分子植物育种(网络版)，2020 年，第 18 卷，第 1 篇，第 1–11 页

研究报告
Reaserch Report

茎瘤芥 GRAS 家族基因鉴定与基因表达分析

蒋龙星　郭佳鑫　孙全　何晓红 *
大数据生物智能重庆市重点实验室, 重庆邮电大学生物信息学院, 重庆, 400065
* 通信作者, hexh@cqupt.edu.cn

摘　要　GRAS 家族蛋白是植物所特有的一类转录因子，在植物基因的表达调控上扮演着重要角色。本研究基于茎瘤芥基因组数据，全基因组鉴定了茎瘤芥 GRAS 基因家族。使用隐马尔可夫模型从茎瘤芥基因组中共鉴定出 102 个 *GRAS* 基因，这些基因不均匀地分布在茎瘤芥的 18 条染色体上并形成了 11 个基因簇，且在茎瘤芥的 A、B 亚基因组中呈不均匀分布。亚细胞定位结果表明，绝大部分 GRAS 蛋白定位于细胞核中，这与转录因子调控基因表达的特点密切相关。利用系统发育分析将其划分为 LISCL、PAT1、DELLA、SCL3、SCR、LS、DLT、SHR、NSP1、SCL32、SCL4/7、NSP2 和 HAM 共 13 个亚家族。蛋白质序列预测结果显示茎瘤芥 GRAS 基因家族亚家族间氨基酸数量、分子量大小、等电点等理化性质存在较大的变异。通过 MCscanX 共线性分析，我们发现茎瘤芥 GRAS 基因家族出现扩张的原因是基因的串联重复和片段重复，在对各类刺激作出适应性响应过程中起着关键作用。结合本实验室茎瘤芥转录组测序结果分析，筛选到 26 个 *GRAS* 基因可能参与茎瘤芥的瘤状茎的形成。这些结果为后期茎瘤芥 *GRAS* 基因功能研究及茎瘤芥育种及产量调控等奠定了一定的基础。

关键词　茎瘤芥, GRAS 基因家族, 系统发育分析, 生物信息学, 转录因子

Genome Identification and Expression Analysis of GRAS Gene Family in *Brassica juncea* var. Tumida L

Jiang Longxing　Guo Jiaxin　Sun Quan　He Xiaohong *
1 Institute of Life Science, Jiyang College of Zhejiang A&F University, Zhuji, 311800, China; 2 Cuixi Academy of Biotechnology, Zhuji, 311800, China
* Corresponding author, hexh@cqupt.edu.cn
DOI: 10.5376/mpb.cn.2020.18.0008

Abstract　GRAS family proteins are plant-specific transcription factors that play important roles in regulating genes expression. In this study, 102 *GRAS* genes were identified from Zha-tsai (*Brassica juncea* var. *tumida Tsen et Lee*, a vegetable for processing the preserved mustard) genome database by hidden Markov model. According to the location information, these genes were unevenly distributed on 18 chromosomes and formed 11 clusters, also unevenly distributed in subgenome A and B. Subcellular localization showed that most GRAS proteins were localized in the cell nucleus, which was closely related to the characteristics that transcription factors regulating gene expression. According to phylogenetic features, the identified *GRAS* genes were divided into 13 subgroups: LISCL, PAT1, DELLA, SCL3, SCR, LS, DLT, SHR, NSP1, SCL32, SCL4/7, NSP2 and HAM. The physicochemical properties prediction revealed that there were some difference in subgroups, such as number of amino acids,

收稿日期：2020 年 3 月 27 日；接受日期：2020 年 4 月 20 日；发表日期：2020 年 4 月 20 日
引用格式: 蒋龙星, 郭佳鑫, 孙全, 何晓红, 2020, 茎瘤芥 GRAS 家族基因鉴定与基因表达分析, 《分子植物育种》(网络版), 18(2): 1-11 (doi: 10.5376/mpb.cn.2020.18.0002) (Jiang L.X., Guo J.X., Sun Q., and He X.H., 2020, Genome identification and expression analysis of GRAS gene family in *Brassica Juncea* var. Tumida L, Fenzi Zhiwu Yuzhong (Molecular Plant Breeding) (Online), 18 (2): 1-11, (doi: 10.5376/mpb.cn.2020.18.0002))

molecular weight, isoelectric point and other physicochemical properties. For the expansion of the *GRAS* genes in *Brassica juncea* var. *tumida Tsen et Lee*, segmental duplication and tandem duplication were main patterns which play the same crucial roles in responding various stress. Combined with the results of transcriptome of the stem in tumorous stem mustard, 26 *GRAS* genes may involved in formation of the stem in tumorous stem mustard. As such, this systematic analysis lays a foundation for further study of the functional characteristics of *GRAS* genes and generates valuable information for improving production of tumorous stem mustard crops.

Keywords *Brassica juncea* var. *tumida Tsen et Lee*, GRAS gene family, Phylogenetic analysis, Bioinformatics, Transcription factor

茎瘤芥(*Brassica juncea* var. *tumida Tsen et Lee*)是一种茎用芥菜，属于芥菜的变种，常称为榨菜。茎瘤芥双子叶植物，十字花科，芸薹属，作为涪陵榨菜的主要原料产物，是重庆地区的优势经济作物。涪陵榨菜因其色、香、味俱佳而畅销国内外(Yang et al., 2016)。

转录因子(transcription factor, TF)是一类与特定的 DNA 序列相结合的蛋白质，从而控制 DNA 的遗传信息转录到信使 RNA 的转录速率(Latchman, 1997)。转录因子的功能是调节基因的表达，以确保生物体整个生命周期中，各基因在正确的时间表达正确的含量(Karin, 1990)。GAI (Peng et al., 1997) (Gibberellic Acid Insensitive)、RGA (Silverstone et al., 1998) (Repressor of GA1-3 mutant)和 SCR (Di Laurenzio et al., 1996) (Scarecrow)是最早被前人发现的三个 *GRAS* 基因，因此以它们的名字命名了这类转录因子家族。

GRAS 基因家族作为植物特异性转录因子之一，在拟南芥中被广泛研究，发现该家族在植物信号转导、生长发育、基因调控以及胁迫相应具有重要的调控作用(Liu and Widmer, 2014)。绝大部分的 GRAS 蛋白都直接定位在细胞核中参与核基因表达调控，仅有少数例外，如 PAT1 蛋白定位于细胞质基质中(Bolle et al., 2000)；在细胞质和细胞核中发现了 SHR 蛋白、SCL 蛋白和 SCL14 蛋白 (Sun et al., 2012; Torres-Galea et al., 2006)。由 GRAS 基因家族编码的蛋白质通常由 400~770 个氨基酸残基组成，其 C- 末端高度保守，而 N- 末端可变(Sun et al., 2011)。此外，仍有少部分 GRAS 蛋白在 C 端含有两个 GRAS 结构域或是一个 GRAS 结构域和一个功能区。

GRAS 蛋白 C 端保守结构域由 5 个模序组成，分别是 LRⅠ(亮氨酸丰富区Ⅰ)、VHⅡD、LRⅡ(亮氨酸丰富区Ⅱ)、PFYRE 和 SAW (Pysh et al., 1999)。VHⅡD 的两侧是两个亮氨酸七肽重复序列(LRⅠ和 LRⅡ)，研究人员也在 bZIP 转录因子中发现过该重复序列，这些重复序列对于研究蛋白质与蛋白质间的相互作用至关重要(Guiltinan and Miller, 1994)。VHⅡD 基序在 GRAS 蛋白中均被发现，被认为是该蛋白家族最保守的结构域，在与 DNA 结合和与蛋白质的结合中起重要作用。

目前研究人员已完成对拟南芥、水稻、杨树、小麦、番茄、甘薯、白菜等物种的 *GRAS* 基因的鉴定及分析工作，其大多研究方法是通过多序列比对构建 GRAS 基因家族的系统发育树，通过系统发育树拓扑结构将 GRAS 蛋白分为 8 至 14 个亚类不等，但茎瘤芥 GRAS 基因家族系统发育分析及大部分成员的功能目前还没有研究报道。

本研究通过生物信息学方法在全基因组水平鉴定了茎瘤芥的 GRAS 家族基因，结合近缘物种进行家族分类，并对其进行结构分析、亚细胞定位、染色体定位、顺式作用元件、串联重复等分析，同时结合实验室转录组数据，鉴定 GRAS 在茎瘤芥的茎膨大期间的表达分析，为后期茎瘤芥 *GRAS* 基因功能研究及茎瘤芥育种及产量调控等提供参考。

1 结果与分析

1.1茎瘤芥 GRAS 基因家族鉴定

在全基因组中，利用 Blatstp 比对和 HMMER 软件搜索结果合并后，共筛选出 103 个 GRAS 转录因子，然后去除冗余，排除重复序列，将剩余序列上传至 NCBI CDD 数据库、Pfam 数据库和 SMART 数据库进行结构域鉴定，最终确定 102 条序列作为茎瘤芥 GRAS 蛋白(图 1)，远多于模式生物拟南芥的 GRAS 蛋白数目 34 个及水稻中是 GRAS 蛋白数目 60 个。BUSCA 网站亚细胞定位结果发现，有 20 个茎瘤芥 GRAS 蛋白定位于叶绿体，BjuA046833 和 BjuA044545 在叶绿体外膜，BjuA026354 和 BjuB006058 在细胞间隙，BjuB020811 在内膜系统，BjuB028332 在线粒体，BjuB046154 在细胞质，其余 74 个定位于细胞核内。细胞核作为遗传与代谢的调控中心，而超过 70%的 GRAS 蛋白定位于细胞核中，进一步验证了 GRAS 蛋白作为转录因子调节个体中奢侈基因的表达。

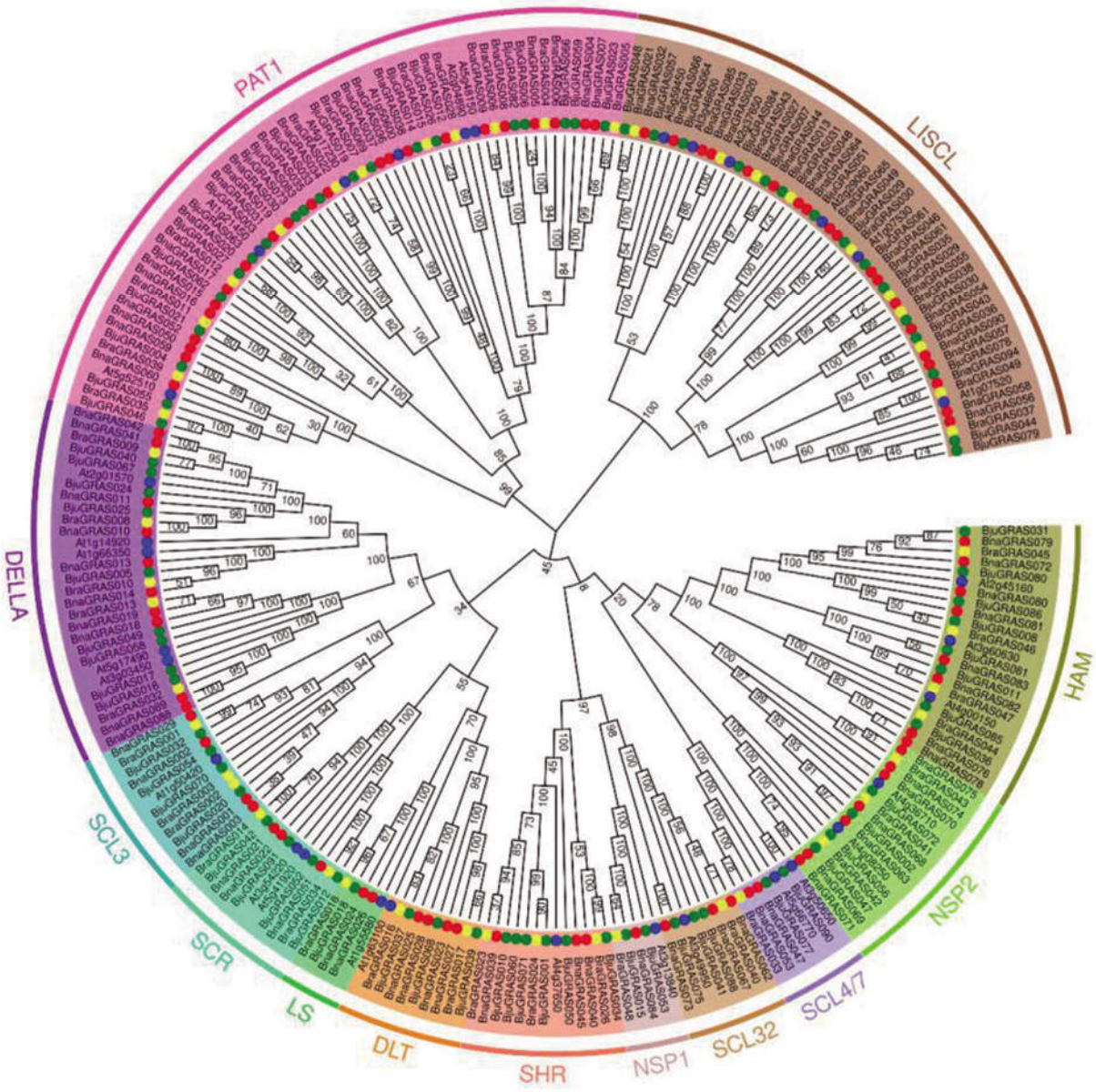

图 1 茎瘤芥, 拟南芥, 芜菁及甘蓝型油菜的 GRAS 基因家族的系统发育树

Figure 1 Phylogenetic tree of the GRAS gene family of *Brassica juncea* var. *tumida Tsen et Lee*, *Arabidopsis*, *Brassica rapa* and *Brassica napus*

