



大豆固氮相关的硫酸盐转运基因进化分析

陈磊, 邵志伟, 高兴, 郑宇轩, 朱新宇

(南通大学 生命科学学院, 江苏 南通 226019)

摘要: 在植物进化过程中, 不同亚家族的硫酸盐转运基因在功能上发生了广泛的分化, 因而有必要理解其背后的分子进化机制。为研究与大豆固氮相关的硫酸盐转运基因的分子进化机制, 本研究从基因结构和选择压力两个方面对涉及根瘤固氮功能分化的大豆硫酸盐转运基因进行进化分析。基因结构分析鉴别了保守和可变的基因结构, 并与固氮相关的大豆硫酸盐转运基因的表达歧异建立关联。选择压力分析发现了大豆固氮相关的 *GmSULTR2;3* 序列存在两段正选择区域, 可能与它的固氮根瘤分化有关。本研究结果提供了大豆固氮相关的硫酸盐转运基因分子进化的新认识, 有助于对其功能的进一步研究。

关键词: 大豆; 硫酸盐转运体; 功能分化; 分子进化

Evolutionary Analysis on Nitrogen Fixation Related Sulfate Transporter Genes in Soybean

CHEN Lei, SHAO Zhi-wei, GAO Xing, ZHENG Yu-xuan, ZHU Xin-yu

(School of Life Sciences, Nantong University, Nantong 226019, China)

Abstract: The functional diversification of plant sulfate transporters (SULTRs) have extensively occurred among SULTR sub-families. It is intriguing to further investigate the molecular evolution behind the functional divergences. In this study, the gene structures and selection pressure analyses were performed to investigate the molecular evolutionary mechanism of the nitrogen fixation related functional diversification of soybean *SULTR* genes. Conserved and variable *SULTR* gene structures were recognized across different clades, and were associated with the expression divergences of nitrogen fixation related *SULTR* genes. Two positive selection regions operated on nitrogen fixation related *GmSULTR2;3* gene were identified, which might be responsible for the transformation of this gene into root nodule genes. These findings provide new insights into the molecular evolution of the nitrogen fixation related *SULTR* genes in soybean, and help to further functional investigation.

Keywords: Soybean; Sulfate transporters; Functional divergence; Molecular evolution

作为植物的必需养分, 土壤中硫元素通过植物细胞跨膜蛋白硫酸盐转运体 (sulfate transporters, *SULTR*) 吸收, 并在植物体内转运^[1]。被子植物的植物硫酸盐转运体包含约 10 个跨膜螺旋和 1 个 C 末端的 STAS 结构域^[2-3]。分子系统发育学研究已经将植物硫酸盐转运体分类为 4 个亚家族 (*SULTR1-4*)^[4], 其中的 *SULTR4* 亚家族被分辨为 *SULTRs* 的祖先形式。在陆地植物进化中, 亚家族 *SULTR1/2/3* 进一步分割成 *SULTR1/2* 和 *SULTR3* 两个分支, 然后, *SULTR1/2* 再扩展成两个被子植物分支 (*SULTR1* 和 *SULTR2*)。

在豆科植物中, 2 个 *SULTR* 基因被发现与根瘤形成过程有关。大豆 *nodulin 70* (*GmN70*, *GmSULTR2;3*) 被发现表达在根瘤出现的时间^[5], 随后进一步证实其主要表达在成熟的固氮根瘤中, 并且定位在共生小体的膜上^[6]。百脉根的 *Sst1* 基因 (*Lj2g3v0776860*) 能够将胞浆中的硫酸盐运输到胞内的根瘤菌中^[7]。*GmN70* 和 *Sst1* 基因在大豆和百

脉根转录组和蛋白质组分析中被进一步鉴别为候选根瘤基因^[8-9]。拟南芥的 *SULTR* 基因的功能研究已经较为充分, 其 *SULTR* 基因通常表达在特定的细胞或亚细胞区室中, 功能上, 它们相互协调地从土壤中获得硫酸盐, 然后在植物体内转运。在拟南芥中, *SULTR1* 亚家族的两个高亲和力成员, *SULTR1;1* 和 *SULTR1;2* 负责从土壤中获得硫酸盐^[10]。*SULTR2* 亚家族的两个低亲和力成员, *SULTR2;1* 和 *SULTR2;2* 负责运输硫酸盐到木质部^[11]。*SULTR3* 亚家族的成员功能上呈现多样化, 如 *SULTR3;1*、*SULTR3;2* 和 *SULTR3;3* 倾向表达在叶组织^[12], *SULTR3;5* 在 *SULTR2;1* 的介导下能强化硫酸盐根到芽的运输^[13], *SULTR3;2*、*SULTR3;3*、*SULTR3;4* 和 *SULTR3;5* 能为正在成熟的胚提供硫酸盐^[14]。*SULTR4* 亚家族的两个成员 *SULTR4;1* 和 *SULTR4;2* 共同定位在液泡膜, 介导硫酸盐从液泡中流出^[15]。

