(+)	-300(+), -600(+), -851(+)
PAP	ACTCAT	bZIP		-895(+)
PAP	ATGAGT	bZIP	-298(-), -553(-)	-301(-), -601(-), -852(-)
POA	CCTTTT	DOF	-345(-)	
(CA) _n	CNAACAC	DOF	-198(+)	-197(+)
P-box	TGCAAAG	DOF	-137(+)	
P-box	TGTAAAGT	DOF	-313(+)	-316(+), -615(+)
BO	TGTCA	Homeo-domain	-796(+)	
BO	TGACA	Homeo-domain		-665(+), -738(+), -888(+)
AACA	AACAAAC	MYB	-191(+)	-187(+)
AACA	AACCA	MYB	-943(+)	
AACA	AACTAAC	MYB		-660(+)
AACA	GTTAGTT	MYB	-498(-)	-799(-)
TOA	TATCCA	MYB	-282(+), -537(+), -773(+)	-836(+)
TOA	TGGATA	MYB	-172(-), -389(-)	-392(-), -692(-)
I-box	GATAAG	MYB	-170(+), -387(+)	-169(+), -466(+)
I-box	CTTATC	MYB	-709(-), -875(-)	-838(-)
CCAAT	CCAAT	NF-Y	-395(+), -827(+), -1 057(+)	-398(+), -698(+)
CCAAT	ATTGG	NF-Y	-291(-), -519(-)	
SEF4	ATTTTTTA	SEF	-861(+)	
SEF4	ATTTTTTG	SEF		-626(+)
SEF4	GTTTTTTG	SEF		-127(-)
TATA	TATAAA	TBP	-104(+), -675(+)	-103(+), -805(+)
TATA	TTTATA	TBP	-1 002(-), -1 043(-)	
W-box	TGACT	WRKY	-845(+)	
W-box	AGTCA	WRKY	-225(-), -295(-), -499(-), -550(-), -946(-)	-221(-), -298(-), -598(-), -849(-), -933(-)

注: +表示 motif 位于启动子序列正链, -表示 motif 位于启动子序列负链。

Note: + indicates the sequence of motif located on the sense strand of promoter. - indicates the sequence of motif located on the antisense strand of promoter.

3 讨论

不同小麦材料中的 ω -醇溶蛋白基因拷贝数不同,研究表明,六倍体普通小麦中的 ω -醇溶蛋白基因数量在15~18个^[7]。根据重新组装的基因组信息、2-DE(双向蛋白电泳)和RNA-seq信息,‘CS’共有5种能够转录并且表达的 ω -醇溶蛋白基因^[16],因此在A-PAGE上可以看到5条清晰的 ω -醇溶蛋白条带(图1)。“Fielder”在A-PAGE上共有7条 ω -醇溶蛋白条带,推测至少应有7种能够表达的 ω -醇溶蛋白,因此根据‘CS’中 ω -醇溶蛋白基因的上下游序列信息,对‘Fielder’中的 ω -醇溶蛋白基因的编码区和启动子序列进行克隆,共得到了11条序列,其中1条为移码突变的假基因,2条含有提前终止密码子,8条具有完整编码框。由于 ω -醇溶蛋白有限的水解位点,质谱分析相对困难^[23],更为重要的是 ω -醇溶蛋白基因等位变异广泛,但数据库中完整的基因数量较少,大部分预测的蛋白分子量较小,可能是丢失了一部分中间重复区^[10]。本研究采用单克隆测序的方法,能够准确知道每条DNA序列的长度,并且通过拼接不同位置的测序结果,大大降低了缺失中间重复区的可能。这种通过简单的PCR进行克隆的方法,避免了2-DE和质谱等繁琐的试验过程,即可获得 ω -醇溶蛋白基因序列,简单易行且成本较低。缺点是需要避开高GC含量的中间重复区设计出合适的引物进行测序,并且由于重复单元较多,需要设计多对引物对测序结果反复验证,避免拼接结果缺失部分重复单元,另外得到的基因序列并不能与A-PAGE上的条带一一对应。

含奇数个半胱氨酸的醇溶蛋白可能作为链内终止剂,参与谷蛋白大聚体的形成^[3]。本研究克隆的 ω -醇溶蛋白基因中,MN441506推导的氨基酸序列含有1个半胱氨酸和甲硫氨酸,MN441502在C末端含有1个半胱氨酸,这2个 ω -醇溶蛋白可能通过分子间二硫键对谷蛋白大聚体的形成及稳定性产生影响,进而影响小麦品质性状。

*Glu-1*基因启动子中的顺式作用元件可以组成保守的顺式作用调控模块(conserved cis-regulatory modules,CCRM)对基因表达进行精确调控,并且表达量较高的x型HMW-GS除了保守的CCRM外,还含有能够被R2R3-MYB转录因子识别的motif^[16,22]。LMW-GS基因启动子也含有保守的非编码调控元件,大部分的s-和i型LMW-GS基因呈

现逐渐增加的表达模式,而m型LMW-GS基因则呈先升高再降低的表达模式,不同基因表达模式的差异与启动子区motif的数量和分布密切相关^[17]。本研究克隆的 ω -醇溶蛋白基因启动子区也含有大量保守的motif,不同种类基因之间motif的数量和分布不同(表3),相同种类基因之间也在一些motif上有差异。其中由GCN4和P-box这2种motif组成的endosperm box对储藏蛋白基因的表达起到非常重要的作用,LMW-GS基因启动子区GCN4和P-box motif的数量和组合多态性使同一类型的基因呈现不同的表达模式^[17]。在本研究克隆的 ω -醇溶蛋白基因中也发现了类似的现象, ω 1,2-醇溶蛋白基因启动子区含有这2种motif组成的3种组合类型,每种类型所含的motif种类和数量不同。 ω 1,2-醇溶蛋白基因启动子只含有1个典型的endosperm box,而 ω 5-醇溶蛋白基因则含有2个。这些motif的组合形式也可能会造成 ω -醇溶蛋白基因之间的差异表达,有待进一步研究。

参考文献 References

- [1] Altenbach S B, Kothari K M. Omega gliadin genes expressed in *Triticum aestivum* cv Butte 86: Effects of post-anthesis fertilizer on transcript accumulation during grain development [J]. *Journal of Cereal Science*, 2007, 46(2): 169-177
- [2] Masci S, Lafiandra D, Porceddu E, Lew E J L, Tao H P, Kasarda D D. D-Glutelin subunits: N-terminal sequences and evidence for the presence of cysteine [J]. *Cereal Chemistry*, 1993, 70(5): 581-585
- [3] Masci S, Egorov T A, Ronchi C, Kuzmicky D D, Kasarda D D, Lafiandra D. Evidence for the presence of only one cysteine residue in the D-type low molecular weight subunits of wheat glutenin [J]. *Journal of Cereal Science*, 1999, 29(1): 17-25
- [4] Zhang Y, Luo G, Liu D, Wang D, Yang W, Sun J, Zhang A, Zhan K. Genome-, transcriptome- and proteome- wide analyses of the gliadin gene families in *Triticum urartu* [J]. *PLoS One*, (2015-07-01). DOI: 10.1371/journal.pone.0131559
- [5] Barro F, Iehisa J C M, Giménez M J, García-Molina M D, Ozuna C V, Comino I, Sousa C, Gil-Humanes J. Targeting of prolamins by RNAi in bread wheat: Effectiveness of seven silencing-fragment combinations for obtaining lines devoid of coeliac disease epitopes from highly immunogenic gliadins [J]. *Plant Biotechnology Journal*, 2016, 14(3): 986-996
- [6] Hsia C C, Anderson O D. Isolation and characterization of wheat ω -gliadin genes [J]. *Theoretical and Applied Genetics*, 2001, 103(1): 37-44

- [7] Sabelli P A, Shewry P R. Characterization and organization of gene families at the Gli-1 loci of bread and durum wheats by restriction fragment analysis[J]. *Theoretical and Applied Genetics*, 1991, 83(2): 209-216
- [8] Tatham A S, Shewry P R. The S-poor prolamins of wheat, barley and rye: Revisited[J]. *Journal of Cereal Science*, 2012, 55(2): 79-99
- [9] Waga J, Skoczowski A. Development and characteristics of ω -gliadin-free wheat genotypes[J]. *Euphytica*, 2014, 195(1): 105-116
- [10] Anderson O D, Gu Y Q, Kong X, Lazo G R, Wu J. The wheat omega-gliadin genes: Structure and EST analysis[J]. *Functional and Integrative Genomics*, 2009, 9(3): 397-410
- [11] Dupont F M, Vensel W H, Chan R, Kasarda D D. Characterization of the 1B-type omega-gliadins from *Triticum aestivum* cultivar Butte[J]. *Cereal Chemistry*, 2000, 77(5): 607-614
- [12] Sabelli P A, Shewry P R. Characterization and organization of gene families at the Gli-1 loci of bread and durum wheats by restriction fragment analysis[J]. *Theoretical and Applied Genetics*, 1991, 83(2): 209-216
- [13] Dupont F M, Vensel W H, Tanaka C K, Hurkman W J, Altenbach S B. Deciphering the complexities of the wheat flour proteome using quantitative two-dimensional electrophoresis, three proteases and tandem mass spectrometry[J]. *Proteome Science*, (2019-05-16). DOI: 10.1186/1477-5956-9-10
- [14] Wan Y, Gritsch C S, Hawkesford M J, Shewry P R. Effects of nitrogen nutrition on the synthesis and deposition of the ω -gliadins of wheat[J]. *Annals of Botany*, 2014, 113(4): 607-615
- [15] Wang D, Li D, Wang J, Zhao Y, Wang Z, Yue G, Liu X, Qin H, Zhang K, Dong L, Wang D. Genome-wide analysis of complex wheat gliadins, the dominant carriers of celiac disease epitopes[J]. *Scientific Reports*, (2017-03-16). DOI: 10.1038/srep44609
- [16] Huo N, Zhang S, Zhu T, Dong L, Wang Y, Mohr T, Hu T, Liu Z, Dvorak J, Luo M, Wang D, Lee J, Altenbach S, Gu Y Q. Gene duplication and evolution dynamics in the homeologous regions harboring multiple prolamins and resistance gene families in hexaploid wheat[J]. *Frontiers in Plant Science*, (2018-05-23). DOI: 10.3389/fpls.2018.00673
- [17] Ravel C, Fiquet S, Boudet J, Dardevet M, Vincent J, Merlino M, Michard R, Martre P. Conserved cis-regulatory modules in promoters of genes encoding wheat high-molecular-weight glutenin subunits[J]. *Frontiers in Plant Science*, (2014-11-12). DOI: 10.3389/fpls.2014.00621
- [18] Juhász A, Makai S, Sebestyén E, Tamás L, Balázs E. Role of conserved non-coding regulatory elements in LMW glutenin gene expression[J]. *PLoS One*, (2011-12-29). DOI: 10.1371/journal.pone.0029501
- [19] Guo W, Yang H, Liu Y, Gao Y, Ni Z, Peng H, Xin M, Hu Z, Sun Q, Yao Y. The wheat transcription factor TaGAMyB recruits histone acetyltransferase and activates the expression of a high-molecular-weight glutenin subunit gene[J]. *Plant Journal*, 2015, 84(2): 347-359
- [20] Boudet J, Merlino M, Plessis A, Gaudin J, Dardevet M, Perrochon S, Alvarez D, Risacher T, Martre P, Ravel C. The bZIP transcription factor SPA heterodimerizing protein represses glutenin synthesis in *Triticum aestivum*[J]. *Plant Journal*, 2019, 97(5): 858-871
- [21] 王祖华, 杨瑞先, 姬云波. 小麦醇溶蛋白酸性聚丙烯酰胺凝胶电泳方法的优化及应用[J]. *中国农学通报*, 2009, 20(25): 54-57
- Wang Z H, Yang R X, Ji Y B. Improvement and its application of acid polyacrylamide gel electrophoresis of wheat gliadin[J]. *Chinese Agricultural Science Bulletin*, 2009, 20(25): 54-57 (in Chinese)
- [22] Li J, Wang K, Li G, Li Y, Zhang Y, Liu Z, Ye X, Xia X, He Z, Cao S. Dissecting conserved cis-regulatory modules of Glu-1 promoters which confer the highly active endosperm-specific expression via stable wheat transformation[J]. *The Crop Journal*, 2019, 7(1): 8-18
- [23] Cho K, Beom H, Jang Y, Altenbach S B, Vensel W H, Simon-Buss A, Lim S, Kim M G, Lee J. Proteomic profiling and epitope analysis of the complex alpha-, gamma-, and omega-gliadin families in a commercial bread wheat[J]. *Frontiers in Plant Science*, (2018-09-18). DOI: 10.3389/fpls.2018.00818

责任编辑: 吕晓梅