

茄子谷氨酸酯-1-半醛 2,1 氨基变位酶 基因克隆和序列分析

余桃^{1,2}, 孙晶¹, 刘莉¹, 邹梦倩¹, 王倩^{1*}, 张爱慧^{1*}

(1. 金陵科技学院园艺园林学院, 江苏 南京 210038; 2. 安徽农业大学园艺学院, 安徽 合肥 230000)

摘要: 基于茄子基因组数据库, 以番茄同源基因序列为探针, 通过电子克隆手段获得茄子谷氨酸酯-1-半醛 2,1 氨基变位酶基因, 命名为 SmGSA。对该基因编码蛋白从氨基酸组成、蛋白的各级结构及其理化性质等方面进行了预测和分析。结果表明, 该基因包含一个长度为 1 044 bp 的开放阅读框, 共编码 347 个氨基酸, 蛋白质分子式预测为 $C_{3137}H_{5232}N_{1044}O_{1322}S_{211}$, 分子量大小为 85.5 KDa, 为亲水性蛋白; α -螺旋、延伸链和无规卷曲是该蛋白结构的主要构成元件; 该蛋白中共存在 9 处跨膜区, 内部到外部螺旋发现 4 个, 外部到内部螺旋发现 5 个。

关键词: 茄子; GSA 基因; 电子克隆; 生物信息学分析

中图分类号: S641.1; S603

文献标识码: A

文章编号: 1672-755X(2019)03-0073-05

Gene Cloning and Sequence Analysis of Glutamate-1-semialdehyde 2,1-aminomutase in Eggplant

YU Tao^{1,2}, SUN Jing¹, LIU Li¹, ZOU Meng-qian¹, WANG Qian^{1*}, ZHANG Ai-hui^{1*}

(1. Jinling Institute of Technology, Nanjing 210038, China; 2. Anhui Agricultural University, Hefei 230000, China)

Abstract: This study was based on the genome database of eggplants (*Solanum melongena* L.). The tomato homologous gene sequence was used as a probe to obtain the Glutamate-1-semialdehyde 2,1-aminomutase (GSA) gene of eggplant within silico cloning method and named it SmGSA. This study tried to predict and analyze the protein encoded by the gene from the aspects of amino acid composition, protein structure and physicochemical properties. The results showed that the gene contained an open reading frame with a length of 1044 bp, encoding 347 amino acids in total. The protein formula was predicted to be $C_{3137}H_{5232}N_{1044}O_{1322}S_{211}$ and weighs 85.5 kilodaltons (KDa), indicating that the protein was hydrophilic. Alpha-helix, extended chain and random curl were the main components of the protein structure. There were nine transmembrane regions in the protein, including four internal-external helices and five outward-to-inner spiral.

Key words: eggplant; GSA gene; in silico cloning; bioinformatics analysis

收稿日期: 2019-04-13

基金项目: 江苏省高等学校大学生创新创业训练计划项目 (201713573068X); 江苏省自然科学基金青年项目 (BK20130093)

作者简介: 余桃 (1996—), 女, 贵州遵义人, 硕士研究生, 主要从事园艺作物遗传育种与分子生物学研究。

通信作者: 王倩 (1984—), 女, 河北邢台人, 讲师, 博士, 主要从事园艺作物遗传育种与生物信息学研究。

张爱慧 (1971—), 女, 山东巨野人, 副教授, 硕士, 主要从事设施园艺教学与科研工作。

茄子(*Solanum melongena* L.)为茄科茄属一年生草本植物,是重要的世界性蔬菜作物。茄子的营养价值很高,其所含的维生素 P 能增强人体细胞间的黏着力^[1-2],增强毛细血管的弹性,降低脆性及渗透性,有防止出血和抗衰老等功能^[3]。

植物叶绿素的生物合成过程复杂,涉及多种酶参与。有研究指出拟南芥中参与合成叶绿素的酶达 15 种^[4],如谷氨酰-tRNA 还原酶、Mg-螯合酶 H 亚基、原叶绿素酸酯氧化还原酶和谷氨酸酯-1-半醛 2,1 氨基变位酶等。从 L-谷氨酰-tRNA 到原卟啉 $\dot{U}$ (proto $\dot{U}$)的生物合成步骤中,谷氨酸酯-1-半醛 2,1 氨基变位酶(GSA)参与并发挥了重要的作用^[5]。叶绿素合成相关酶基因的克隆研究,对于从分子水平研究蔬菜作物光合作用机理具有重要意义。目前已有大豆^[5]、拟南芥^[6]、油菜^[7]等多个作物克隆出 GSA 基因,而有关茄子 GSA 基因的克隆及分析尚未见报道。

植物叶绿体的生长发育代谢是一个较为复杂的调控网络系统,鉴定出茄子中与叶绿体发育相关的基因功能,对于研究园艺作物叶绿体生长发育代谢网络中各基因的关系以及揭示茄子叶片中叶绿素合成的分子调控机制都具有重要指导意义。

1 试验方法

1.1 电子克隆获得茄子 SmGSA 基因序列

在美国国立生物技术信息中心(NCBI)网站的基因数据库中搜索番茄同源基因序列,将番茄 GSA 基因(Gene ID:543967)序列与整个茄科作物基因数据库(<https://solgenomics.net/>)进行比对,首先选择工具栏(Tools)内的 Blast 程序放入该番茄同源序列,将比对得到的多条序列分别通过 BLAST 软件进行在线分析,并根据基因注释分析确定茄子 GSA 基因,命名为 SmGSA。

1.2 基因序列分析及同源比对

通过 NCBI 开放读码框搜索软件 ORF finder(<http://www.ncbi.nlm.nih.gov/gof/gorf.html>)寻找克隆到的茄子 GSA 基因序列开放阅读码。进一步使用 DNAMAN 软件将其与烟草、番茄等茄科作物的 GSA 基因进行同源性比对分析。

1.3 SmGSA 基因编码蛋白的生物信息学分析

使用相关在线软件 SMS(<http://www.bio-soft.net/sms/>)对茄子 SmGSA 的基因序列进行蛋白质翻译,使用在线软件 EXPASY(<http://www.expasy.org/tools/protparam.html>)进行一级蛋白结构预测,使用在线软件 PRABI(https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_mlrc.html)进行二级蛋白结构预测,使用 SWISS-MODEL(<https://www.swissmodel.expasy.org/>)软件进行三级蛋白结构预测。TMpred 软件(http://www.ch.embnet.org/software/TMPRED_form.html)用来预测蛋白跨膜结构,ProtScale 软件(<http://web.expasy.org/protscale/>)用来分析蛋白的亲疏水性理化性质等。

2 结果与分析

2.1 SmGSA 基因序列及同源性分析

使用在线软件 SMS 中的序列翻译工具分析可知,茄子 SmGSA 基因全长 1 044 bp,含有 1 个编码 347 个氨基酸的完整开放阅读框,序列如图 1,起始密码子 ATG 位于 1 bp 处,终止子 TAG 位于 1 044 bp 处。该序列未包含 3'非编码区(Untranslated Regions,UTR)和 5'非编码区。

利用 DNAMAN 软件分析该基因与番茄(AK319622.1)、野生种潘那利番茄(XM_015217690.1)、辣椒(XM_016708785.1)、烟草(XM_016653534.1)、野生烟草(XM_019375913.1)、栓皮栎(XM_024071748.1)的同源性,结果分别为 94%、94%、93%、91%、90%、80%。

进一步通过软件构建茄子 SmGSA 与数据库中已有的 6 种 GSA 氨基酸序列的系统进化树,由图 2 可知整个系统进化树分为 2 支,其中茄子 SmGSA 与番茄、野生种番茄亲缘关系最近,与烟草和辣椒等茄科作物聚为一大支,但在茄科作物分支中与辣椒亲缘关系较远,茄子 SmGSA 与栓皮栎亲缘关系最远。

1 M K K G T S F G A P C L L E N T L A E M
1 ATGAAGAAAGGAACGAGCTTCGGTGCATGCTCCTTGAAAATACACTGGCAGAGATG
21 V I S A V P S I E M V R F V N S G T E A
61 GTTATTTCAAGCTGTTCCAGCATAGAAATGGTTCGATTTGTCAACTCTGGTACAGAGCA
41 C M G V L R L A R A F T G R S K I V K F
121 TGTATGGAGTGCTACGGCTGGCTGCTGCTTTCACTGGCGTTCTAAGATCGTCAAAATTT
61 E G C Y H G H A D P F L V K A G S G V A
181 GAGGGTTGTTACCATGGTCAATGCTGATCCATTTCTGTTAAAGCAGGCGTGGGGTTGCC
81 T L G L P D S P G V P K A A T S E T L T
241 ACCGTCGGTCTCCCGCACTCCCGAGGAGTTCGGAAGGCTGCTAGTTCGGAACCTCTGACA
101 A P Y N D I S A V E R L F E E N K G E I
301 GCACCATACAATGATATTTCTGCTGTTGAGAGGCTCTTTGAGGAGAACAAAGGAGAAATTT
121 A A L I L E P V V G N A G F I P P K L E
361 GCCGCTCAATTCTCGAGCCTGTTGTGGGAAATGCTGGCTTCATCCCAAGCTAGAGAA
141 F L A A I R K I T K E N G A L L I F D E
421 TTTCTTGCTGCTATTCGCAAGATCAGCAAGAGAAATGGCGCACTTCTTATTTTCGATGAA
161 V M T G F R L A Y G G A Q E Y F G I T P
481 GTTATGACTGGATTTAGGCTTGCTTATGGCGAGCTCAAGAGTATTTGGAATACTCCA
181 D L T T L G K I I G G G L P V G A Y G G
541 GATTGACAACTCGGGAAGATTTAGTGGTGGCTTGCCAGTAGGTGCATACGGAGGA
201 R R D I M E L V A P A G P M Y Q A G T L
601 AGACGGGATATCATGGAATTTGGTAGCACTGCAGGACCAATGTATCAAGCTGGGACCTA
221 S G N P L A M T A G I H T L K R L Q G Q
661 AGTGGAAATCCATTGGCCATGACTGCTGGAATTCACACGCTTAAGCGGTTACAGGACAG
241 G T Y E H L D K I T A E L T Q G I L D A
721 GGAACATATGAACACTTGGACAAGATCAGCTGAACCTACACAAGGAATCTTGGATCGG
261 G K K T G H A M C G G Y I R G M F G F F
781 GGAAGAAGACTGGTCAATGCTGCGCGGGTACATAAGGGGATGTTTGGCTTCTTC
281 F T D G P V H N F S D A K K S E T E K F
841 TTCACGGATGGTCCCGTCCATCACTTTTCGGATGCAAGAAGAGTGAACAGAAAAATTT
301 G R F Y R G M L E E G V Y F A P S Q F E
901 GGCAGATTCTATAGAGGATGCTGGAGGAAGGTGTATTTTGGCTCATCAGATTTGAG
321 A G F T S L A H S P E D I Q K T V A A A
961 GCTGGATTACTAGCTTAGCAGACTCTCCAGAGGATATTCAAAAGACTGTAGCTGCTGCT
341 E K V L K K I *
1021 GAAAAAGTTTGAAGAAATTTAA

图 1 电子克隆获得的茄子 SmGSA 基因序列及其推导的氨基酸序列

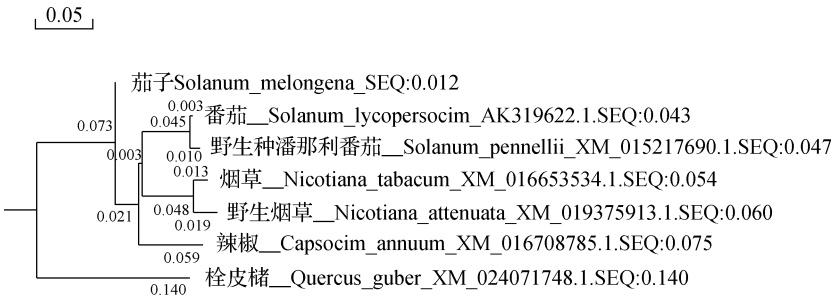


图 2 植物中 GSA 基因的系统进化树

2.2 SmGSA 蛋白一级结构预测和分析

茄子 SmGSA 基因序列在线翻译的结果表明:茄子 SmGSA 基因预测分子式为 C_{3 137} H_{5 232} N_{1 044} O_{1 322} S₂₁₁,共翻译 347 个氨基酸,蛋白质分子量为 85. 5 KDa,限制性酶切位点最多为 AluI ag|ct,分布位置在 18、71、415、515、650、752、975、1 013 bp,其他一级结构预测数据如表 1。

表 1 茄子 SmGSA 蛋白的一级结构预测数据结果

一级结构特性	预测结果
编码氨基酸个数	347
理论等电点(PI)	5. 08
分子量(MW)/Da	85 500
不稳定系数(II)	46. 99
脂肪族指数(AI)	27. 20
亲水性平均值(GRAVY)	0. 705

茄子 SmGSA 基因一级蛋白结构的预测和数据分析的结果显示:Ala(A)284 个,占 27. 2%;Cys(C) 211 个,占 20. 2%;Gly(G)272 个,占 26. 1%;Thr(T)277 个,占 26. 5%。带电残基的总数(Asp+Glu+Arg+Lys)为 0 个,该蛋白质不含有任何 Trp 残基。Tiny G,A,S 出现 556 次,占比 53%。

2.3 SmGSA 蛋白二级结构预测和分析

二级结构是指多肽链沿一个方向排列成具有周期性结构的构象,它是多肽主链中有规则的重复构象。作为蛋白质分子重要的组成部件,二级结构的预测对于蛋白质三维结构的预测具有十分重要的意义。利

用在线分析软件 PRABI 对 SmGSA 蛋白二级结构进行预测,发现 α -螺旋、延伸链和无规卷曲为该蛋白结构的主要构成元件。其中以 α -螺旋和无规卷曲为主,比例共为 85%左右,无规卷曲占主导(表 2)。

表 2 茄子 SmGSA 蛋白二级结构构成元件比例

名称	数量	比例/%
α -螺旋(Hh)	133	38.33
延伸链(Ee)	52	14.99
无规卷曲(Cc)	162	46.69

2.4 SmGSA 蛋白三级结构预测和分析

获得的蛋白质序列通过利用在线软件 SWISS MODEL 寻找与目标序列匹配的进化相关结构,预测结果共发现了 1 029 个模板,从中挑选出匹配度最高的序列,该序列的全球性模型质量估测(GMQE)得分为 0.98(GMQE 得分越接近 1,质量越高),因此预测出来的三级蛋白结构具有很高的准确性。蛋白 3D 模型如图 3 所示,该蛋白三级结构多为 α -螺旋和无规卷曲,与二级结构结果一致。



图 3 茄子 SmGSA 蛋白的 3D 模型图

2.5 蛋白跨膜分析

利用 TMpred 软件在线预测茄子 SmGSA 基因蛋白跨膜结构,预测结果见图 4,评估分值大于 500 被认为是最有可能的跨膜结构域,因此根据软件预测结果,该蛋白中存在 9 处跨膜区。其中氨基酸内部到外部螺旋发现 4 个,分别为氨基酸 154~170 位置处、180~200 位置处、207~229 位置处、267~292 位置处;氨基酸外面到内部螺旋发现 5 个,分别为氨基酸 30~51 位置处、120~143 位置处、178~200 位置处、214~234 位置处、267~290 位置处。预测跨膜能力最强的螺旋是 120~143 氨基酸位置处。说明该蛋白可能具有跨膜转运信号,可以与生物膜结合并参与到细胞信号传递过程中。

2.6 蛋白质亲疏水性分析

用已获得的茄子 GSA 基因氨基酸序列,利用 Protscale 网站软件,在线分析茄子 SmGSA 基因蛋白的亲疏水性。得到线性模型图(图 5),横坐标表示 347 个氨基酸,纵坐标正值为疏水性,负值为亲水性。结果表明,茄子 SmGSA 蛋白为亲水性蛋白质,该结果与一级结构预测结果相一致。

3 结论与讨论

自从第一代 Sanger 测序技术诞生以来,DNA 测序技术经历了第二代高通量测序技术^[8]、第三代单分子测序平台^[9],以及以纳米孔测序为代表的第四代测序技术。第四代测序对生命科学领域具有里程碑意义,引领了科学研究模式的革新以及研究思维的转变^[10]。随着电子信息技术的不断发展,相关研究不断深入,在不久的将来电子克隆在植物领域的应用将会更加广泛和深入。

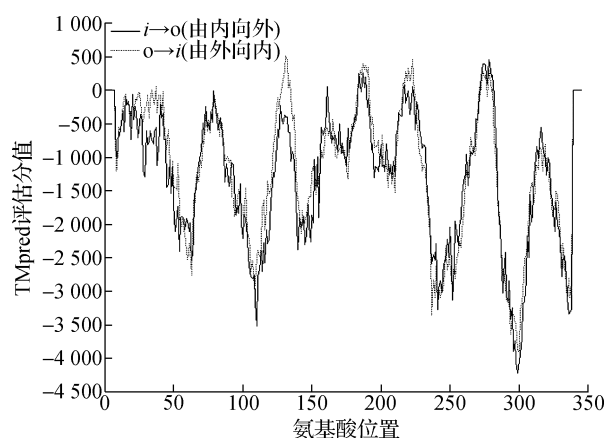


图4 茄子 SmGSA 蛋白的跨膜区分分析

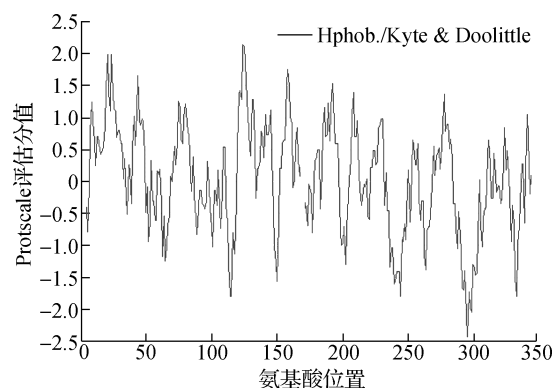


图5 茄子 SmGSA 蛋白的亲疏水性分析

茄子是我国重要的蔬菜作物之一,目前主要研究方向多集中于栽培生理和病虫害防治等方面,对于茄子基因测序和生物信息学分析等相对缺乏。随着基因组学和转录组学测序数据的飞速增长,利用生物信息学分析手段对其各类数据进行分析预测和信息挖掘,对茄子重要基因分离和功能验证具有重要意义。

叶绿素是植物光合作用中的重要色素,叶绿素生物合成大致分为两部分,分别为从 L-谷氨酰-tRNA 到原卟啉 IX 的生物合成和原卟啉 IX 到叶绿素的生物合成,这些生物合成需要酶的参与。其中谷氨酸酯-1-半醛 2,1 氨基变位酶(GSA)基因参与第一部分的合成,其编码的谷氨酸酯-1-半醛 2,1 氨基变位酶作用是催化谷氨酸酯-1-半醛形成 σ -氨基酮戊酸(ALA)。而 ALA 的合成是四吡咯物质合成途径的限速步骤,它决定了叶绿素合成途径的总流量,因此 GSA 基因是该生物合成途径中关键基因之一,分离克隆 GSA 基因对进一步深入研究具有重要作用。

本研究获得了茄子 SmGSA 基因完整开放阅读框并对其翻译的蛋白质结构和理化性质进行了预测和分析,研究结果既可为 SmGSA 基因下一步的功能鉴定及表达奠定基础,又从电子克隆技术和生物信息学分析等方向为茄子遗传育种应用提供了一定的理论依据。

参考文献:

- [1] 崔群香,郝振萍,张爱慧,等. 茄子枝条扦插技术研究及其在茄子育种中的应用[J]. 金陵科技学院学报,2017,33(3):73-76
- [2] 张开畅,王涛,陈艺群,等. 不同颜色茄子果实的若干营养品质分析[J]. 亚热带农业研究,2016,12(1):45-49
- [3] 汪文兴. 茄子的营养价值与设施高产高效栽培技术[J]. 现代园艺,2014(6):38
- [4] 王平荣,张帆涛,高家旭,等. 高等植物叶绿素生物合成的研究进展[J]. 西北植物学报,2009,29(3):629-636
- [5] 姚嘉龙,郭蓓,谢皓. 叶绿素合成关键酶基因表达的半定量 RT-PCR 分析[J]. 基因组学与应用生物学,2015,34(3):593-598
- [6] ZHONG S W, HUI S, XI Y P, et al. Ethylene is crucial for cotyledon greening and seedling survival during de-etiolation [J]. Plant Signaling and Behavior, 2010, 5(6):739-742
- [7] LIU D, KONG D D, FU X K, et al. Influence of exogenous 5-aminolevulinic acid on chlorophyll synthesis and related gene expression in oilseed rape de-etiolated cotyledons under water-deficit stress[J]. Photosynthetica, 2016, 54(3):1-8
- [8] 闫绍鹏,杨瑞华,冷淑娇,等. 高通量测序技术及其在农业科学研究中的应用[J]. 中国农学通报,2012,28(30):171-176
- [9] 张得芳,马秋月,尹佟明,等. 第三代测序技术及其应用[J]. 中国生物工程杂志,2013,33(5):125-131
- [10] 田李,张颖,赵云峰. 新一代测序技术的发展和应[J]. 生物技术通报,2015,31(11):1-8

(责任编辑:湛 江)