基因重复是基因功能分化的基础^[16-17]。植物 *SULTR* 基因经历多次基因重复形成了功能广泛分

收稿日期: 2018-04-26

基金项目: 江苏省自然科学基金面上项目 (BK20151269)。

第一作者简介: 陈磊 (1994 -), 男, 硕士, 主要从事植物分子进化研究。E-mail: 16090003@yjs.ntu.edu.cn。

通讯作者: 朱新宇 (1963 -), 男, 博士, 教授, 主要从事植物分子进化研究。E-mail: zhuxinyu@ntu.edu.cn。

化的基因家族^[4],然而,功能歧异背后的分子进化机理还有待进一步研究。目前,已经采用了多种方法研究功能分化的分子进化基础,其中,基于序列比对的氨基酸变异类型的统计学方法^[18],基于密码子变异的选择压力测试方法^[19],以及基于基因结构变异的功能分化的分析方法^[20],是通常采用的有效的分析手段。在本研究中,我们聚焦大豆 *SULTR* 基因的固氮功能分化,从基因结构和选择压力变化两个方面探索大豆 *SULTR* 基因的固氮功能分化的分子进化基础,为深入研究 *SULTR* 基因在大豆生长、发育、根瘤固氮等方面的功能作用提供可靠基础。

1 材料与方法

1.1 材料

最近版本的大豆 (*Glycine max*) 基因组注释数据从 Phytozome 平台 (<https://phytozome.jgi.doe.gov/pz/portal.html>)^[21] 下载。使用大豆蛋白序列 GLYMA18G02230 (*GmSULTR2;3*) 作为查询序列,通过 HMMER3.0^[22] 搜索下载的大豆基因组数据,搜索 E 值设置为 0.01。击中的候选序列通过 SMART 平台 (<http://pfam.sanger.ac.uk/search>)^[23] 检查,以确保包含 STAS 结构域。

1.2 方法

1.2.1 多序列比对与系统发育分析 使用 MUSCLE 3^[24] 对大豆 *SULTRs* 的全长蛋白质序列进行比对,手工检查比对结果以删除模糊比对区域和自衍插入缺失位点。相应的密码子比对用 PAL2NAL 程序^[25] 构建。使用 MEGA7^[26] 测试最优进化模型和优化参数,然后用 PhyML3.1^[27] 构建蛋白质最大似然系统发育树 (ML),以 *SULTR4* 亚家族作为外类群^[4]。树内部节点的可靠性评估采用基于 100 次重抽样的 bootstrap 方法。

1.2.2 基因结构分析 由于基因结构对基因家族的功能进化具有重要意义^[28-29],本研究检查大豆 *SULTR* 基因的内含子数量和相位。大豆 *SULTR* 基因的编码区序列 (CDS) 和相应的基因组序列从 Phytozome^[21] 获取,外显子-内含子结构信息通过 GSDS 程序 (<http://gsds.cbi.pku.edu.cn/>)^[30] 鉴别。

1.2.3 选择压力评估 蛋白质序列中氨基酸具有不同的非同义替代速率 (Ka) 和同义替代速率 (Ks),计算平行同源基因对的 ω ($= Ka/Ks$) 比率值,判断氨基酸所处的选择压力,大于 1 说明存在正选择,小于 1 提示负选择。为了鉴别 *SULTR2* 亚家族中发生固氮功能分化的 *GmSULTR2;3* (*N70*) 成员与亚家族内部的其它非固氮功能成员之间是否发生正选择歧异进化,使用 KaKs_Calculator 2.0^[31] 计算这些平行同源基因对的 Ka/Ks 比率。在 *SULTR2* 亚家族中,共选择 4 个编码区完整的成员用于计算与

GmSULTR2;3 (*N70*) 之间的 Ka/Ks 比率,计算参数设定如下: method = γ - MYN^[32], Window Length = 45 bp, Step Length = 6 bp。在计算之前,比对中的空位 (GAP) 栏被删除。

2 结果与分析

2.1 基因组尺度的 *SULTR* 基因的鉴别

基于最新的大豆基因组注释数据,共抽提了 29 条 *SULTR* 基因序列 (表 1)。一般被子植物基因组包含 10 ~ 12 个 *SULTR* 基因同源拷贝^[4],大豆基因组包含两倍多的拷贝数量。相比以前报道的 28 个同源拷贝数量^[33],我们鉴别的大豆 *SULTR* 基因拷贝数量增加 1 个拷贝 (GLYMA03G09435, *GmSULTR1;1c*) (表 1)。这个增加的拷贝属于 *SULTR1* 亚家族,与其它 *SULTR1* 成员比较,它是唯一定位在 3 号染色体的同源拷贝,因此确认是一个新鉴别的拷贝 (表 1)。