1.2 GRAS 蛋白系统发育树的构建

为预测茎瘤芥 *GRAS* 基因的各亚分类和进化发育关系，使用已鉴定的茎瘤芥 *BjuGRAS* 基因与已报道的拟南芥、芸薹及甘蓝型油菜的 *GRAS* 基因进行了 ClustalW 多序列比对，删除比对结果并不理想的某些 GRAS 蛋白序列后，利用 30 个拟南芥、81 个茎瘤芥、48 个芜菁和 88 个甘蓝型油菜 *GRAS* 基因进行了系统发育分析。结果见图 1，四个物种 *GRAS* 基因依据进化树拓扑结构可分为 13 个亚类(Tian et al., 2004; Sun et al., 2011; Liu and Widmer, 2014; Shan et al., 2020)。

我们根据 GRAS 家族成员的结构特征和功能分化，茎瘤芥中的 *GRAS* 基因进一步划分为 13 个亚家族它们分别是：LISCL、PAT1、DELLA、SCL3、SCR、LS、DLT、SHR、NSP1、SCL32、SCL4/7、NSP2 和 HAM。通过系统发育树拓扑结构中可看出，亚家族 PAT1 包含 14 个茎瘤芥 *GRAS* 基因，而亚家族 LS 仅含有 1 个茎瘤芥 *GRAS* 基因，推测茎瘤芥 *GRAS* 基因在各个亚家族间分布存在较大差异。

1.3 茎瘤芥 *GRAS* 基因结构和氨基酸保守域分析

茎瘤芥 GRAS 转录因子基因编码区长度介于 209 bp~19 602 bp 之间，基因家族成员蛋白质的氨基

酸序列平均长度为 513 个残基。其中，基因号 BjuA-024156 的氨基酸残基数目最多达到了 1243 个，而基因号 BjuA026168 的氨基酸残基数目最少为 70 个。ExPASy-ProtParam 网站的预测结果显示，茎瘤芥各 BjuGRAS 蛋白质等电点 pI 介于 4.59~10.07 之间，平均等电点为 5.87，说明大部分茎瘤芥 GRAS 蛋白呈酸性；分子量介于 7.78 kD 至 137.18 kD 之间，平均分子量为 57.21 kD，可以看出茎瘤芥 BjuGRAS 转录因子家族成员理化性质存在较大的组成变异，推测其功能多样性可能较为丰富。

利用 MEME 网站对茎瘤芥 BjuGRAS 蛋白的基序组成及位点分布情况进行了预测，搜索出氨基酸保守域出现概率最大、位点最多的 10 个基序(图 2)。结果显示，预测的 10 个茎瘤芥 BjuGRAS 蛋白基序中基序 5 最为保守，其在 94 个 BjuGRAS 蛋白中均有分布。

使用 TBtools 软件将茎瘤芥 *BjuGRAS* 基因系统发育树、氨基酸保守域及基因结构图进行合并。其中 C- 末端区域和 N- 末端区域的基序分布存在一定差异，即位于 C- 末端区域的基序数量多于位于 N- 末

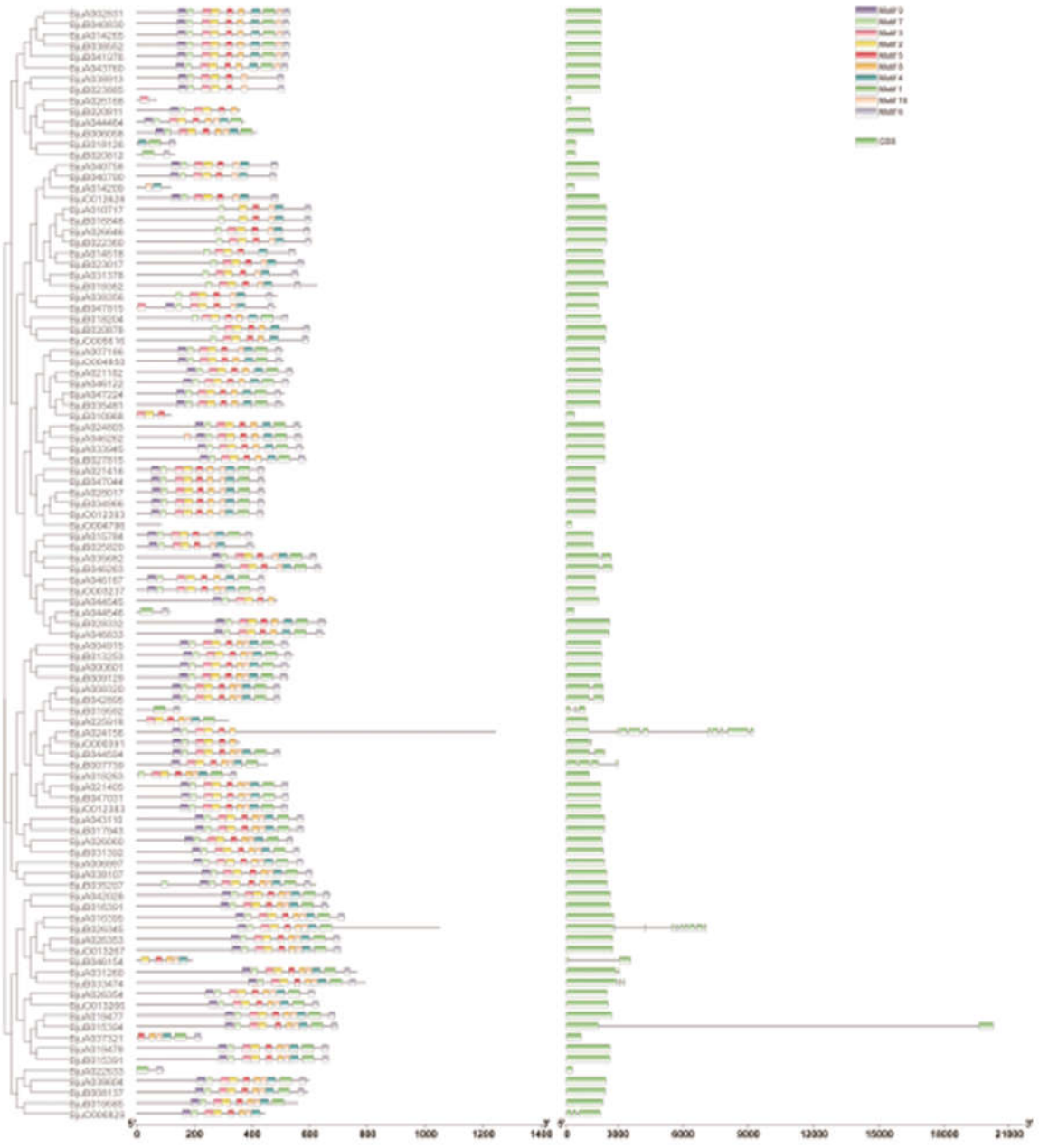

图 2 茎瘤芥 GRAS 基因家族系统进化, 氨基酸保守域及基因结构

Figure 2 Phylogenetic analysis, amino acid conserved domain and gene structure diagram of GRAS gene family in *Brassica juncea* var. *tumida Tsen et Lee*

端区域的基序数量。尽管 *GRAS* 基因的 N- 末端区域的基序可变，但其增加了 *GRAS* 基因功能多样性和生物网络复杂性(Sun et al., 2012; Sun et al., 2011)。不同亚科间基因在 N 末端区域具有不同的基序，但同一亚科中的大多数 GRAS 蛋白具有相似的基序，其为 BjuGRAS 成员在同一亚家族中的紧密进化关系提供了额外的证据，这些基序与相应的 GRAS 结构域匹配(Guo et al., 2017)。基序 9 和 7 在靠近 N- 末端区域一侧的 LHR Ⅰ域中，其后是 VHⅡD 域的基序 3 和 2，LHR Ⅱ结构域中的基序 5 和 8，PFYRE 域中的基序 4、1 和 10，C- 末端区域内的 SAW 结构域中的基序 6。基因结构多样性是基因家族进化的重要组成部分，并进一步支持了系统进化分组(Wei et al., 2016)。茎瘤芥 *GRAS* 基因的内含子数量从 0 个到 10 个不等，在 102 个 *BjuGRAS* 基因中，有 15 个具有内含子，9 个只有一个内含子，87 个没有内含子。

1.4茎瘤芥 *GRAS* 基因染色体定位及顺式作用元件分析

借助 SAMtools 软件和茎瘤芥基因组注释信息，将鉴定得到的茎瘤芥 *GRAS* 基因利用 MG2C 网站进行染色体定位(图 3)，图中各条染色体的长度可由其左侧刻度进行估算。利用前文已鉴定的 102 个 *BjuGRAS* 基因除 *BjuO0004796*、*BjuO0005616*、*BjuO0003237*、*BjuO0006829*、*BjuO0000391* 及 *BjuO0004850* 共六个 Contig 重叠群基因和 *BjuO0012383*、*BjuO0012393*、*BjuO0012628*、*BjuO0013266* 及 *BjuO0013267* 共五个 Scaffold 拼接基因外，其余 91 个 *BjuGRAS* 基因均能定位至茎瘤芥的各条染色体上，在每条染色体上的 *BjuGRAS* 基因数量 2 至 9 个范围内。其中，B07 号染色体分布的 *BjuGRAS* 基因最少仅 2 个；而 A09 号染色体含有 9 个基因，是茎瘤芥 *BjuGRAS* 基因分布最多的染色体。在茎瘤芥的 A、B 两个同源染色体组中，分别含有 50 和 41 个 *GRAS* 基因，暗示茎瘤芥 *GRAS* 基因的含量在 A、B 亚基因组中出现一定的分化，在白菜和黑芥自然杂交并加倍产生茎瘤芥的过程中，同源染色体的保留和丢失在亚基因组间可能存在一定的偏好。

基因家族的扩增和基因组的进化机制主要依靠基因重复事件，其主要重复类型是串联重复和片段重复(Cannon et al., 2004)。基因簇是由重复产生的两个相邻的相关基因或者大量相同基因串联排列而成，属于共同祖先基因的扩增产物，分布在染色体上相对集中的区域。在茎瘤芥 102 个 *BjuGRAS* 基因中，有 22 个基因被确定为串联重复基因占比 24.2%，在图 3 中以红色文字高亮显示，其在染色体上形成 11 个基因簇，在基因家族扩增过程中起着重要作用。顺式作用元件本身不编码任何蛋白质，是存在于基因旁侧序列中能影响基因表达的序列，是转录因子的结合位点，转录因子与之结合，调控结构基因转录的精确起始和转录效率。通过启动子预测数据库 Plant CARE 对茎瘤芥 *BjuGRAS* 基因组上游 1 500 bp 长度的启动子 DNA 序列鉴定出 96 种顺式作用元件，除 TATA-box 启动子及 CAAT-box、GC-motif 增强子等基础作用元件外，还包括多种与生长发育调节、逆境胁迫诱导、光反应调节、激素应答等过程相关等顺式作用元件。本文选取了与激素应答相关的生长素响应(TGA-element, TGA-box)及赤霉素响应元件 (GARE-motif)、光反应调节相关元件(G-box,

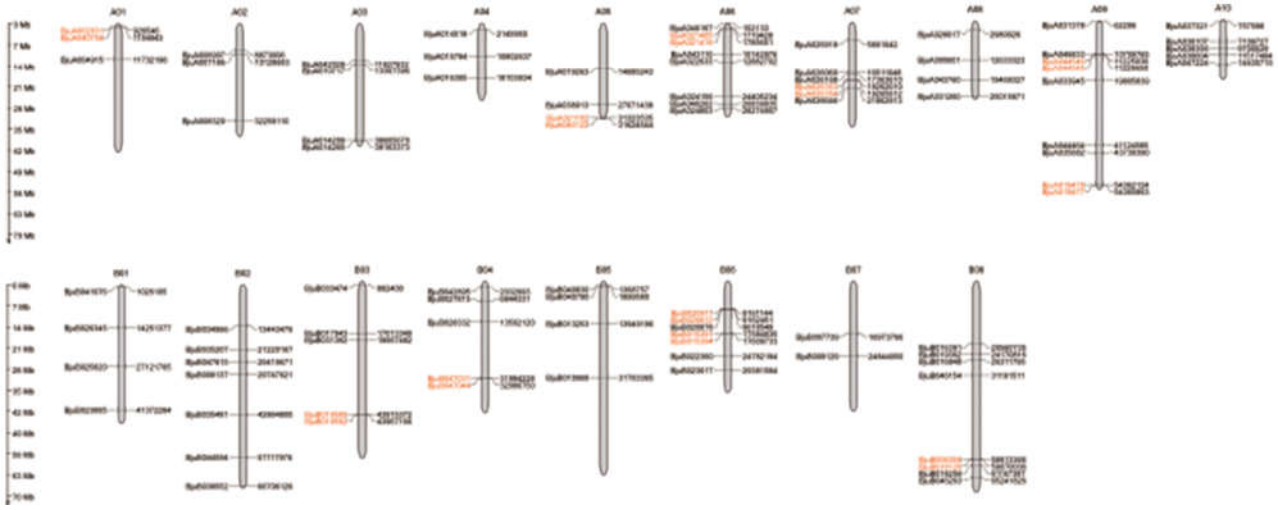

图 3 BjuGRAS 基因染色体定位蜈蚣图

Figure 3 BjuGRAS genes location information on each chromosome

Box II, GT1-motif 等)、与逆境胁迫诱导的脱落酸响应(ABRE)、厌氧诱导(ARE)、防御和压力反应(TC-rich repeats)、低温响应(LTR)、水杨酸响应(TCA-element, SARE)、茉莉酸甲酯响应(TGACG-motif)、生长发育调节相关的昼夜节律控制元件(circadian)与缺氧特异性诱导元件(GC-motif)共 11 种顺式作用元件利用 GSDS 2.0 网站进行结果展示(图 4)。其中，光响应相关顺式作用元件几乎在所有茎瘤芥 *GRAS* 基因的启动子区域中都存在，平均每个 *GRAS* 基因含有 6.8 个，暗示茎瘤芥 GRAS 基因表达可能与光响应调节有关。

1.5 茎瘤芥 *GRAS* 基因串联重复与片段复制分析

分析茎瘤芥 *GRAS* 基因组间共线性关系，对于了解基因间的起源、演化、分化及分类，对于 GRAS 功能研究具有重大的理论意义。图 5 研究了茎瘤芥基因组内各基因间的共线性关系，也反映出各条染色体的加倍信息(Guo et al., 2019)，共有 84 对基因存在共线性关系。其中，灰色部分的连线代表茎瘤芥全基因组中所有基因间存在的共线性关系，绿色连线代表茎瘤芥基因组中出现的染色体加倍及基因重复等现象。BjuA021102&BjuA046122、BjuA024803&BjuA046262、BjuB017943&BjuB031392 三对基因分布于同一染色体上，而 BjuA019478&BjuB015391、BjuA035682&BjuB046263 等存在分布于不同染色体上。通过 Circos 图我们发现 *GRAS* 基因家族在茎瘤芥中由于串联重复和片段重复发生了扩增和基因组进化，在 A、B 亚基因组间大量基因存在染色体加倍及基因重复现象，导致茎瘤芥的 GRAS 基因家族数量远大于拟南芥和水稻等模式生物。

1.6 茎瘤芥 *GRAS* 基因差异表达

结合本实验茎瘤芥转录组测序结果，Blast 比对找出 102 条 *GRAS* 基因，其中 79 条在茎瘤芥的茎中均由表达，说明 79 条 *GRAS* 基因参与茎的发育。以大叶芥突变株茎(无膨大茎) 22 周大叶芥茎(DY_0)为对照组，茎瘤芥播种后 18 周(茎未膨大前, YA1_0)，20 周(茎开始膨大前一周, YA2_0)，22 周(茎膨大后一周, YA3_0)，25 周(茎膨大后一个月, YA4_0)的新鲜茎做为实验组，抽提总 RNA 进行转录组测序，通过实验组与对照组比较，发现 *BjuGRAS020*、*BjuGRAS-032*、*BjuGRAS054*、*BjuGRAS070* 等 4 个基因随着茎膨大，表达量逐渐上调，*BjuGRAS031*、*BjuGRAS039*、*BjuGRAS010*、*BjuGRAS037*、*BjuGRAS060*、*BjuGRAS-027*、*BjuGRAS068*、*BjuGRAS071*、*BjuGRAS096* 等 9 个

图 4 茎瘤芥 BjuGRAS 基因顺式作用元件分析

Figure 4 Cis-element analysis of BjuGRAS genes

基因在茎膨大过程中，与大叶芥突变株茎相比，表达量逐渐下调。*BjuGRAS021*、*BjuGRAS014*、*BjuGRAS050*、*BjuGRAS033*、*BjuGRAS057*、*BjuGRAS048*、*BjuGRAS-085*、*BjuGRAS011*、*BjuGRAS058*、*BjuGRAS016*、*BjuG-*

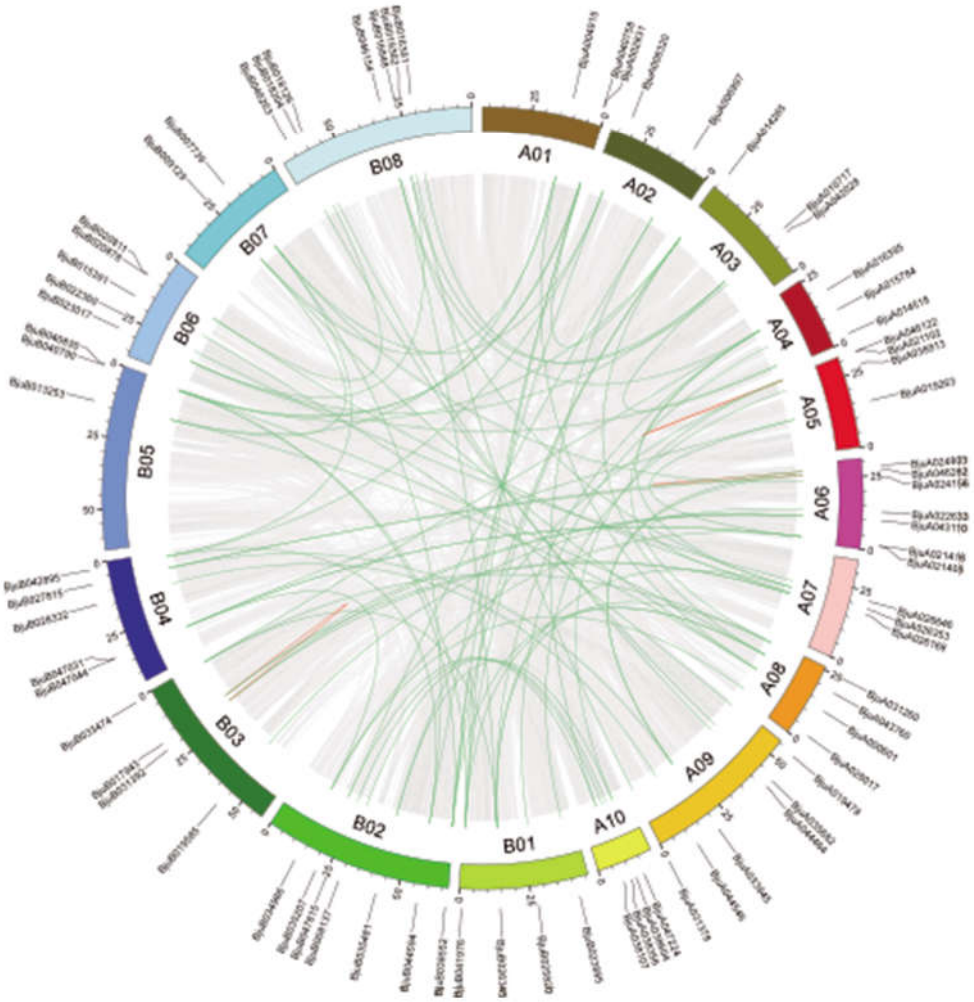

图 5 茎瘤芥 BjuGRAS 基因 MCScanX 分析结果
Figure 5 Collinear analysis of BjuGRAS genes

RAS035、*BjuGRAS064*、*BjuGRAS100* 等 13 个基因在茎瘤芥膨大后逐渐下调，说明这 26 个 *BjuGRAS* 在茎瘤芥的茎膨大过程中发挥重要作用。

2 讨论

本研究对茎瘤芥的 GRAS 基因家族进行了生物信息学分析，通过隐马尔科夫模型和基因组注释文件在茎瘤芥中发现了 102 个 *GRAS* 基因，远多于拟南芥(34 个)和水稻(60 个)等模式生物中 *GRAS* 基因家族成员的数目。我们推测这是由于茎瘤芥是由白菜和黑芥自然杂交后再经加倍形成的异源四倍体的变种，因此在茎瘤芥的 A、B 亚基因组中存在一些同源基因。同时，茎瘤芥发生了 GRAS 基因家族的扩增和基因组的进化，导致茎瘤芥的 GRAS 基因家族数量远大于拟南芥和水稻等模式生物。基因家族扩增和基因组进化的主要方式是串联重复和片段重复，*GRAS* 基因通过串联重复和片段复制保留在茎瘤芥的基因组中，在对各类刺激作出适应性响应过程中起着关键作用(Hanada et al., 2008; Jiang et al., 2010)。通过染色体定位发现，茎瘤芥的 A、B 亚基因组中分别含有 50 和 41 个 *GRAS* 基因，因此我们推断在白菜和黑芥自然杂交并加倍产生茎瘤芥的过程中，同源染色体的保留和丢失在 A、B 亚基因组间可能存在一定的偏好。此外，茎瘤芥 BjuGRAS 转录因子家族成员在等电点、分子量、氨基酸数量等理化性质方面存在较大的组成变异，我们推测 BjuGRAS 作为转录因子蛋白，其功能多样性可能较为丰富。通过亚细胞定位结果发现，茎瘤芥 GRAS 蛋白在细胞核、叶绿体、叶绿体外膜、细胞间隙、内膜系统、线粒体和细胞质中均有分布，但绝大部分 GRAS 蛋白定位于细胞核内，在一定程度上验证了 GRAS 蛋白的功能多样性。

茎瘤芥 *GRAS* 基因的内含子数量从 0 个到 10 个不等，在 102 个 BjuGRAS 基因中，有 15 个具有内含子，9 个只有一个内含子，87 个没有内含子。无内含

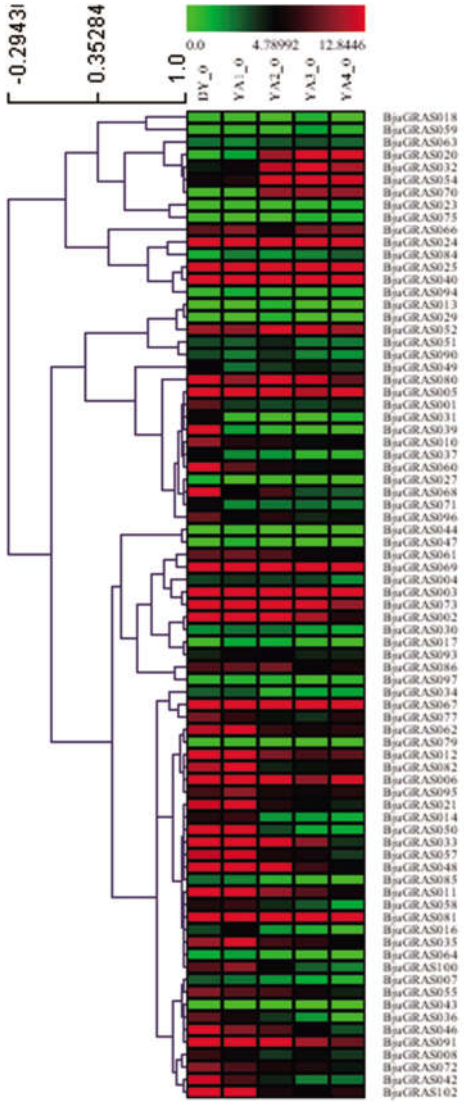

图 6 GRAS 基因在茎瘤芥的瘤茎膨大前后的表达特征

Fig6 Expression profile of BjuGRAS of development of the stem in tumorous stem mustard

子的 GRAS 家族，可能起源于陆地植物的早期，而无内含子的特性可能源于微管植物的早期，后续该特性在基因家族复制和扩展中逐渐放大，具有独特的进化历程。基因上游 1.5 kb 左右的启动子区域顺式作用元件通过响应不同的环境信号调节相应基因的转录过程，对植物的生长发育过程产生重要影响(Liu et al., 2013)。我们通过 Plant CARE 数据库预测发现茎瘤芥 *GRAS* 基因启动子区域共有 96 种 8 868 个顺式作用元件，主要包括与激素应答相关的生长素响应及赤霉素响应元件、光反应调节相关元件、与逆境胁迫诱导的脱落酸响应、厌氧诱导、防御和压力反应、低温响应、水杨酸响应、茉莉酸甲酯响应、生长发育调节相关的昼夜节律控制元件与缺氧特异性诱导元件等顺式作用元件，表明茎瘤芥 *GRAS* 基因在生长发育调节、光反应调节、逆境胁迫诱导、激素应答、组织特异性表达等过程起着重要作用。

3 材料与方法

3.1茎瘤芥 *GRAS* 基因的全基因组鉴定

芸薹属数据库网站(Cheng et al., 2011) (BRAD, http://brassicadb.org/brad/)下载茎瘤芥(*Brassica juncea* L.)、甘蓝型油菜(*Brassica napus*)基因组序列、编码序列和蛋白质序列及芜菁(*Brassica rapa*)的 *GRAS* 基因序列。从 Phytozome (JGI) (Nordberg et al., 2014)数据库网站(https://phytozome.jgi.doe.gov)下载已报道的拟南芥(Arabidopsis thaliana) GRAS 蛋白多肽序列(peptide)。利用 Pfam 数据库(http://pfam.xfam.org/)下载编号为 PF03514 的 GRAS 保守结构域隐马尔科夫模型文件，在 Bio-Linux 系统中使用 HMMER 3.0 软件(http://hmmer.janelia.org/)提供的 hmmsearch 命令，对茎瘤芥的注释基因组序列进行 GRAS 转录因子相关结构域的搜索鉴定。使用 ClustalW (Larkin et al., 2007)多序列比对结果文件构建茎瘤芥特异的隐马尔可夫模型，并对前述序列展开二次搜索，基于 E-value<0.001 筛选，去除重复序列。将前述序列文件利用 NCBI CDD 保守域数据库(https://www.ncbi.nlm.nih.gov/cdd/)、Pfam 数据库 (http://pfam.xfam.org/) 和 SMART 数据库(http://smart.embl.de/)对候选蛋白质序列中 GRAS 保守结构域展开搜索，删除缺失 GRAS 结构域的候选基因，最终获得茎瘤芥 GRAS 转录因子蛋白序列信息。将已经鉴定的茎瘤芥 GRAS 蛋白序列文件上传至 BUSCA 亚细胞成分注释器(Savojardo et al., 2018) (http://busca.biocomp.unibo.it/)进行亚细胞定位分析。

3.2 茎瘤芥 *GRAS* 基因序列比对图的展示及系统发育树的构建

利用 ClustalW 软件，使用默认参数对 GRAS 转录因子蛋白序列进行多序列比对，将比对结果文件导入 ESPript 3 (Robert and Gouet, 2014) 网站(http://espript.ibcp.fr/)绘制基因序列比对图。利用 MEGA X (Kumar et al., 2018)软件，利用删除空位后生成的拟南芥、茎瘤芥、芜菁及甘蓝型油菜 GRAS 蛋白多肽序列的 ClustalW 比对结果文件通过邻接法(neighbor-joining method, NJ)对已报道的拟南芥、水稻及预测的茎瘤芥 GRAS 蛋白质结构域序列进行进化树构建，选择 p-distance 模型，并设置 Boot-strap 参数为 1000，其余参数默认。最后通过 EvolView (Subrama

nian et al., 2019) 网站(https://www.evolgenius.info/evolview/)对系统发育树进行美化。

3.3 茎瘤芥 *GRAS* 基因结构和氨基酸保守域分析将

经过鉴定的茎瘤芥 GRAS 蛋白序列上传至 ExPASyProtParam 网站(http://web.expasy.org/protparam/)分析每条 GRAS 蛋白序列的生化信息 (氨基酸数量, 分子量和等电点等)。通过 MEME (Bailey et al., 2009) 网站(http://meme-suite.org/tools/meme)对蛋白质序列中保守 Motif 序列进行预测，寻找 10 个 Motif，并设置预测 Motif 宽度范围 6~50 个残基，其余使用默认参数。通过 Gene Structure Display Server 2.0 (Hu et al., 2015) (http://gsds.cbi.pku.edu.cn/) 网站对茎瘤芥 GRAS 基因结构进行分析，展示外显子 Exon、蛋白质编码区 CDS、内含子 intron 等结构。使用 TBtools 软件将 GRAS 转录因子蛋白系统发育树、Motif 序列预测结构图和基因结构图进行合并。

3.4茎瘤芥 *GRAS* 基因染色体定位及顺势作用元件分析

借助 SAMtools (Li et al., 2009)工具提取茎瘤芥各条染色体长度信息。使用 Perl 脚本提取 BjuGRAS 各基因位置信息。利用 MapGene2Chromosome (Chao et al., 2015) 网站(http://mg2c.iask.in/mg2c_v2.1/)构建 *BjuGRAS* 基因在茎瘤芥各条染色体上的物理图谱。利用 Perl 语言脚本程序提取茎瘤芥 *GRAS* 基因组上游 1 500 bp 长度的启动子 DNA 序列，并将其提交至启动子预测数据库 Plant CARE (Lescot et al., 2002) (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) 进行顺式作用元件预测，整理预测结果，并将结果文件使用 GSDS 2.0 网站和 TBtools 软件进行绘图。

3.5 茎瘤芥 *GRAS* 基因串联重复与片段复制分析

利用茎瘤芥各 *GRAS* 基因间 Blast (Altschul et al., 1990)同源比对结果分析茎瘤芥 *GRAS* 基因间串联重复情况。基因串联重复鉴定使用以下两个标准：①在相对于较长的基因的情况下，两个基因的比对率大于 70%，且比对相似性大于 70%；②两个基因在染色体上的位置小于 100 kb (Vatansever et al., 2016)。将满足上述条件的基因手动筛选后在染色体物理图谱上标出。使用 MCScanX (Wang et al., 2012)软件利用共线性分析方法，对基因的加倍与复制现象进行分析。准备染色体信息文件、共线性基因及区块信息文件、标明各基因间共线性关系注释文件及主配置文件，基于 Bio-Linux 系统命令行使用 Circos 绘图软件(http://circos.ca/)对共线性分析结果进行可视化展示。

3.6 茎瘤芥 *GRAS* 基因差异表达

大叶芥突变株茎(无膨大茎)为对照组，22 周大叶芥茎(DY_0)，茎瘤芥播种后 18 周(茎未膨大前, YA1_0)，20 周(茎开始膨大前一周, YA2_0)，22 周(茎膨大后一周, YA3_0)，25 周(茎膨大后一个月, YA4_0)的新鲜茎做为实验组，抽提总 RNA 进行转录组测序，所有实验材料均由重庆市涪陵农科所提供。

作者贡献

蒋龙星是本研究的执行人，完成本论文的数据分析和论文初稿撰写；郭佳鑫和孙全负责试验设计和论文修改定稿；何晓红是项目的构思人和负责人，指导试验设计和论文修改定稿。全体作者都阅读并同意最终的文本。

致谢

本研究由重庆市基础与前沿研究计划项目(cstc2015jcyjA80006)项目资助。

参考文献

Altschul S.F., Gish W., Miller W., Myers E.W., and Lipman D.J., 1990, Basic local alignment search tool, J. Mol. Biol., 215 (3): 403-410

Bailey T.L., Boden M., Buske F.A., Frith M., Grant C.E., Clementi L., Ren J., Li W.W., and Noble W.S., 2009, MEME SUITE: tools for motif discovery and searching, Nucleic Acids Res., 37(Web Server Issue): W202-W208

Bolle C., Koncz C., and Chua N.H., 2000, PAT1, a new member of the GRAS family, is involved in phytochrome A signal transduction, Genes Dev., 14(10): 1269-1278Cann

on S.B., Mitra A., Baumgarten A., Young N.D., and May G., 2004, The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana, BMC Plant Biol., 4: 10

Chao J.T., Kong Y.Z., Wang Q., Sun Y. H., Gong D.P., Lv J., and Liu G.S., 2015, MapGene2Chrom, a tool to draw gene physical map based on Perl and SVG languages, Yi Chuan, 37(1): 91-97

Cheng F., Liu S.Y., Wu J., Fang L., Sun S.L., Liu B., Li P.X., Hua W., and Wang X.W., 2011, BRAD, the genetics and genomics database for Brassica plants, BMC Plant Biol., 11: 136

Di Laurenzio L., Wysocka-Diller J., Malamy J.E., Pysh L., Helar-

iutta Y., Freshour G., Hahn M.G., Feldmann K.A., and Benfey P.N., 1996, The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the Arabidopsis root, Cell, 86(3): 423-433

Guiltinan M.J., and Miller L., 1994, Molecular characterization of the DNA-binding and dimerization domains of the bZIP transcription factor, EmBP-1, Plant Mol. Biol., 26 (4): 1041-1053

Guo P. C., Wen J., Yang J., Ke Y.Z., Wang M.M., Liu M.M., Ran F., Wu Y.W., Li P.F., Li J.N., and Du H., 2019, Genome-wide survey and expression analyses of the GRAS gene family in Brassica napus reveals their roles in root development and stress response, Planta, 250(4): 1051-1072

Guo Y., Wu H., Li X., Li Q., Zhao X., Duan X., An Y., Lv W., and An H., 2017, Identification and expression of GRAS family genes in maize (Zea mays L.), PLoS One, 12 (9): e0185418Hanada K., Zou C.,

Lehti-Shiu M.D., Shinozaki K., and Shiu S.H., 2008, Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli, Plant Physiol., 148(2): 993-1003

Hu B., Jin J.P., Guo A.Y., Zhang H., Luo J.C., and Gao G., 2015, GSDS 2.0: an upgraded gene feature visualization server, Bioinformatics (Oxford, England), 31(8): 1296-1297

Jiang S.Y., Ma Z.G., and Ramachandran S., 2010, Evolutionary history and stress regulation of the lectin superfamily in higher plants, BMC Evol. Biol., 10: 79

Karin M., 1990, Too many transcription factors: positive and negative interactions, New Biol., 2(2): 126-131

Kumar S., Stecher G., Li M., Knyaz C., and Tamura K., 2018, MEGA X: Molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol., 35(6): 1547-1549

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., and Higgins D. G., 2007, Clustal W and Clustal X version 2.0, Bioinformatics (Oxford, England), 23(21): 2947-2948

Latchman D.S., 1997, Transcription factors: an overview, Int. J. Biochem. Cell Biol., 29(12): 1305-1312Le

scot M., Déhais P., Thijs G., Marchal K., Moreau Y., Van de Peer Y., Rouzé P., and Rombauts S., 2002, PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences, Nucleic Acids Res., 30(1): 325-327

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., and Durbin R., 2009, The sequence Alignment/Map format and SAMtools, Bioinformatics (Oxford, England), 25(16): 2078-2079

Liu X. Y., and Widmer A., 2014, Genome-wide comparative analysis of the GRAS gene family in populus, Arabidopsis and Rice, Plant Mol. Biol. Rep., 32(6): 1129-1145

Liu Y.K., Zhang D., Wang L., and Li D.Q., 2013, Genome-Wide analysis of Mitogen-Activated protein kinase gene family in maize, Plant Mol. Biol. Rep., 31(6): 1446-1460

Nordberg H., Cantor M., Dusheyko S., Hua S., Poliakov A., Shabalov I., Smirnova T., Grigoriev I.V., and Dubchak I., 2014, The genome portal of the Department of Energy Joint Genome Institute: 2014 updates, Nucleic Acids Res., 42 (Database issue): D26-D31

Peng J., Carol P., Richards D.E., King K.E., Cowling R.J., Murphy G.P., and Harberd N.P., 1997, The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin responses, Genes Dev., 11(23): 3194-3205

Pysh L.D., Wysocka-Diller J.W., Camilleri C., Bouchez D., and Benfey P.N., 1999, The GRAS gene family in Arabidopsis: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes, Plant J., 18(1): 111-119

Robert X., and Gouet P., 2014, Deciphering key features in protein structures with the new ENDscript server, Nucleic Acids Research, 42(Web Server issue): W320-W324

Savojardo C., Martelli P.L., Fariselli P., Profiti G., and Casadio R., 2018, BUSCA: an integrative web server to predict subcellular localization of proteins, Nucleic Acids Res., 46 (W1): W459-W466

Shan Z., Luo X., Wu M., Wei L., Fan Z., and Zhu Y., 2020, Genome-wide identification and expression of GRAS gene family members in cassava, BMC Plant Biol., 20(1): 46

Silverstone A.L., Ciampaglio C.N., and Sun T., 1998, The Arabidopsis RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway, Plant Cell, 10(2): 155-169

Subramanian B., Gao S., Lercher M. J., Hu S.N., and Chen W.H., 2019, Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees, Nucleic Acids Res., 47(W1): W270-W275

Sun X.L., Jones W.T., and Rikkerink E.H., 2012, GRAS proteins: the versatile roles of intrinsically disordered proteins in plant signalling, Biochem. J., 442(1): 1-12

Sun X., Xue B., Jones W.T., Rikkerink E., Dunker A.K., and Uversky V.N., 2011, A functionally required unfoldome from the plant kingdom: intrinsically disordered N-terminal domains of GRAS proteins are involved in molecular recognition during plant development, Plant Mol. Biol., 77(3):

Tian C.G., Wan P., Sun S.H., Li J.Y., and Chen M.S., 2004, Genome-wide analysis of the GRAS gene family in rice and Arabidopsis, Plant Mol. Biol., 54(4): 519-532

orres-Galea P., Huang L.F., Chua N.H., and Bolle C., 2006, The GRAS protein SCL13 is a positive regulator of phy-

tochrome-dependent red light signaling, but can also modulate phytochrome A responses, Mol. Genetics Genomics, 276(1): 13-30

Vatansever R., Koc I., Ozyigit I.I., Sen U., Uras M.E., Anjum N. A., Pereira E., and Filiz E., 2016, Genome-wide identification and expression analysis of sulfate transporter (SULTR) genes in potato (Solanum tuberosum L.), Planta, 244 (6): 1167-1183

Wang Y.P., Tang H.B., Debarry J.D., Tan X., Li J.P., Wang X.Y., Lee T.H., Jin H.Z., Marler B., Guo H., Kissinger J.C., and Paterson A.H., 2012, MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity, Nucleic Acids Res., 40(7): e49

Wei Y.X., Shi H.T., Xia Z.Q., Tie W.W., Ding Z.H., Yan Y., Wang W.Q., Hu W., and Li K.M., 2016, Genome-Wide Identification and Expression Analysis of the WRKY Gene Family in Cassava, Front. Plant Sci., 7: 25

Yang J.H., Liu D.Y., Wang X.W., Ji C.M., Cheng F., Liu B. N., Hu Z.Y., Chen S., Pental D., Ju Y.H., Yao P., Li X.M., Xie K., Zhang J.H., Wang J.L., Liu F., Ma W.W., Shopan J., Zheng H.K., Mackenzie S.A., and Zhang M.F., 2016, The genome sequence of allopolyploid Brassica juncea and analysis of differential homoeolog gene expression influencing selection, Nat. Genetics, 48(10): 1225-1232