

# 玉米穗行数相关基因 *GRMZM2G398848* 的生物信息学分析

孙培元

(黑龙江省农业科学院 齐齐哈尔分院, 黑龙江 齐齐哈尔 161041)

**摘要:** 研究采用生物信息学的方法, 对一个玉米穗行数相关基因 *GRMZM2G398848* 的理化性质、跨膜结构域、功能域、二级结构、三级结构等方面进行了预测和分析。结果表明: 该基因编码蛋白由 9 195 个原子构成, 其相对分子质量为 65.681 1 kDa, 是一个不稳定亲水性蛋白, 包含 1 个 F-box 超亲家族和 2 个 AMN1 超亲家族, 无跨膜结构域, 无信号肽, 由  $\alpha$  螺旋、无规则卷曲、延伸链和  $\beta$  折叠构成。

**关键词:** 玉米; 穗行数; *GRMZM2G398848* 基因; 生物信息学

**中图分类号:** S513 **文献标识码:** A **文章编号:** 1002-2767(2016)11-0004-04 DOI:10.11942/j.issn1002-2767.2016.11.0004

穗行数(kernel row number, KRN)是玉米产量的主要构成因素, 其由小穗成对分生组织数目决定。该性状为数量性状遗传, 具有较高的广义遗传力, 因此对于玉米高产育种来说解析其遗传方式具有十分重要的意义<sup>[1]</sup>。Lu 等以掖478 $\times$ 丹340 的 F<sub>2,3</sub> 群体为材料, 在 bnlg339-umc1865 标记区间上(bin7.03)检测到一个控制玉米穗行数的主效 QTL, 能够解释 17.86% 的表型变异, 并在 7 个多样性环境中稳定表达<sup>[2]</sup>。周强等利用元分析方法将 26 个不同双亲分离群体定位的 176 个穗行数 QTL 整合到参考图谱 IBM 2008 Neighbors 上, 发掘出 25 个“一致性”与效应值大于 4.5 的 QTL, 其中分别在 bin1.09、bin2.07、bin4.09、bin5.01、bin5.03、bin7.02、bin9.06 和 bin10.04 上检测到控制穗行数的主效 QTL<sup>[3]</sup>。Liu 等对在 bin4.08 上控制玉米穗行数的主效 QTL(KRN4)进行了精确定位和图位克隆, 同时对其进行了基因表达分析和关联分析验证, 发现 UB3 中的 A/GSNP(S35)与 KRN4 位点关联显著<sup>[4]</sup>。Brown 等通过全基因组关联分析和联合连锁的方法, 利用巢式关联群体(NAM)检测到了 36 个与玉米穗行数相关的 QTL<sup>[5]</sup>。张焕欣等以 203 份主要玉米自交系为材料, 采用分布于玉米基因组的 41 101 个 SNP 标记对穗行数进行关联分析, 在显著 SNP 位点 LD 区域内成功发掘出 4 个候选基因<sup>[6]</sup>。

目前, 已有诸多学者利用分子生物学手段对玉米穗行数进行研究。然而, 大多数研究均集中在玉米穗行数主效 QTL 的挖掘, 而关于控制穗行数基因功能的研究则鲜有报道。随着生物信息学的高速发展以及玉米基因组测序的完成, 利用生物信息学相关软件进行快速有效地分析已成为当今在基因结构及功能研究领域主流的研究方法<sup>[7-8]</sup>。因此, 研究以玉米基因组中一个玉米穗行数相关基因 *GRMZM2G398848* 的序列为材料, 采用生物信息学的分析方法针对该基因的结构及功能进行系统性分析和预测, 旨在研究该基因在玉米生长发育中的调节机制, 以期对玉米穗行数相关基因的功能研究提供理论依据。

## 1 材料与方法

### 1.1 材料

根据张焕欣对玉米穗行数进行的全基因组关联分析结果<sup>[6]</sup>, 研究选取 *GRMZM2G398848* 基因作为试验材料, 其序列来自于美国国立生物技术信息中心网站(National Center for Biotechnology Information, NCBI)(<http://www.ncbi.nlm.nih.gov/>), 登录号为 DAA43782.1。

### 1.2 方法

试验采用在线软件 Protparam(<http://web.expasy.org/protparam/>)分析 *GRMZM2G398848* 基因编码蛋白的基本理化性质; 采用在线分析工具 NCBI-CDS(<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi/>)对其保守结构域进行分析; 采用在线软件 PSORT(<http://psort.hgc.jp/form.html>)对其进行亚细胞定位; 采用在线软

收稿日期: 2016-08-21

作者简介: 孙培元(1991-), 男, 黑龙江省绥化市人, 硕士, 研究实习生, 从事玉米分子遗传育种研究。E-mail: 562337271@qq.com。

件 TMHMM Server v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) 寻找其跨膜结构; 采用在线分析工具 SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>) 对其信号肽进行预测; 采用 NetPhos 2.0 Serve 数据库 (<http://www.cbs.dtu.dk/services/NetPhos/>) 对其磷酸化位点进行预测; 利用 Coils 在线分析工具 ([http://www.ch.embnet.org/software/COILS\\_form.html](http://www.ch.embnet.org/software/COILS_form.html)) 对其形成卷曲螺旋的倾向性进行预测; 采用 SOPMA 数据库 ([http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_sopma.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)) 对其二级结构进行分析; 采用在线软件 SWISS-MODEL (<http://swissmodel.expasy.org/>) 对其三级结构进行分析; 采用 MEGA 6.0 软件将其同其它植物蛋白进行同源比对并构建系统进化树。

## 2 结果与分析

### 2.1 GRMZM2G398848 基因编码蛋白的理化性质预测

利用分析软件 Protparam 对 GRMZM2G398848 基因编码蛋白的理化性质预测结果显示, 该蛋白

分子式为  $C_{2895}H_{4586}N_{820}O_{863}S_{31}$ , 由 9 195 个原子构成, 相对分子质量 65.681 1 kDa, 理论等电点(pI)为 5.62, 带负电荷的氨基酸残基总数(Asp+Glu)为 72 个, 带正电荷的氨基酸残基总数(Arg+Lys)为 59 个, 总平均亲水性系数为 -0.040, 属于亲水性蛋白, 理论推导半衰期是 30 h, 不稳定系数为 49.38, 这说明该蛋白是不稳定蛋白。

### 2.2 GRMZM2G398848 基因编码蛋白保守结构域分析

结构功能域是介于二级与三级结构之间、且具有一定功能的紧密结构单位, 是在蛋白质三级结构内的独立折叠单元。利用 NCBI-CDS 在线分析工具预测 GRMZM2G0398848 基因编码蛋白中可能存在的保守结构域(见图 1)。结果表明, 该基因编码蛋白包含 2 个 AMN1 超亲家族(AMN1 superfamily)和 1 个 F-box 超亲家族(F-box superfamily)。F-box 结构域, 即 SCF 复合体的 F-box 亚基, 它是 E3 连接酶识别底物蛋白所必需的亚基蛋白, 通过介导一系列生长素反应来参与植物生长发育的调控。

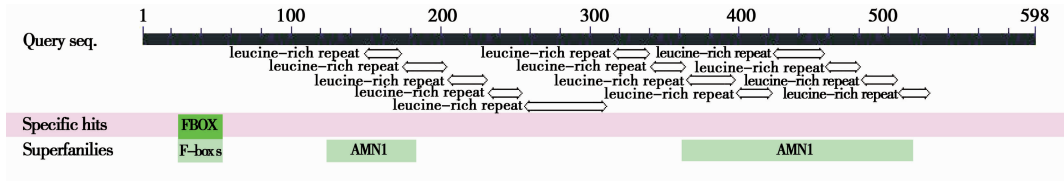


图 1 GRMZM2G398848 基因编码蛋白的保守结构域

Fig.1 The conserved domains of GRMZM2G398848 coding protein

### 2.3 GRMZM2G398848 基因编码蛋白磷酸化位点预测

蛋白质磷酸化在细胞的信号传递过程中起到不可忽视的作用。利用 NetPhos 2.0 对 GRMZM2G398848 基因编码蛋白的磷酸化位点进行预测。结果表明, 在整个多肽链中的氨基酸磷酸化位点(>0.5)一共有 26 个, 其中包括 21 个丝氨酸(Ser)磷酸化位点、4 个苏氨酸(Thr)磷酸化位点和 1 个酪氨酸(Tyr)磷酸化位点。研究表明, 在该蛋白的磷酸化过程中, 丝氨酸磷酸化从中起到主导性作用。因此, 推测在 GRMZM2G398848 基因编码蛋白中丝氨酸磷酸化对其行使功能存在重要作用。

### 2.4 GRMZM2G398848 基因编码蛋白亚细胞定位

利用在线软件 PSORT 对 GRMZM2G398848

基因编码蛋白进行亚细胞定位(见表 1)。结果表明, 该蛋白定位于细胞质的可能性最大, 其可信值为 0.450; 微体(过氧化物酶体)其次, 可信值达 0.300; 该蛋白在线粒体基质空间和叶绿体类囊膜上也有分布, 但可信值较低, 均为 0.100。由此推测该蛋白属于细胞质蛋白。

表 1 GRMZM2G398848 基因编码蛋白的亚细胞定位

Table 1 Predicted subcellular localization of GRMZM2G398848

亚细胞结构 Destination of cell compartment	可信度 Certainty
细胞质 Cytoplasm	0.450
微体(过氧化物酶体) Microbody( Peroxisome)	0.300
线粒体基质空间 Mitochondrial matrix space	0.100
叶绿体内囊体膜 Chloroplast thylakoid membrane	0.100

## 2.5 GRMZM2G398848 基因编码蛋白跨膜区域及信号肽预测

利用 TMHMM Server v. 2. 0 在线分析软件对 GRMZM2G398848 基因编码蛋白进行跨膜结构域分析。结果表明,该蛋白不存在跨膜结构域,位于膜外区的可能性最大(可能性=1. 1),为非跨膜蛋白类。利用网络分析工具 SignalP 4. 1 分析 GRMZM2G398848 基因编码蛋白信号肽发现,C 值(剪切位点值)为 0. 116,在该蛋白中未发现可能的信号肽序列。由此说明,GRMZM2G398848 基因编码蛋白为非分泌型蛋白,因此其并没有在细胞中发生迁移。参考亚细胞定位结果,该蛋白可能是在细胞质中合成后,直接在细胞质中行使功能。

卷曲螺旋 (coiled coil, CC) 在蛋白质结构与功能的关系上具有极其重要的作用, 是一类由 2 股或 2 股以上右手  $\alpha$  螺旋相互缠绕而形成的平行或反平行左手超螺旋结构。利用 Coils 分析工具, 以 window=14、21 和 28 为参数, 按照几率 > 50% 就可形成螺旋的规则, 对 GRMZM2G398848 基因编码蛋白形成卷曲螺旋的倾向性进行预测。结果表明, 该蛋白在 windows=21 时极易形成卷曲螺旋, 在 windows=14 时可能形成卷曲螺旋, 而在 windows=28 时不能形成卷曲螺旋结构。

## 2.6 GRMZM2G398848 基因编码蛋白二级结构预测

利用 SOPMA 服务器对 GRMZM2G398848 基因编码蛋白的二级结构进行预测分析(见图 2)。结果表明,该蛋白的二级结构主要是由  $\alpha$ -螺旋(46.82%)组成,其次是无规则卷曲(30.94%),再次是由延伸链区(14.55%)和  $\beta$ -折叠(7.69%)构成。这说明该蛋白为 mixed 型蛋白。

### 2.7 GRMZM2G398848 基因编码蛋白三级结构预测

蛋白质空间结构的预测对于生物学研究具有重大意义。对于氨基酸线性序列来说,只有按照特定的空间构象对其进行折叠,才能发挥该序列所应有的生物学功能和活性。因此,蛋白质的空间结构的预测对于蛋白质的功能及其机制方面的了解会起到很大程度上的帮助作用。*GRMZM2G398848* 基因编码蛋白的三级结构模

型见图 3。该蛋白主要由  $\alpha$ -螺旋盘绕而成,其次是  $\beta$ -折叠,中间则由无规则卷曲连接。这与之前二级结构预测结果相吻合。

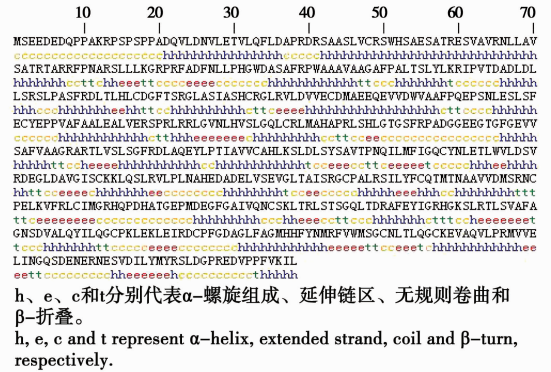


图 2 GRMZM2G398848 基因编码蛋白的二级结构

Fig. 2 Secondary structure of GRMZM2G398848

coding protein

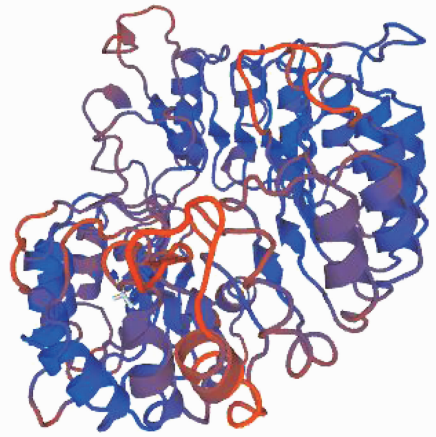


图 3 GRMZM2G398848 基因编码蛋白的三级结构

Fig. 3 The predicted 3-D structure of *GRMZM2G398848*

coding

## 2.8 GRMZM2G398848 基因编码蛋白系统进化树的构建

为了研究该基因编码蛋白的起源和进化关系,对其编码的氨基酸序列在 NCBI 网站中进行 Blast,最终选取 8 条与之相似度较高的序列并下载,通过 MEGA6.0 软件中的 ClustalW 功能进行氨基酸序列比对并采取邻位连接法(Neighbor-Joining, NJ),构建 GRMZM2G398848 蛋白与其它 8 种蛋白间的系统进化树。由图 4 可知,GRMZM2G398848 蛋白与菠萝 F-box 基因 *FBX14* 编码蛋白聚为一类,其在进化上的分组可能揭示其在功能上的异同。

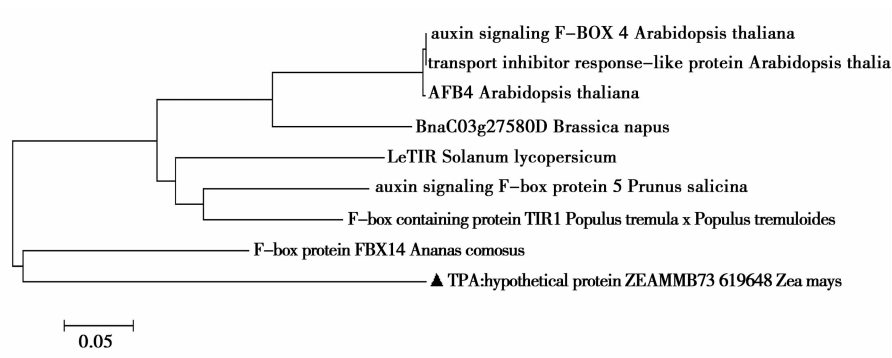


图 4 GRMZM2G398848 基因编码蛋白系统进化树

Fig. 4 Phylogenetic tree of GRMZM2G398848

### 3 结论与讨论

玉米是世界主要的粮饲作物之一,其产量在所有作物中占据首位。我国则在推动世界玉米生产上起到重要作用,国家统计局数据显示,截止至2014 年我国玉米单位面积产量可达 5 817 kg·hm<sup>-2</sup>,播种面积共计 3 707. 6 万 hm<sup>2</sup>,现已在我国占据着不可或缺的地位<sup>[9-11]</sup>。因此,对于育种工作者来说,确保我国玉米生产安全、持续、稳定地发展责任重大。随着时间的推移,人口增长、能源短缺以及畜牧业的发展等问题导致玉米的市场需求量在日益增长。因此,如何提高玉米单位面积产量便成为了解决这些问题的关键点。众所周知,玉米单位面积产量由单位面积有效穗数、穗粒数和粒重构成,其中,穗粒数又包括穗行数和行粒数。穗行数的遗传方式属于数量性状遗传,具有较高的广义遗传力,可通过决定穗粒数进而对产量造成影响<sup>[12]</sup>。因此,针对穗行数的研究在玉米高产育种中有着十分重要的意义。

本研究借助相关数据库和软件对一个与玉米穗行数相关的基因 GRMZM2G398848 进行了详尽的生物信息学分析和预测。其中,在进行保守结构域分析时,发现该基因属于 F-box 蛋白。研究表明,TIR1 编码的 F-box 蛋白为生长素受体,其与 AUX/IAA 作用会导致 AUX/IAA 为 SCFTIR1E3 介导的蛋白降解途径所降解<sup>[13-15]</sup>。这说明 GRMZM2G398848 基因编码蛋白可能与生长素的调控有关,进而对玉米穗行数产生影响。在 GRMZM2G398848 蛋白与其它物种蛋白间的系统进化树构建结果则表明,GRMZM2G398848 基因编码蛋白与菠萝 F-box 基因 FBX14 编码蛋白的功能与反应机制较为类似,可以根据基因 FBX14 进而推测 GRMZM2G398848 基因编码蛋白的功能,这对未来进一步研究该基因的功能奠

定了一定的基础。

#### 参考文献:

[1] 白胜双,彭勃,王超楠. 玉米穗行数遗传基础的研究进展[J]. 天津农业科学,2016(5):8-10.

[2] Lu Ming,Xie Chuanxiao,Li Xinhai ,et al. Mapping of quantitative trait loci for kernel row number in maize across seven environments [J]. Molecular Breeding, 2011, 28 (2): 143-152.

[3] 周强,王平喜,程备久,等. 玉米穗行数性状 QTL 的元分析[J]. 玉米科学,2014,22(2): 35-40.

[4] Liu Lei,Du Yanfang,Shen Xiaomeng,et al. KRN4 controls quantitative variation in maize kernel row number[J]. Plos Genetics,2015,11(11):e1005670.

[5] Brown P J,Upadavayula N,Mahone G S,et al. Distinct genetic architectures for male and female inflorescence traits of maize[J]. Plos Genetics,2011,7(11):1276-1280.

[6] 张焕欣,翁建峰,张晓聪,等. 玉米穗行数全基因组关联分析[J]. 作物学报,2014(1):1-6.

[7] 于滔,王成波,曹士亮,等. 玉米 bZIP 转录因子的生物信息学分析[J]. 黑龙江农业科学,2016(4):1-5.

[8] 孙培元,王俊强,韩业辉,等. 玉米 RNA 结合蛋白 GRMZM2G052926 基因的生物信息学分析[J]. 黑龙江农业科学,2016(3):7-12.

[9] 习银生,杨丽. 2014 年玉米市场形势分析与 2015 年展望[J]. 农业展望,2015,11(1):8-14.

[10] 杨军,仇焕广,董婉璐,等. 2014 年国内外玉米市场分析 & 2015 年展望[J]. 农业展望,2015,11(3):10-14.

[11] 杨军,仇焕广,董婉璐,等. 2013 年国内外玉米供需分析及 2014 年展望[J]. 农业展望,2014,10(2):4-8.

[12] 张焕欣. 玉米穗行数关联分析与连锁定位[D]. 北京:中国农业科学院,2014.

[13] Dharmasiri N,Dharmasiri S,Estelle M. The F-box protein TIR1 is an auxin receptor[J]. Nature,2005,435: 441-445.

[14] Gray W M,del Pozo J C,Walker L,et al. Identification of an SCF ubiquitin-ligase complex required for aulin response in Arabidopsis thaliana[J]. Genes & Development, 1999,13: 1678-1691.

[15] Cardozo T,Pagano M. The SCF ubiquitin ligase: insights into a molecular machine[J]. Nature reviews Molecular cell biology,2004,5:739-751.

# 油用向日葵 *FAD2-1* 基因 RT-PCR 体系的建立与优化

周 菲,刘 岩,黄绪堂,梁春波,王文军,李 岑,郭永利  
(黑龙江省农业科学院 经济作物研究所,黑龙江 哈尔滨 150086)

**摘要:**为了建立向日葵 *FAD2-1* 基因高效特异的 PCR 扩增体系,以油用向日葵 86-1 为材料提取总 RNA,并反转录合成 cDNA,以 cDNA 第一条链为模板对 *FAD2-1* 基因全长编码序列的 PCR 体系进行了优化。结果表明:根据已公布的 *FAD2-1* 序列设计 3 对特异性引物,每对引物设计了 5 个退火温度,筛选出适于该基因 PCR 扩增的引物及其退火温度,并将 PCR 扩增条件中循环数缩短为 30 次,延伸时间缩短为 7 min,此改进提高了 PCR 反应的效率。

**关键词:**油用向日葵;*FAD2-1*;RT-PCR

**中图分类号:**S565.5 **文献标识码:**A **文章编号:**1002-2767(2016)11-0008-03 **DOI:**10.11942/j.issn1002-2767.2016.11.0008

向日葵属菊科向日葵属,是世界上最重要的油料作物之一,油用型向日葵生产的葵花油不饱和脂肪酸含量高于大豆、花生、油菜等油料作物,是优质的保健食用油。 $\Delta^{12}$ -脂肪酸脱氢酶基因 *FAD2-1*,其主要作用是在单不饱和脂肪酸油酸(Oleic Acid,OA,18:1 $\Delta^9$ )的第 12 和 13 位碳原子之间插入一个双键,形成含有 2 个双键的多不饱和脂肪酸-亚油酸(Linoleic Acid,LA,18:2 $\Delta^9,12$ )<sup>[1-2]</sup>。为建立向日葵 *FAD2-1* 基因高效特异的 PCR 扩增体系,本研究以油用向日葵 86-1 籽仁为材料,对扩增 *FAD2-1* 基因全长编码序列的 RT-PCR 反应进行优化,为后续克隆该基因全长奠定基础。

## 1 材料与方法

### 1.1 材料

以黑龙江省农业科学院经济作物研究所自育的油用向日葵保持系 86-1 的籽仁为试验材料,于开花后 12 d 进行取材,迅速放入液氮中冷冻,之后冻存于-80℃冰箱中备用。

**收稿日期:**2016-09-20  
**基金项目:**黑龙江省农业科技创新工程资助项目(2014QN020);哈尔滨市科学技术局科技创新人才资助项目(2013RFQYJ027);哈尔滨市科学技术局科技创新人才资助项目(2013RFQYJ034);国家向日葵现代产业技术体系资助项目(CARS-16)  
**第一作者简介:**周菲(1984-),女,辽宁省丹东市人,硕士,助理研究员,从事向日葵研究工作。E-mail:zhoufei66666@163.com。

# Bioinformatics Analysis of the Gene *GRMZM2G398848* Related to Kernel Row Number in Maize

SUN Pei-yuan  
(Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar, Heilongjiang 161041)

**Abstract:** The characters of the gene *GRMZM2G398848* related to kernel row number was analyzed by bioinformatics tools, including the physical and chemical properties, transmembrane domain, subcellular localization, functional domain and 3D structure. The results showed that this protein was an unstable hydrophilic protein containing 9 195 atoms. The molecular weight of the protein containing a F-box superfamily and two AMN1 superfamilies was 65.681 1 kDa. It was no signal peptide, no transmembrane domain, and contained alpha-helix, random coil, extended strand and  $\beta$  folding.

**Keywords:** maize; kernel row number; *GRMZM2G398848*; bioinformatics analysis