表 1 大豆硫酸盐转运体基因列表
Table 1 List of soybean *SULTR* genes

Phytozome 访问号 Phytozome accession	别名 Alias	Uniprot 访问号 Uniprot accession
GLYMA04G11582	<i>GmSULTR1;1a</i>	K7KJB8
GLYMA06G11140	<i>GmSULTR1;1b</i>	I1KA20
GLYMA03G09435	<i>GmSULTR1;1c</i>	K7KDA3
GLYMA08G19240	<i>GmSULTR1;2a</i>	I1KUE7
GLYMA15G05760	<i>GmSULTR1;2b</i>	I1MDV4
GLYMA13G02060	<i>GmSULTR1;3a</i>	I1LW07
GLYMA14G34221	<i>GmSULTR1;3b</i>	K7M7J3
GLYMA08G14700	<i>GmSULTR2;1a</i>	I1KT54
GLYMA11G36210	<i>GmSULTR2;1b</i>	I1LN03
GLYMA18G02240	<i>GmSULTR2;1c</i>	I1MYU4
GLYMA06G11150	<i>GmSULTR2;2a</i>	I1KA21
GLYMA14G34210	<i>GmSULTR2;2b</i>	I1MAQ9
GLYMA13G02080	<i>GmSULTR2;2c</i>	K7LX04
GLYMA18G02230	<i>GmSULTR2;3, N70</i>	I1MYU2
GLYMA03G31310	<i>GmSULTR3;1a</i>	I1JNX0
GLYMA19G34160	<i>GmSULTR3;1b</i>	I1N9K0
GLYMA02G16365	<i>GmSULTR3;2a</i>	K7K8D7
GLYMA10G03460	<i>GmSULTR3;2b</i>	I1L867
GLYMA07G34370	<i>GmSULTR3;3a</i>	I1KM60
GLYMA20G02080	<i>GmSULTR3;3b</i>	I1ND92
GLYMA07G34360	<i>GmSULTR3;3c</i>	I1KM59
GLYMA07G00840	<i>GmSULTR3;4a</i>	I1KG82
GLYMA08G22120	<i>GmSULTR3;4b</i>	I1KV89
GLYMA13G43670	<i>GmSULTR3;4c</i>	I1M5M2
GLYMA15G01710	<i>GmSULTR3;4d</i>	I1MCM8
GLYMA07G09710	<i>GmSULTR3;5a</i>	I1KIS8
GLYMA09G32110	<i>GmSULTR3;5b</i>	I1L4G4
GLYMA18G52270	<i>GmSULTR4;1</i>	I1N522
GLYMA02G10590	<i>GmSULTR4;2</i>	I1JDS8

2.2 序列比对与系统发育分析

在排除模糊区域和自衍插入后,蛋白质比对产生一个 602 个氨基酸残基长度的矩阵和一个相应的 1 806 个核苷酸长度的密码子。在 AIC 标准下蛋白质比对的最优进化模型为 LG 模型,优化参数为: I = 0.018 9、G = 1.083 3。最大似然法系统发育分析 (ML) 产生一个最大似然分值为 -49 006. 233 54 的系统发育树(图 1)。

系统发育分析重现了以前的植物 *SULTR* 基因

的系统发育关系^[4]。大豆 *SULTR* 基因被清楚地划分为 4 个亚家族(*SULTR* 1~4),每个亚家族具有高度的 bootstrap 支持值 (BP > 85%)。*SULTR*3 亚家族基因扩展为 4 个亚支,分别编码 *SULTR*3;1/3;2、*SULTR*3;3、*SULTR*3;4 和 *SULTR*3;5。图 1 中,*SULTR*3 亚家族中的 *SULTR*3;5 分支,及 *SULTR*2 亚家族中的 *GmSULTR*2;3(N70) 分支具有相对长的分支长度,说明在进化过程中积累了相对更多的氨基酸替代。

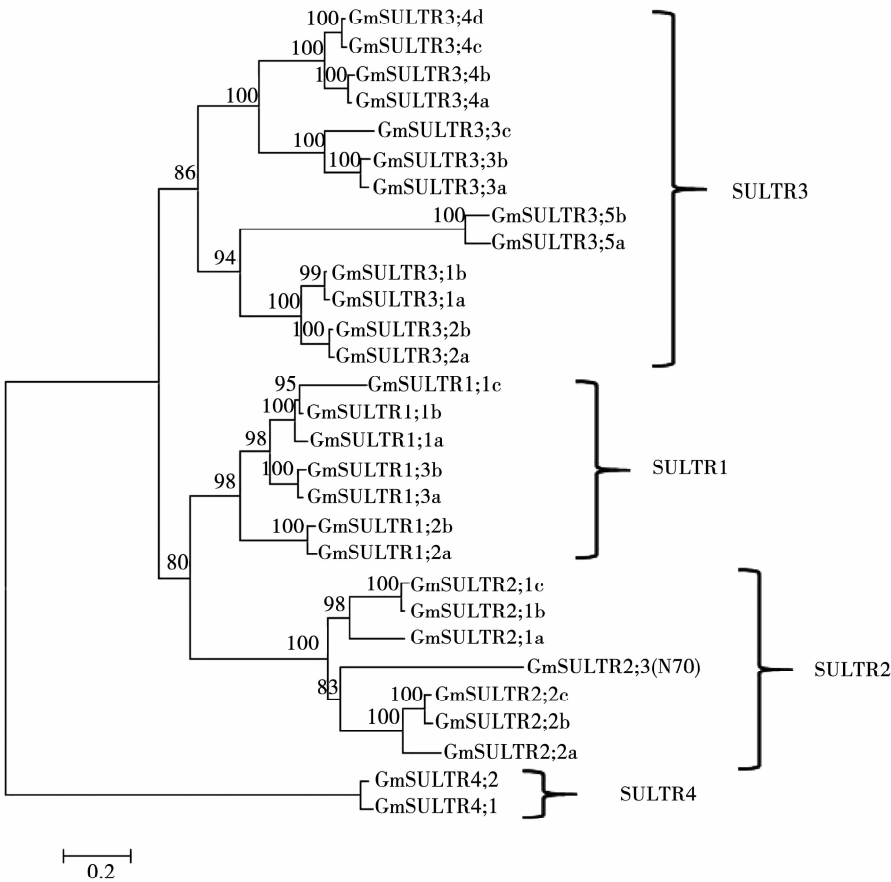


图 1 大豆硫酸盐转运蛋白的最大似然树

Fig 1 Maximum likelihood (ML) tree of soybean *SULTR* proteins

2.3 基因结构分析

所有的大豆 *SULTR* 基因包含内含子,总共 335 个内含子存在于 29 大豆 *SULTR* 基因中,内含子数量从 8 到 16 不等,其中,*SULTR*4 亚家族的两个成员具有 16 个内含子,相比其它亚家族成员的内含子数量明显更多(表 2)。尽管内含子的数量和相位 (phase) 在亚家族内部和之间存在歧异,但在每个亚家族或分支内部保守的外显子 - 内含子结构清晰可辨(表 2)。在 *SULTR*1~3 三个亚家族中,共同保守的基因结构为“12000010210”,这个结构在 *SULTR*1、*SULTR*2 以及 *SULTR*3 亚家族内部的 *SULTR*3;

3 和 *SULTR*3;4 亚支,其最左侧增加了一个 0 相位内含子。在 335 个内含子中,192 (57%) 属于相位 0, 90 (27%) 属于相位 1, 53 (16%) 属于相位 2 (表 2), 类似于以前对水稻中 21 570 个基因的内含子分析报告,即 57.3% 属于相位 0, 21.5% 属于相位 1, 21.2% 属于相位 2^[34]。此外,*SULTR*4 亚家族的内含子数量和相位组成与其它 3 个亚家族存在显著差异,意味着 *SULTR*1/2/3 谱系在与祖先 *SULTR*4 分化后经历了内含子丢失,这一现象支持“内含子早”的理论^[35-36]。

表 2 大豆 *SULTR* 基因的内含子数量和相位的统计

Table 2 Statistic of the phase and number of introns in soybean *SULTR* genes

分支 Clade	基因名 Gene name	每一种相位的内含子数量 No. of introns in each phase			总的内含子数量 Total No. of introns	内含子的相位构成 Intron phase constitution
		0	1	2		
<i>SULTR1</i>	<i>GmSULTR1;2a</i>	7	3	2	12	012000010210
	<i>GmSULTR1;2b</i>	7	3	2	12	012000010210
	<i>GmSULTR1;1b</i>	7	3	2	12	012000010210
	<i>GmSULTR1;1c</i>	4	3	1	8	01110200
	<i>GmSULTR1;1a</i>	7	4	1	12	011000010210
	<i>GmSULTR1;3b</i>	6	2	1	9	012000010
	<i>GmSULTR1;3a</i>	7	3	2	12	012000010210
<i>SULTR2</i>	<i>GmSULTR2;1b</i>	6	3	2	11	01200010210
	<i>GmSULTR2;1c</i>	6	3	2	11	01200010210
	<i>GmSULTR2;1a</i>	6	3	2	11	01200010210
	<i>GmSULTR2;3, N70</i>	6	3	2	11	01200010210
	<i>GmSULTR2;2b</i>	7	3	1	11	01200010010
	<i>GmSULTR2;2c</i>	7	3	1	11	01200010010
	<i>GmSULTR2;2a</i>	6	3	2	11	01200010210
<i>SULTR3;1/2</i>	<i>GmSULTR3;2a</i>	6	3	2	11	12000010210
	<i>GmSULTR3;2b</i>	6	3	2	11	12000010210
	<i>GmSULTR3;1b</i>	6	3	2	11	12000010210
	<i>GmSULTR3;1a</i>	6	3	2	11	12000010210
<i>SULTR3;3</i>	<i>GmSULTR3;3c</i>	6	4	2	12	012000010211
	<i>GmSULTR3;3a</i>	7	3	2	12	012000010210
	<i>GmSULTR3;3b</i>	7	3	2	12	012000010210
<i>SULTR3;4</i>	<i>GmSULTR3;4c</i>	7	3	2	12	012000010210
	<i>GmSULTR3;4d</i>	7	3	2	12	012000010210
	<i>GmSULTR3;4b</i>	7	3	2	12	012000010210
	<i>GmSULTR3;4a</i>	6	4	2	12	112000010210
<i>SULTR3;5</i>	<i>GmSULTR3;5b</i>	6	2	2	10	1200000210
	<i>GmSULTR3;5a</i>	6	3	2	11	12000010210
<i>SULTR4</i>	<i>GmSULTR4;2</i>	10	4	2	16	0120000010001210
	<i>GmSULTR4;1</i>	10	4	2	16	0120000010001210
总数 Total		192	90	53	335	

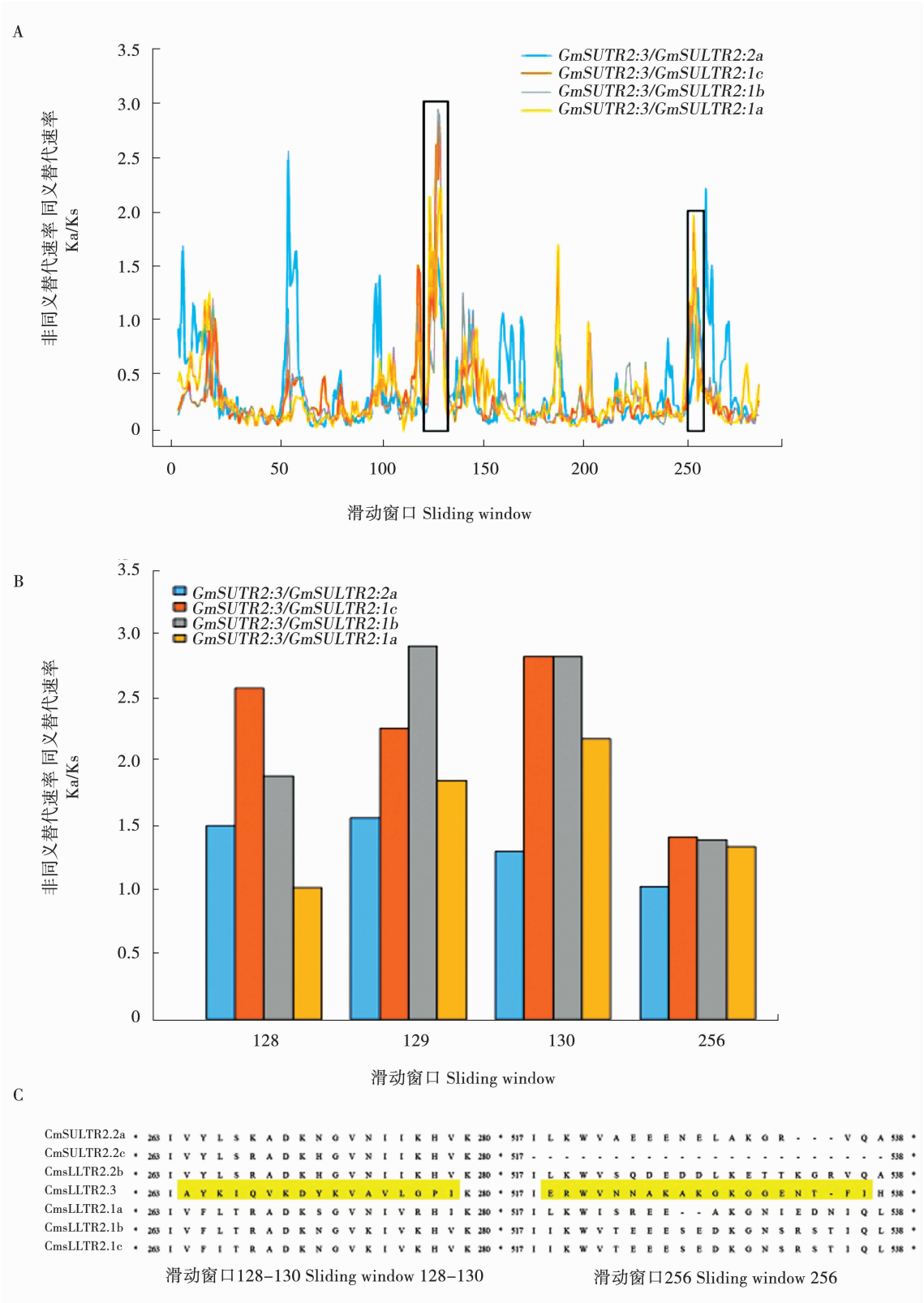
粗体字表示保守的内含子相位构成。

The bold letters represent the conserved intron constitution in each clade.

2.4 选择压力评估

通过同义 (Ks) 和非同义 (Ka) 比率的计算可以确定平行同源基因对之间的选择压力类型。利用滑动窗口方法研究发现在 4 个平行同源基因对之间的总体上处于负选择压力之下 ($\omega < 1$), 但能够鉴别 2 个共同的正选择区域, 其 ω ($= Ka/Ks$) 比率在 4 个平行同源基因对之间均大于 1, 最高达到 2.86 (图 2A)。我们放大这 2 个区域, 发现其对应的窗口分别为 128 – 130 和 256 (图 2B)。我们进一步确定这

两个窗口对应的 *SULTR2* 亚家族的蛋白质比对区段, 并在 *GmSULTR2;3* (*N70*) 序列中鉴别了相应的氨基酸残基区段 (黄色标注) (图 2C)。很容易发现, 这两个黄色区段的氨基酸残基与其它成员 (非根瘤功能分化) 的氨基酸残基绝大部分是不同的, 提示发生了连续的非同义替代, 这两个区段可能是 *GmSULTR2;3* 发生根瘤固氮功能分化的潜在因素, 同时不排除存在其它导致根瘤固氮功能分化因素, 如启动子区的顺式作用元件的歧异。



A: *GmSULTR2;3* 基因与 4 个平行同源基因之间的 K_a/K_s 比率; B: 在平行同源基因对之间发生正选择的滑动窗口; C: 在固氮相关基因 *GmSULTR2;3* 上发生正选择的 2 个氨基酸区段。

A: K_a/K_s ratio between *GmSULTR2;3* gene and its four paralogs; B: The sliding windows with positive selection occurred between four pairs of paralogs; C: Two regions of positive selection operated on the nitrogen fixation related *GmSULTR2;3* gene.

图 2 固氮相关基因 *GmSULTR2;3* 与平行同源基因之间的选择压力分析

Fig. 2 Selection pressure analysis between nitrogen fixation related *GmSULTR2;3* gene and its paralogs

3 讨论

植物具有复杂的涉及硫代谢过程的转运网络^[37]。高等植物基因组通常包含10个以上的 *SULTR* 同源基因^[4],而大豆基因组比其它被子植物基因组包含两倍多的 *SULTR* 基因,这可能由于经历的两次全基因组重复^[38-39],或者由于适应性进化的需要^[40]。大豆 *SULTR* 基因在进化过程获得了为根瘤内共生小体提供硫养分的功能创新^[11-12],因此,理解这种功能分化背后的进化机制非常重要,这有助于设计更有效的实验方案展开进一步的研究。

大豆的外显子-内含子结构变化在3个大豆 *SULTR* 亚家族中被识别(表2)。在亚家族 *SULTR1* 中,3个基因, *GmSULTR1;1a*、*GmSULTR1;1c* 和 *GmSULTR1;3b*,与保守的结构“012000010210”比较发生了歧异(表2),这3个基因在大豆 RNA-Seq 表达地图(Atlas of *Glycine max*)^[41]中没有被检索到,因此,我们推测它们的基因结构歧异可能是未被检索到的原因。在 *SULTR2* 亚家族中, *GmSULTR2;2b* 和 *GmSULTR2;2c* 两个基因与保守的结构“012000010210”比较发生了歧异(表2),它们的功能变化值得关注。在 *SULTR3* 亚家族中, *GmSULTR3;3c*、*GmSULTR3;4a* 和 *GmSULTR3;5b* 这3个基因与保守的结构“012000010210”比较发生了歧异(表2),其中的 *GmSULTR3;5b* 和 *GmSULTR3;5a* 是百脉根 *SULTR* 基因(命名为 *Sst1*)共直系同源基因(co-orthologs)^[14,15],而 *Sst1* 基因功能上为百脉根的胞内根瘤菌提供硫养分^[13]。结果显示在大豆 RNA-Seq 表达地图(Atlas of *Glycine max*)^[41]中,大豆 *GmSULTR3;5b* 和 *GmSULTR3;5a* 在根瘤中具有极高的表达水平,不同的是, *GmSULTR3;5a* 在不同组织中广泛表达,如根、种子、花和叶中,而 *GmSULTR3;5b* 仅限于在根瘤中表达,推测这种变化可能与 *GmSULTR3;5b* 的基因结构歧异有关(表2)。以前的研究认为基因的外显子-内含子结构歧异常常导致平行同源基因(paralogs)的功能分化^[42],因而,大豆 *SULTR* 基因中这些发生基因结构歧异的基因,它们的功能可能也发生了变化。

非同义与同义替代比率 ω ($=K_a/K_s$) 决定了序列编码区的适应性选择的类型。在 *SULTR2* 亚家族,固氮相关基因 *GmSULTR2;3* 与4个平行同源基因之间的平均 K_a/K_s 比率为0.32,指示着这4个平行同源基因对处于强的负选择压力下。然而,在4个平行同源基因对之间存在两个区段具有一致的正选择类型,平均 K_a/K_s 比率为1.89,指示这两个区段在蛋白质水平上可能发生了功能歧异,导致 *GmSULTR2;3* 基因的进化速率加速(图1)。因此,推测这两个正选择区段可能使得大豆 *GmSULTR2;3*

进化成为负责为根瘤内共生小体提供硫养分的新功能,以适应它的环境需要。重复基因的进化可能有4种命运:保持功能不变(functional conservation),亚功能化(sub-functionalization),新功能化(neo-functionalization)和假基因化(pseudogenization)^[43]。本研究结果表明, *GmSULTR2;3* 中存在的这两个正选择区段代表部分新功能化的进化命运,而3个 *SULTR1* 亚家族的基因,即 *GmSULTR1;1a*、*GmSULTR1;1c* 和 *GmSULTR1;3b*,由于歧异的基因结构不能探测它们的基因表达,可能已经假基因化。

4 结论

本研究对大豆硫酸盐转运基因根瘤固氮功能分化的进化机制进行了探索。基因结构分析显示,涉及根瘤固氮的大豆 *GmSULTR2;3* (*N70*) 基因保持保守的基因结构,而涉及根瘤固氮功能的百脉根硫酸盐转运基因 *Sst1* 的大豆直系同源基因 *GmSULTR3;5b* 显示了基因结构歧异,并反映在它的表达歧异上。选择压力分析发现,大豆 *GmSULTR2;3* (*N70*) 序列中存在两段正选择区域,推测这两段区域可能与它的根瘤固氮的功能分化存在潜在关联。本研究结果有助于对大豆硫酸盐转运基因在生长、发育和根瘤固氮等方面的深入研究。

参考文献

[1] Takahashi H, Kopriva S, Giordano M, et al. Sulfur assimilation in photosynthetic organisms: Molecular functions and regulations of transporters and assimilatory enzymes[J]. Annual Review of Plant Biology, 2011, 62: 157-184.

[2] Rouached H, Berthomieu P, El Kassis E, et al. Structural and functional analysis of the C-terminal STAS (sulfate transporter and anti-sigma antagonist) domain of the *Arabidopsis thaliana* sulfate transporter SULTR1.2 [J]. Journal of Biological Chemistry, 2005, 280(16): 15976-15983.

[3] Sharma A K, Rigby A C, Alper S L. STAS domain structure and function[J]. Cell Physiol Biochemical, 2011, 28(3): 407-422.

[4] Takahashi H, Buchner P, Yoshimoto N, et al. Evolutionary relationships and functional diversity of plant sulfate transporters[J]. Frontiers in Plant Science, 2012, 2: 119.

[5] Kouchi H, Hata S. Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development [J]. Molecular Genetics and Genomics, 1993, 238(1-2): 106-119.

[6] Vincill E D, Szczygłowski K, Roberts D M. GmN70 and LjN70. Anion transporters of the symbiosome membrane of nodules with a transport preference for nitrate[J]. Plant Physiology, 2005, 137: 1435-1444.

[7] Krusell L, Krause K, Ott T, et al. The sulfate transporter SST1 is crucial for symbiotic nitrogen fixation in *Lotus japonicus* root nodules[J]. The Plant Cell, 2005, 17(5): 1625-1636.

[8] Libault M, Farmer A, Brechenmacher L, et al. Complete transcriptome of the soybean root hair cell, a single-cell model, and its

- alteration in response to *Bradyrhizobium japonicum* infection[J]. Plant Physiology, 2010, 152(2): 541-552.
- [9] Clarke V C, Loughlin P C, Gavrin A, et al. Proteomic analysis of the soybean symbiosome identifies new symbiotic proteins[J]. Molecular & Cellular Proteomics, 2015, 14(5): 1301-1322.
- [10] Yoshimoto N, Takahashi H, Smith F W, et al. Two distinct high-affinity sulfate transporters with different inducibilities mediate uptake of sulfate in *Arabidopsis* roots[J]. The Plant Journal, 2002, 29(4): 465-473.
- [11] Takahashi H, Watanabe-Takahashi A, Smith F W, et al. The roles of three functional sulphate transporters involved in uptake and translocation of sulphate in *Arabidopsis thaliana* [J]. The Plant Journal, 2000, 23(2): 171-182.
- [12] Cao M J, Wang Z, Wirtz M, et al. SULTR3;1 is a chloroplast-localized sulfate transporter in *Arabidopsis thaliana* [J]. The Plant Journal, 2013, 73(4): 607-616.
- [13] Kataoka T, Hayashi N, Yamaya T, et al. Root-to-shoot transport of sulfate in *Arabidopsis*. Evidence for the role of SULTR3;5 as a component of low-affinity sulfate transport system in the root vasculature[J]. Plant Physiology, 2004, 136(4): 4198-4204.
- [14] Zuber H, Davidian J C, Aubert G, et al. The seed composition of *Arabidopsis* mutants for the group 3 sulfate transporters indicates a role in sulfate translocation within developing seeds [J]. Plant Physiology, 2010, 154(2): 913-926.
- [15] Kataoka T, Watanabe-Takahashi A, Hayashi N, et al. Vacuolar sulfate transporters are essential determinants controlling internal distribution of sulfate in *Arabidopsis* [J]. The Plant Cell, 2004, 16(10): 2693-2704.
- [16] Zhang J. Evolution by gene duplication: An update[J]. Trends in Ecology & Evolution, 2003, 18(6): 292-298.
- [17] Panchy N, Lehti-Shiu M, Shiu S H. Evolution of gene duplication in plants[J]. Plant Physiology, 2016, 171(4): 2294-2316.
- [18] Gu X. Statistical methods for testing functional divergence after gene duplication[J]. Molecular Biology and Evolution, 1999, 16(12): 1664-1674.
- [19] Kondrashov F A, Rogozin I B, Wolf Y I, et al. Selection in the evolution of gene duplications[J]. Genome Biology, 2002, 3(2): RESEARCH0008.
- [20] Xu G X, Guo C C, Shan H Y, et al. Divergence of duplicate genes in exon-intron structure[J]. Proceedings of the National Academy of Sciences of the United States of America, 2012, 109(4): 1187-1192.
- [21] Goodstein D M, Shu S, Howson R, et al. Phytozome: A comparative platform for green plant genomics [J]. Nucleic Acids Research, 2012, 40(Database issue): D1178-D1186.
- [22] Finn R D, Clements J, Arndt W, et al. HMMER web server; 2015 update [J]. Nucleic Acids Research, 2015, 43 (W1): W30-W38.
- [23] Letunic I, Doerks T, Bork P. SMART: Recent updates, new developments and status in 2015 [J]. Nucleic Acids Research, 2015, 43(Database issue): D257-D260.
- [24] Edgar R C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput[J]. Nucleic Acids Research, 2004, 32(5): 1792-1797.
- [25] Suyama M, Torrents D, Bork P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments[J]. Nucleic Acids Research, 2006, 34 (Web Server issue): W609-W612.
- [26] Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets[J]. Molecular Biology and Evolution, 2016, 33(7): 1870-1874.
- [27] Guindon S, Dufayard J F, Lefort V, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0 [J]. Systematic Biology, 2010, 59(3): 307-321.
- [28] Stoltzfus A, Logsdon J M, Palmer J D, et al. Intron 'sliding' and the diversity of intron positions[J]. Proceedings of the National Academy of Sciences of the United States of America, 1997, 94(20): 10739-10744.
- [29] Rogozin I B, Sverdlov A V, Babenko V N, et al. Analysis of evolution of exon-intron structure of eukaryotic genes[J]. Briefings in Bioinformatics, 2005, 6(2): 118-134.
- [30] Hu B, Jin J, Guo A Y, et al. GSDS 2.0: An upgraded gene feature visualization server[J]. Bioinformatics, 2015, 31(8): 1296-1297.
- [31] Wang D, Zhang Y, Zhang Z, et al. KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies[J]. Genomics, Proteomics & Bioinformatics, 2010, 8(1): 77-80.
- [32] Wang D P, Wan H L, Zhang S, et al. Gamma-MYN: A new algorithm for estimating Ka and Ks with consideration of variable substitution rates[J]. Biology Direct, 2009, 4: 20.
- [33] Ding Y, Zhou X, Zuo L, et al. Identification and functional characterization of the sulfate transporter gene GmSULTR1;2b in soybean[J]. BMC Genomics, 2016, 17: 373.
- [34] Lin H, