

分子植物育种
Molecular Plant Breeding
ISSN 1672-416X, CN 46-1068/S

《分子植物育种》网络首发论文

题目： 蓖麻 GASA 基因家族全基因鉴定及生物信息学分析
作者： 尹冬雪，李国瑞，曲毅鹏，张春玲，由佳庆，汤凯，罗婷，陈永胜
网络首发日期： 2023-03-29
引用格式： 尹冬雪，李国瑞，曲毅鹏，张春玲，由佳庆，汤凯，罗婷，陈永胜. 蓖麻 GASA 基因家族全基因鉴定及生物信息学分析[J/OL]. 分子植物育种, <https://kns.cnki.net/kcms/detail/46.1068.S.20230328.1740.037.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

研究报告

Research Report

蓖麻 GASA 基因家族全基因鉴定及生物信息学分析

尹冬雪¹ 李国瑞^{1,2,3} 曲毅鹏¹ 张春玲¹ 由佳庆¹ 汤凯¹ 罗婷¹ 陈永胜^{1,2,3*}

1 内蒙古民族大学,蓖麻育种国家民委重点实验室,通辽, 028000;2 中国科学院,长春光学精密机械与物理研究所,长春, 130000;

3 内蒙古民族大学,内蒙古自治区高校蓖麻产业工程技术研究中心/内蒙古自治区蓖麻育种重点实验室/内蒙古自治区蓖麻产业协同创新中心/蓖麻产业技术创新内蒙古自治区工程研究中心, 通辽, 028000

*通信作者, chenys_2000@163.com

摘要 GASA (GA-Stimulated in Arabidopsis)家族是植物特有的转录因子家族, 对植物整个生长发育周期有着关键作用。蓖麻 GASA 家族全基因组的鉴定和分析还未见报道。本研究借助生物信息学手段对其基因家族进行成员鉴定、理化性质、蛋白保守结构域及进化分析。结果表明: 基于蓖麻基因组数据共鉴定到 13 个蓖麻 GASA 基因家族成员, 编码 88~245 个氨基酸; 13 个蛋白均为碱性亲水蛋白, 且大多数为不稳定蛋白; 二级结构中无规则卷曲占比最多; 亚细胞定位预测蛋白主要定位在细胞核和高尔基体; 种内进化分析显示 RcGASA1 和 RcGASA8 具有片段重复; 模体分析共鉴定到 4 个 motif。本研究结果将为进一步分析 RcGASA 调控蓖麻生长发育作用机制及生物学功能、蓖麻矮化性状的分子鉴定以及矮化资源创新利用深入研究提供参考依据。

关键词 蓖麻; GASA; 基因家族; 生物信息学分析

Identification and Bioinformatics Analysis of GASA Gene Family in Castor Oil Plant

Yin Dongxue¹ Li Guorui^{1,2,3} Qu Yipeng¹ Zhang Chunling¹ You Jiaqing¹ Tang Kai¹ Luo Ting¹

Chen Yongsheng^{1,2,3*}

基金项目: 本研究由应用光学国家重点实验室 2020 年度开放课题项目(SKLA02020001A19)、中央引导地方科技发展资金项目(2021ZY0016)、内蒙古自然科学基金项目(2021MS03086)、内蒙古自治区高等教育科研项目(NJZY20121)、内蒙古自治区蓖麻产业协同创新中心开放基金项目(MDK2021012)共同资助

1 Key Laboratory of Castor Breeding of the State Ethnic Affairs Commission, Inner Mongolia Minzu University, Tongliao, 028000; 2 Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, 130000; 3 Inner Mongolia Industrial Engineering Research Center of Universities for Castor / Inner Mongolia Key Laboratory of Castor Breeding / Inner Mongolia Collaborative Innovation Center for Castor Industry / Inner Mongolia Engineering Research Center of Industrial Technology Innovation of Castor, Inner Mongolia Minzu University, Tongliao, 028000

*Corresponding author, chenys_2000@163.com

Abstract The GA-stimulated in Arabidopsis gene family is a specific transcription factor in the plant that plays a key role in the whole growth and development cycle. While the whole genome analysis of the GASA gene family has not been reported in castor. In this study, bioinformatics was used to analyze the family members' member prediction, protein structure and evolutionary analysis. A total of 13 castor GASA genes were identified based on the castor genome, they encode 88~245 amino acids; These proteins are all basic proteins and hydrophilic proteins, and most of them are unstable proteins; Random curls account for the largest proportion of secondary structures; Subcellular localization prediction proteins are predominantly localized in the nucleus and the Golgi apparatus; Intraspecific evolutionary analysis showed fragment duplication in RcGASA1 and RcGASA8; A total of 4 motifs were identified in the model analysis. The results of this study will provide a reference for further analysis of the mechanism and biological function of *RcGASA* in regulating the growth and development of castor beans, the molecular identification of dwarf traits of castor beans, and the in-depth research on the innovative utilization of dwarf resources.

Keywords Castor; GASA; Gene family; Bioinformatics analysis

GASA(Gibberellic Acid-Stimulated in Arabidopsis)基因家族是一类受赤霉素(GA)调节的 DELLA 下游的转录因子, GASA 基因家族在植物中分布广泛, 在植物生长发育和激素信号转导过程中发挥极其重要的作用(Wang et al., 2009; Haruta et al., 2014)。GASA 蛋白通常在植物幼嫩和生长旺盛的组织和器官中高表达(Ben-Nissan et al., 2004), 参与种子萌发、花果发育、侧根形成、激素信号转导和植物非生物胁迫。

首个 GASA 基因家族成员是在番茄(*Lycopersicon esculentum*)的赤霉素缺失突变体 *gib1* 植株中被分离鉴定的 *GAST1*(GA-stimulated transcript 1) (Shi et al., 1992), 之后陆续在多种植物中完成了 GASA 全基因家族的分离和鉴定: 拟南芥(*Arabidopsis thaliana*)鉴定到 15 个成员(张盛春和王小菁, 科学通报, (22): 2760-2767)、小麦(*Triticum aestivum*)35 个(吕亮杰等, 2018)、水稻(*Oryza sativa*)9 个(Furukawa et al., 2006)、狗尾草(*Setaria*

viridis)12 个(冯冠楠等, 2022)、黑杨(*Populus deltoides*×*Populus nigra*)18 个(赵腾等, 2012)、葡萄(*Vitis vinifera* L.)14 个(Ahmad et al., 2020)、非洲菊(*Gerbera hybrida*)5 个(刘小飞等, 2019)、菜豆(*Phaseolus vulgaris* L.)23 个(Büyüket al., 2021), 这些研究多围绕家族成员鉴定、基因结构、系统进化和表达分析展开。结果都表明 GASA 蛋白基因结构相对稳定, 可大致分为 3 部分: N 端一般由 18~29 个氨基酸残基信号肽序列组成; 信号肽后面紧邻 7~31 个极性氨基酸残基组成的亲水区域, 不同的 GASA 蛋白在这段区域内的序列变化较大; C 端则是由 60 个左右氨基酸组成的保守结构域, 包含 12 个完全保守的半胱氨酸残基(GASA 结构域), 一般 1 个 GASA 基因只拥有 1 个 GASA 结构域, 该结构域的关键氨基酸发生缺失或改变会导致 GASA 无法发挥基因功能(Aubert et al., 1998; 钟春梅和王小菁, 2016)。

蓖麻为大戟科蓖麻属一年生或多年生草本植物, 是世界十大油料作物之一, 有“绿色石油之称”, 可在航空航天、石油、化工、医药等众多行业中广泛应用(赵秀平等, 2020)。我国蓖麻株高在 150~300 cm 之间, 过高的株高使得无法利用机械进行一次性大面积的收割, 而依靠人工收割就导致种植蓖麻的成本大大提高(李敬忠等, 2018, 农业科技通讯, (10): 198-200), 蓖麻的年产量也随之逐年减少, 进行蓖麻矮化工作是解决当前蓖麻产量困境的有效方式, 蓖麻的矮化还可增强蓖麻抗性, 如: 抗倒伏、抗旱等(孙华军等, 2016)。在风兰(2020)的研究中已经证实蓖麻 GASA 基因家族中的 *RcGASA9* 在蓖麻株高发育中具有重要的调控作用, 目前对蓖麻 GASA 基因功能研究的报道较少, 对蓖麻 GASA 全基因家族进行生物信息学分析的研究尚未见报道。对 *RcGASA* 家族进行鉴定和生物信息学分析有助于为 *RcGASA* 调控蓖麻生长发育作用机制及生物学功能、蓖麻矮化性状的分子鉴定以及矮化资源创新利用深入研究提供参考依据。

1 结果与分析

1.1 蓖麻 GASA 基因家族成员的鉴定

从在线软件 Pfam 下载 GASA 结构域(PF02704)的隐马尔科夫模型, 运用 Hmmer 程序在蓖麻全基因组数据中进行 BLAST 比对, 并运用在线软件 SMART 验证检索蛋白的保守结构域, 将 HMMER 和 BLAST 比对结果进行冗余序列的删除, 最终鉴定到 13 个 *RcGASA* 家族成员, 按照染色体顺序依次命名为 *RcGASA1*~*RcGASA13*, 借助 Primer Premier 5.0 软件设计基因 CDs 编码序列(Coding sequence)的 PCR 引物, 并通过 NCBI-blast 验证引物特异性, 引物序列见表 1。

表 1 RcGASA 基因 PCR 引物

Table 1 PCR reactions primers of RcGASA gene

引物名称	正向序列(5'-3')	反向序列(5'-3')
Primername	Forward number (5'-3')	Reverse primer (5'-3')
<i>RcGASA1</i>	GGTTAAAGGAGCTAACAGAAGGC	GCACTTGGTCTTGTTCATGAG
<i>RcGASA2</i>	CCAGCAAATATTGATGGAGGGAG	CAAGGACATACATGCTTGTGCC
<i>RcGASA3</i>	TCCTCTTAATGCATGAGCTCCAG	AGACGTCTCGATTGCCAGAAGTA
<i>RcGASA4</i>	GGTGTGCTTTGTTCTCACTTCCT	TCATGTCTCTGTAACAAGGGCAC
<i>RcGASA5</i>	GGCCATCTTCAAGACTCTACTGT	GTGGAACACAGTCACATCGGACA
<i>RcGASA6</i>	ATCTCCAAGCCATTGATTGCCTC	ATTCATGGTTACCGGAAGTTCCC
<i>RcGASA7</i>	TGATGATCATCCAAGTGGCGTGA	TAGCAGGGACAGGCATCATAGTT
<i>RcGASA8</i>	TGGTGTAGAGGTGGAAACAGGA	TGAGTGGTCATCTCAGTGTAGCA
<i>RcGASA9</i>	CAGCTTGGGAGAAGTCTCATCAT	TGCCATTGATGGGACAAGGTAGT
<i>RcGASA10</i>	CACACCAGTAATCAAGCCACCAA	AATCTGCTCTTGACCTGACTGGT
<i>RcGASA11</i>	CCCTTCTCTGTTGGTCACTTTC	CAAGGACAAACTTGCTTGTACC
<i>RcGASA12</i>	CTTCTTCTCCTCCTGTTTGCCA	TAGCATGGACTCTTCCTTGTG
<i>RcGASA13</i>	GATGATGATGAGGAAGGCAAGCT	GAAGGTACACACAAGCACTTAGC

1.2 蓖麻 GASA 蛋白理化性质的分析

借助 ExPASy 对 GASA 基因家族蛋白进行分子量、等电点和氨基酸等信息的预测(表 2)。RcGASA 蛋白的氨基酸序列的长度在 88~245 aa 之间, 分子量在 9.72~27.85 kD 之间, 长度和分子量最大和最小分别是 RcGASA9 和 RcGASA4; RcGASA 蛋白的等电点在 7.97~9.89 之间, 13 个基因家族成员等电点全部大于 7, 说明 RcGASA 蛋白是碱性蛋白; RcGASA 蛋白的不稳定指数在 26.72~76.98 之间, RcGASA4、RcGASA6、RcGASA8 和 RcGASA11 的不稳定指数小于 40, 说明这 4 个 RcGASA 蛋白是稳定蛋白, 其它 9 个 RcGASA 蛋白的等电点均大于 40, 说明其它 9 个 RcGASA 蛋白是不稳定蛋白; RcGASA 蛋白的

脂肪指数在 42.27~92.60 之间；RcGASA 蛋白的总平均亲水性在-0.556~-0.01 之间，说明 RcGASA 蛋白均是亲水蛋白。

表 2 RcGASA 蛋白的基本信息

Table 2 Basic information of RcGASA protein

基因名称	基因号	氨基酸数	分子量(kD)	等电点	不稳定指数	脂肪指数	总平均亲水系数
Gene name	Gene ID	No. of AA	MW (kD)	pI	Instability index	Aliphatic index	GRAVY
<i>RcGASA1</i>	Rc01T000758.1	105	11.59	9.03	44.86	67.71	-0.069
<i>RcGASA2</i>	Rc02T003752.1	113	12.67	9.45	66.20	79.38	-0.187
<i>RcGASA3</i>	Rc02T003753.1	90	10.02	8.73	59.89	76.00	-0.097
<i>RcGASA4</i>	Rc02T004517.1	88	9.72	8.91	26.72	53.07	-0.184
<i>RcGASA5</i>	Rc02T004871.1	100	10.86	8.49	57.40	92.60	0.324
<i>RcGASA6</i>	Rc02T004872.1	101	10.88	9.16	38.41	78.42	-0.01
<i>RcGASA7</i>	Rc02T004873.1	106	11.01	8.91	57.18	78.40	0.103
<i>RcGASA8</i>	Rc04T007999.1	93	10.10	9.00	35.84	64.95	-0.17
<i>RcGASA9</i>	Rc05T009567.5	245	27.85	9.06	76.98	65.35	-0.556
<i>RcGASA10</i>	Rc05T009569.1	234	24.76	9.89	69.86	67.44	-0.295
<i>RcGASA11</i>	Rc05T012305.1	95	10.63	9.20	34.73	64.63	-0.051
<i>RcGASA12</i>	Rc06T013577.1	100	11.14	7.97	40.31	51.80	-0.072
<i>RcGASA13</i>	Rc10T022796.1	97	11.01	9.28	51.08	42.27	-0.255

1.3 RcGASA 蛋白结构的分析

经过在线软件 Prabi 的分析, 显示 RcGASA 蛋白的二级结构均含有 α -螺旋、延伸链、 β -转角和无规则卷曲这四种结构元件, 只有 RcGASA5 特殊, 不具有延伸链, 仅有其他 3 种结构。四种结构元件中整体含量最高的是无规则卷曲, 含量在 50.00%~76.50%; 其次是 α -螺旋, 含量在 12.38%~43.33%; 再次是延伸链, 排除 RcGASA5 其他基因的含量在 2.83%~15.10%; 最低的是 β -转角, 含量在 2.56%~6.45%。RcGASA 蛋白亚细胞定位的预测结果表明, RcGASA1、RcGASA4、RcGASA6、RcGASA11、RcGASA12 和 RcGASA13 蛋白定位于高尔基体, RcGASA5、RcGASA8、RcGASA9 和 RcGASA10 蛋白定位于细胞核, 其它 3 个基因定位于不止 1 个位置(表 3), 为进一步探究蓖麻 GASA 蛋白的结构, 使用 SWISS-MODEL 软件在蛋白二级结构的基础上将 GASA 蛋白进一步修饰为三级结构, 对具有代表性的蓖麻 RcGASA 蛋白的同源建模进行展示(图 1)。

表 3 RcGASA 蛋白二级结构和亚细胞定位

Table 3 Secondary structure and subcellular location of RcGASA protein

基因名称	α -螺旋(%)	延伸链(%)	β -转角(%)	无规则卷曲(%)	亚细胞定位
Gene name	α -helix (%)	Extended strand (%)	β -turn (%)	Random coil (%)	Subcellular location
<i>RcGASA1</i>	12.38	8.57	3.81	75.24	高尔基体 Golgi apparatus
<i>RcGASA2</i>	31.86	9.73	3.54	54.87	高尔基体; 细胞核 Golgi apparatus; Nucleus.
<i>RcGASA3</i>	43.33	3.33	3.33	50.00	细胞膜; 高尔基体; 细胞核 Cell membrane; Golgi apparatus; Nucleus
<i>RcGASA4</i>	31.82	7.95	2.27	57.95	高尔基体 Golgi apparatus

<i>RcGASA5</i>	30.00	0.00	4.00	66.00	细胞核 Nucleus
<i>RcGASA6</i>	40.59	3.96	2.97	52.48	高尔基体 Golgi apparatus
<i>RcGASA7</i>	23.58	2.83	3.77	69.81	细胞膜; 细胞壁; 高尔基体; 细胞核 Cell membrane; Cell wall; Golgi apparatus; Nucleus
<i>RcGASA8</i>	23.66	9.68	6.45	60.22	细胞核 Nucleus
<i>RcGASA9</i>	15.51	15.10	3.27	66.12	细胞核 Nucleus
<i>RcGASA10</i>	14.10	6.84	2.56	76.50	细胞核 Nucleus
<i>RcGASA11</i>	34.74	6.32	5.26	53.68	高尔基体 Golgi apparatus
<i>RcGASA12</i>	38.00	8.00	4.00	50.00	高尔基体 Golgi apparatus
<i>RcGASA13</i>	28.87	4.12	5.15	61.86	高尔基体 Golgi apparatus



图 1 蓖麻 RcGASA 基因家族的蛋白三级结构

Figure 1 Protein tertiary structure of the castor RcGASA gene family

1.4 RcGASA 蛋白质保守基序分析

通过 MEME 和 TBtools 分析 RcGASA 蛋白的序列进行分析后获得 4 个 motif, 根据保守性的强弱依次命名为 motif 1~ motif 6, 构建了 RcGASA 蛋白的系统进化树和保守基序图(图 2)。基序, 亦称模序、模体, 是指 DNA、蛋白质等生物大分子中的保守序列, 这部分结构域是介于二级和三级结构之间的另一种结构层次, 也称超二级结构(孙利霞等, 2015)。从 RcGASA 蛋白的保守基序可知, 13 个 RcGASA 蛋白均含有 motif 1 和 motif 3, 说明这两个基序在 RcGASA 蛋白中高度保守。*RcGASA2*、*RcGASA5*、*RcGASA7*、*RcGASA8*、*RcGASA13* 这 5 个基因具有全部 4 种 motif, *RcGASA1*、*RcGASA3*、*RcGASA4*、*RcGASA10*、*RcGASA11*、*RcGASA12* 这 6 个基因具有 motif1、motif2、motif3 这三种基序, *RcGASA9* 最为特殊, 仅具有 motif1 和 motif3。将保守基序与进化树结合起来可分析, 不同进化树的分支所含的 motif 也有所区别, 如 motif 20 仅在 *RcGASA20* 与 *RcGASA21* 这一分支中, 说明 RcGASA 蛋白具有一定的保守性。

为探究 *RcGASA* 基因家族的基因结构, 基于蓖麻注释文件, 使用编写的脚本对其基因结构进行分析, 运用在线网页 GSDS 对基因结构进行可视化(图 2)。结果显示, 13 个基因均为断裂基因, 即外显子被内含子隔开, 且每个基因在结构上均存在一定差异。结合 *RcGASA* 基因家族蛋白进化树分析可知, 同一类之间的结构较为相似, 其同源性越高, 反之, 其同源性越低。在 *RcGASA* 基因家族的编码区中, 包含 1~4 个内含子, 其中, *RcGASA4* 有 1 个内含子, *RcGASA2* 和 *RcGASA10* 均有 3 个内含子, *RcGASA9* 有 4 个内含子, 其余 9 个 *RcGASA* 基因均包含 2 个内含子, 占总 *RcGASA* 基因个数的 69.2%。*RcGASA* 基因编码的蛋白质在蓖麻在进化过程中具有较高的保守性, 编码区没有因 *GASA* 基因结构发生变化产生较大变异。因此, 推测 *RcGASA* 基因结构变化的主要原因在于内含子上, 而不在外显子上。

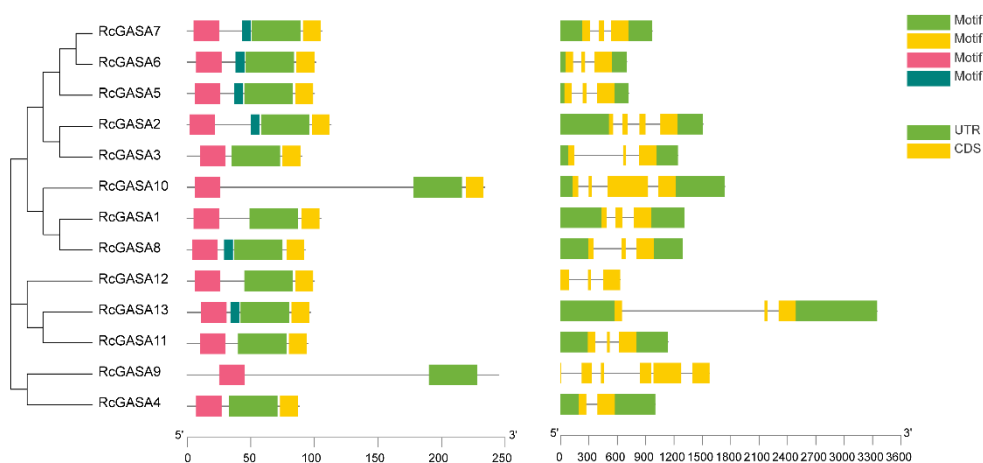


图 2 蓖麻 RcGASA 基因的系统发育树, 蛋白质保守基序和基因结构

Figure 2 Intraspecific phylogenetic trees, protein-conserved motifs and gene structures of RcGASA genes in castor

1.5 RcGASA 复制分析

基因在染色体上的排列分布与染色体在表达过程中的参与程度及在植物生长发育中的重要性密切相关(乐超银等, 2003)。为探究蓖麻 GASA 基因家族的复制模式与进化机制, 进行了染色体定位分析, 以往蓖麻多在 Scaffold 水平进行分析, 难以有效地探明其关系, 得益于测序技术的进步, 二代测序结合三代测序已将蓖麻基因组组装至染色体水平, 13 个蓖麻 *RcGASA* 基因分布在 6 条染色体上, 在 RcChr3、RcChr7、RcChr8、RcChr9 上没有分布。此外, 除 RcChr2 和 RcChr5 染色体上有多个成员分布, 其余染色体上只有 1 个基因。

为了分析蓖麻 *RcGASA* 家族基因与祖先材料的同源进化关系, 采用生物信息学方法具有同源关系的基因进行相关的连线说明(图 3)。MCScanX 结果显示, 在蓖麻 *RcGASA* 基因家族的 13 个成员中, 除 Chr1 染色体上的 *RcGASA1* 与 Chr4 染色体上的 *RcGASA8* 有对应性关系, 其余 11 个基因都与其他染色体上的基因无同源性对应关系, *RcGASA1* 和 *RcGASA8* 可能来源于全基因组复制或片段复制。

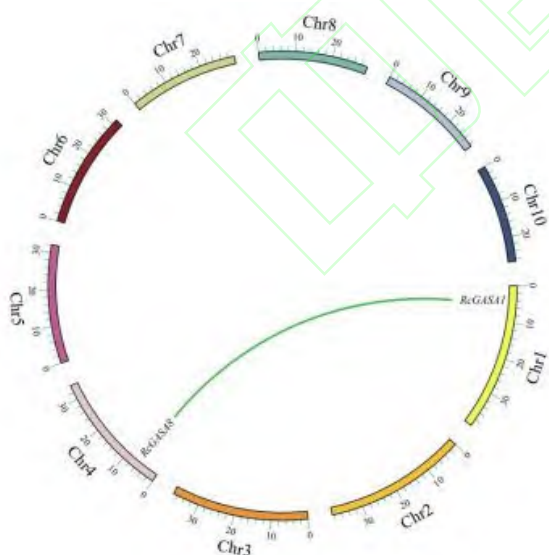
图 3 蓖麻 *RcGASA* 基因同源进化分析

Figure 3 Homologous evolution analysis of *RcGASA* genes in castor bean

1.6 GASA 基因家族进化分析

为了研究蓖麻 GASA 基因家族的进化关系，采用邻接法(NJ)对蓖麻(RcGASA, 13 个)、拟南芥 (AtGASA,15 个)和狗尾草(SvGASA, 12 个)的蛋白质序列进行比对分析，构建种间进化树(图 4)。

进化分析结果显示，这三个物种的所有 GASA 基因可分为 3 个亚族，命名为：G1、G2 和 G3。3 个物种的家族基因都在 3 个亚族中有所分布，其中：13 个 RcGASA 家族基因中有 9 个基因属于 G3 亚族，15 个 AtGASA 家族基因有 6 个分布在 G1 亚族。

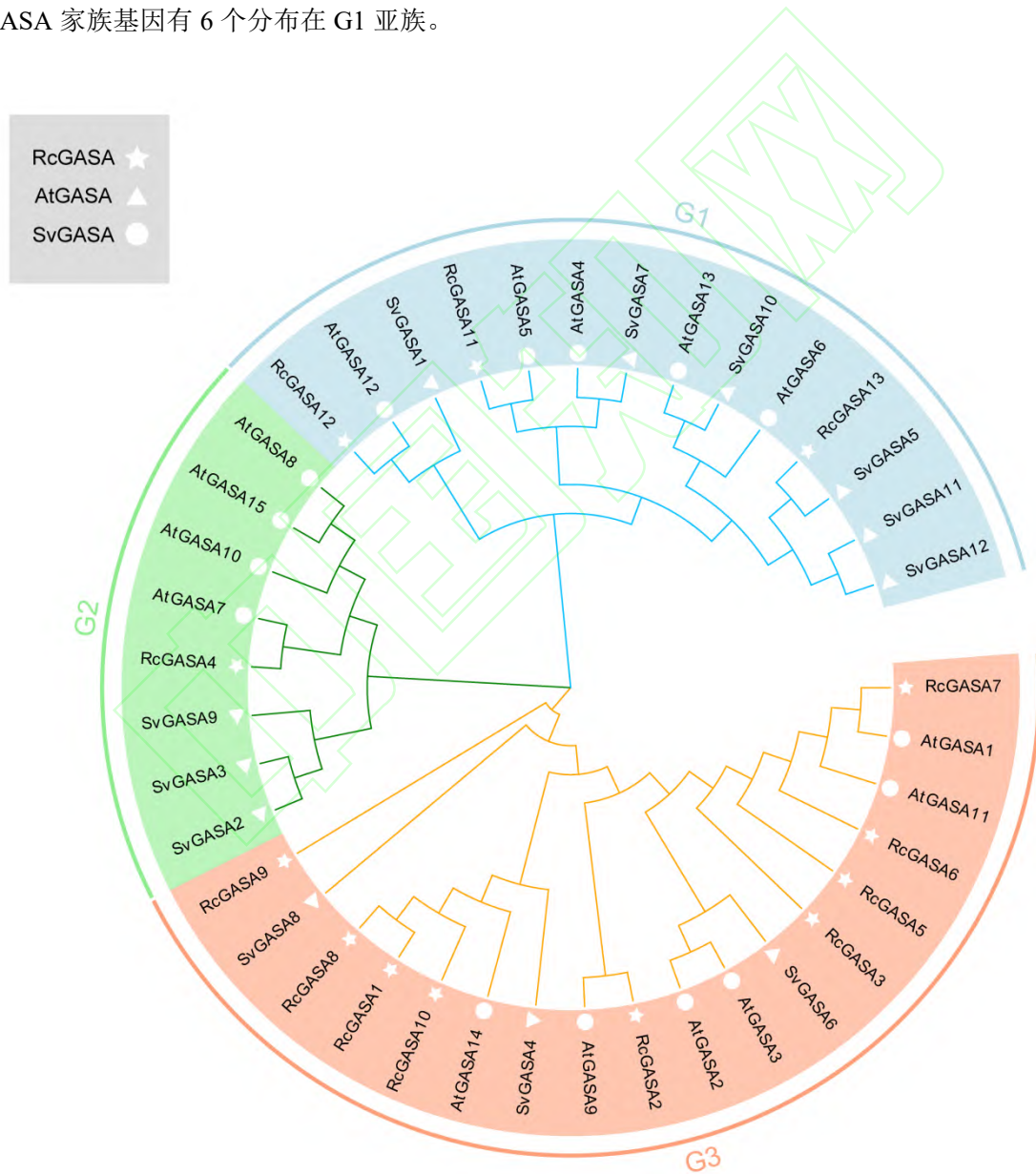


图 4 蓖麻 *RcGASA* 种间系统发育树

Figure 4 Phylogenetic relationships of *RcGASA* genes in castor bean

2 讨论

GASA 是一类富含半胱氨酸的小分子多肽, 在植物生长发育、激素响应及逆境胁迫等多个生理生化途径中发挥着重要作用(Marshall et al., 2011), GASA 在蓖麻中又可能与调控株高有关, 且目前关于 *RcGASA* 基因家族成员的研究又较少。本研究基于蓖麻基因组数据库, 首次鉴定到蓖麻 GASA 家族的 13 个基因, 编码长度为 88~245 个氨基酸, 这与小麦 35 个 *TaGASA* 基因编码 78~264 个氨基酸的结果很相近(吕亮杰等, 2018); 蛋白质的二级结构是判断蛋白质稳定性的重要因素(穆昭艳等, 2012), 其中, α -螺旋、 β -转角是蛋白质的有序结构, 具有稳定性, 而无规则卷曲为蛋白质的无序结构, 为不稳定性(时小东等, 2019), 蓖麻 GASA 基因家族中, 无规则卷曲占比最高, 全部在 50%以上, 无规则卷曲是无序结构, 所以 *RcGASA* 全基因家族均为不稳定蛋白, 这与理化性质中预测的蛋白稳定系数结果一致; 在保守基序和基因结构的分析中, 4 个 motif 的位置和数量除 *RcGASA9* 比较特殊外, 其余都具有一定的相似性, 且进化树的亲缘关系越近, 基因结构越相似。蓖麻 *RcGASA* 家族共 13 个基因, 该成员数目与拟南芥中鉴定到的 15 个家族成员数目相近, 又远少于小麦中的 35 个 GASA 基因, 结合共线性分析可以看出, 蓖麻中无串联重复事件, 仅有 1 次片段复制。

随着越来越多的 GASA 基因家族在不同物种中被鉴定到, 部分基因家族成员的功能也被揭示。在胡杨 (*Populus euphratica*)中共鉴定到 19 个 *PeGASA* 家族成员, 通过 RT-qPCR 研究发现, 该家族基因主要在幼嫩组织和器官中表达, 推测其可能参与叶片的生长发育(Han et al., 2021); 对 5 种棉花品种(*Gossypium herbaceum*, *G.arboreum*, *G.raimondii*, *G.barbadense*, and *G.hirsutum*)进行全基因组分析, 分别鉴定到 19、17、25、33 和 38 个家族成员, 通过转录组和 RT-qPCR 研究分析, 揭示了非生物胁迫下棉花 *GhGASA* 基因可能参与纤维发育和非生物胁迫响应(Qiao et al., 2021); 在花生(*Arachis hypogaea* L.)全基因组中共鉴定到 40 个 *AhGASA* 基因, 转录组学和 RT-qPCR 研究表明 *AhGASA* 基因可能参与控制花生荚的大小(Wu et al., 2022)。在可可树(*Theobroma cocoa* L.)全基因组鉴定到 17 个 *TcGASA* 基因, 在转录组研究中发现 *TcGASAs* 对真菌有响应, 这为开发抗疫霉属黑杆病的可可树品种提供了靶基因(Abdullah et al., 2021)。

综上所述, 本研究通过生物信息学的手段, 从全基因组中鉴定到 13 个蓖麻 *RcGASA* 基因家族成员, 并对其进行理化性质、系统进化、保守结构域、基因结构和共线性进行分析, 这为解析 *RcGASA* 基因家族功能奠定了基础, 并为遗传育种工作提供了参考。

3 材料与amp;方法

3.1 *RcGASA* 基因家族成员的鉴定

蓖麻全基因序列下载自 Oil Plant Database(<http://oilplants.iflora.cn>), 拟南芥 *GASA* 基因家族的基因与蛋白质序列获取于 TAIR 数据库(Araport11, <https://www.arabidopsis.org>), 从 Pfam 数据库(<http://pfam.xfam.org>) 中搜索 *GASA* 基因家族蛋白的保守结构域(PF02704)的隐马尔科夫模型, 下载对应的 hmm 文件, 通过 HMMER 在线软件将蓖麻基因组序列与 hmm 文件进行 BLAST 比对($E\text{-value} \leq 1e^{-5}$), 将 HMMER 和 BLAST 比对结果进行冗余序列的删除, 最终确定家庭成员。

3.2 生物学信息学分析

在线软件 ExPASy (<https://web.expasy.org/protparam>)预测 *RcGASA* 蛋白的理化性质; 在线软件 Prabi (<http://www.prabi.fr>) 分析 *RcGASA* 蛋白的二级结构; 在线软件 SWISS-MODEL (<https://swissmodel.expasy.org>) 预测 *RcGASA* 蛋白的三级结构; 在线软件 Cell-Ploc (<http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2>)预测 *RcGASA* 蛋白的亚细胞定位。

3.3 系统进化树、基因结构、保守结构域的预测和分析

系统进化树和保守基序图由软件 clustalW、phyML、MEGA7 和在线软件 MEME (<http://meme-suite.org/tools/meme>) 及 TBtools (www.tbtools.com) 共同绘制; 利用在线工具 GSDS (<http://gsds.cbi.pku.edu.cn>)对 *RcGASA* 家族成员进行基因结构分析与共线性分析。

作者贡献

尹冬雪负责本研究的实验设计和论文写作; 曲毅鹏和张春玲负责基因家族的鉴定; 汤凯、罗婷、由佳庆负责基因家族系统进化和结构分析; 李国瑞和陈永胜是项目的负责人, 指导实验设计和论文修改。全体作者都阅读并同意最终的文本。

致谢

本研究由应用光学国家重点实验室 2020 年度开放课题项目(SKLA02020001A19)、中央引导地方科技

发展资金项目(2021ZY0016)、内蒙古自然科学基金项目(2021MS03086)、内蒙古自治区高等教育科研项目(NJZY20121)和内蒙古自治区蓖麻产业协同创新中心开放基金项目(MDK2021012)共同资助。

参考文献

- Abdullah, Faraji S., Mehmood F., Malik H.M.T., Ahmed I., Heidari P., and Poczaï P., 2021, The GASA gene family in *Theobroma cacao*: genome wide identification and expression analyses, *Agronomy*, 11(7): 1425.
- Ahmad B., Yao J., Zhang S.L., Li X.M., Zhang X.M., Yadav V., and Wang X.P., 2020, Genome-wide characterization and expression profiling of GASA genes during different stages of seed development in grapevine (*Vitis vinifera* L.) predict their involvement in seed development, *Int. J. Mol. Sci.*, 21(3): 1088.
- Aubert D., Chevillard M., Dorne A.M., Arlaud G., and Herzog M., 1998, Expression patterns of GASA genes in *Arabidopsis thaliana*: the *GASA4* gene is up-regulated by gibberellins in meristematic regions, *Plant Mol. Biol.*, 36(6): 871-883.
- Ben-Nissan G., Lee J.Y., Borohov A., and Weiss D., 2004, GIP, a *Petunia hybrida* GA-induced cysteine-rich protein: a possible role in shoot elongation and transition to flowering, *Plant J.*, 37(2): 229-238.
- Büyüktürk L., Okay A., Górska M., İlhan E., and Aras S., 2021, Identification and characterization of the Pvul-GASA gene family in the *Phaseolus vulgaris* and expression patterns under salt stress, *Turkish J. Bot.*, 45(7): 655-670.
- Feng G.N., An C., Ding Y.J., Zhang X.X., and Zhuang L.L., 2022, Characterization and analysis of GASA gene family and response to abiotic stress in *Setaria viridis*, *Caodi Xuebao (Acta Agrestia Sinica)*, 30(6): 1379-1387. (冯冠楠, 安聪, 丁懿嘉, 张夏香, 庄黎丽, 2022, 狗尾草 GASA 基因家族鉴定及其在不同逆境下表达模式分析, *草地学报*, 30(6): 1379-1387.)
- Feng L., 2020, Transcriptome analysis of high-stalk and dwarf castor and functional studies of *RcBKII* and *RcGASA9* genes, Thesis for M.S., Inner Mongolia Minzu University, Supervisor: Chen Y.S., pp.79-81. (风兰, 2020, 高、矮秆蓖麻转录组分析及 *RcBKII* 和 *RcGASA9* 基因的功能研究, 硕士学位论文, 内蒙古民族大学, 导师: 陈永胜, pp.79-81.)
- Furukawa T., Sakaguchi N., and Shimada H., 2006, Two *OsGASR* genes, rice GAST homologue genes that are abundant in proliferating tissues, show different expression patterns in developing panicles, *Genes Genet. Syst.*, 81(3): 171-180.
- Han S., Jiao Z.Y., Niu M.X., Yu X., Huang M., Liu C., Wang H.L., Zhou Y.Y., Mao W., Wang X.F., Yin W.L., and Xia X.L., 2021, Genome-wide comprehensive analysis of the GASA gene family in *Populus*, *Int. J. Mol. Sci.*, 22(22): 12336.
- Haruta M., Sabat G., Stecker K., Minkoff B.B., and Sussman M.R., 2014, A peptide hormone and its receptor protein kinase regulate plant cell expansion, *Science*, 343(6169): 408-411.
- Le C.Y., Sun M., Chen S.W., and Yu Z.N., 2003, Construction of insecticidal recombinant *Bacillus thuringiensis* using an integrative vector, *Yichuan Xuebao (Journal of Genetics and Genomics)*, (8): 737-742. (乐超银, 孙明, 陈守文, 喻子牛, 2003, 利用整合载体构建苏云金芽胞杆菌杀虫工程菌, *遗传学报*, (8): 737-742.)
- Liu X.F., Wang X.J., and Peng J.Z., 2019, Cloning and expression analysis of GASA family genes in petals of *Gerbera hybrida*, *Yuanyi Xuebao (Acta Horticulturae Sinica)*, 46(11): 2257-2264. (刘小飞, 王小菁, 彭建宗, 2019, 非洲菊花瓣 GASA 家族基因的克隆及表达分析, *园艺学报*, 46(11): 2257-2264.)
- Lü L.J., Chen X.Y., Zhang Y.L., Liu Q., Wang L.M., Ma L., and Li H., 2018, Bioinformatics analysis of wheat GASA gene family,

- Zuowu Zazhi (Crops), (6): 58-67. (吕亮杰, 陈希勇, 张业伦, 刘茜, 王莉梅, 马乐, 李辉, 2018, 小麦 GASA 基因家族生物信息学分析, 作物杂志, (6): 58-67.)
- Marshall E., Costa L.M., and Gutierrez-Marcos J., 2011, Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development, *J. Exp. Bot.*, 62(5): 1677-1686.
- Mu Z.Y., Wang L.P., Zhao Y., Wang N., Li A.X., and Wang X., 2012, Study on frequency of isomannanase locus mutation in *Bacillus subtilis* induced by UV, *Huabei Nongxuebao (Acta Agriculturae Boreali-Sinica)*, 27(4): 36-41. (穆昭艳, 汪立平, 赵勇, 王娜, 李安雪, 汪欣, 2012, 枯草芽孢杆菌紫外诱变异甘露聚糖酶基因突变的研究, *华北农学报*, 27(4): 36-41.)
- Qiao K.K., Ma C.K., Lü J.Y., Zhang C.J., Ma Q.F., and Fan S.L., 2021, Identification, characterization, and expression profiles of the GASA genes in cotton, *Journal of Cotton Research*, 4(1):16.
- Shi L., Gast R.T., Gopalraj M., and Olszewski N.E., 1992, Characterization of a shoot-specific, GA₃- and ABA-regulated gene from tomato, *Plant J.*, 2(2): 153-159.
- Shi X.D., Wu Q., Xiang D.B., Wan Y., and Zhao G., 2019, Bioinformatics analysis of GIS gene family in *Fagopyrum tataricum*, *Xinan Nongye Xuebao (Southwest China Journal of Agricultural Sciences)*, 32(8): 1717-1722. (时小东, 吴琪, 向达兵, 万燕, 赵钢, 2019, 苦荞 GIS 基因家族生物信息学分析, *西南农业学报*, 32(8): 1717-1722.)
- Sun H.J., Li G.R., Li Y., Cong A.Q., Li W., Qi M., and Chen Y.S., 2016, Cloning and analysis of *RcDof* gene associated with castor dwarfing, *Huabei Nongxuebao (Acta Agriculturae Boreali-Sinica)*, 31(4): 63-67. (孙华军, 李国瑞, 李跃, 丛安琪, 李威, 齐蒙, 陈永胜, 2016, 蓖麻矮化相关基因 *RcDof* 的克隆及分析, *华北农学报*, 31(4): 63-67.)
- Sun L.X., Hu X.Z., Li S.B., and Li K., 2015, Recognition of complex super secondary structure $\beta\alpha\beta$ motifs in proteins based on combined sequence feature, *Neimenggu Gongye Daxue Xuebao (Journal of Inner Mongolia University of Technology: Natural Science Edition)*, 34(3): 177-183. (孙利霞, 胡秀珍, 李少波, 李昆, 2015, 基于组合的序列特征识别蛋白质复杂超二级结构 $\beta\alpha\beta$ 模体, *内蒙古工业大学学报: 自然科学版*, 34(3): 177-183.)
- Wang L., Wang Z., Xu Y.Y., Joo S.H., Kim S.K., Xue Z., Xu Z.H., Wang Z.Y., and Chong K., 2009, OsGSR1 is involved in crosstalk between gibberellins and brassinosteroids in rice, *Plant J.*, 57(3): 498-510.
- Wu Y., Sun Z.Q., Qi F.Y., Zhao M.B., Dong W.Z., Huang B.Y., Zheng Z., and Zhang X.Y., 2022, Comprehensive analysis of GASA family members in the peanut genome: identification, characterization, and their expressions in response to pod development, *Agronomy*, 12(12):3067.
- Zhao T., Xia X.L., and Yin W.L., 2012, Cloning and functional analysis of GASA gene in poplar, *Guangdong Nongye Kexue (Guangdong Agricultural Sciences)*, 39(8): 138-140, 144, 4. (赵腾, 夏新莉, 尹伟伦, 2012, 黑杨 GASA 基因的克隆和功能分析, *广东农业科学*, 39(8): 138-140, 144, 4.)
- Zhao X.P., Li G.R., Feng L., Han W.Y., Wang S., Sun J.X., Yan X.Y., Huang F.L., and Chen Y.S., 2020, Bioinformatics analysis of castor gibberellin receptor protein RcGID1B, *Fenzi Zhiwu Yuzhong (Molecular Plant Breeding)*, 18(8): 2497-2502. (赵秀平, 李国瑞, 风兰, 韩雯毓, 王双, 孙佳欣, 闫星伊, 黄凤兰, 陈永胜, 2020, 蓖麻赤霉素受体蛋白 RcGID1B 的生物信息学分析, *分子植物育种*, 18(8): 2497-2502.)
- Zhong C.M., and Wang X.J., 2016, Progress in cysteine-rich gibberellic acid-stimulated Arabidopsis protein, *Zhiwu Xuebao (Chinese Bulletin of Botany)*, 51(1): 1-8. (钟春梅, 王小菁, 2016, 富含半胱氨酸的 GASA 小分子蛋白研究进展, *植物学报*, 51(1): 1-8.)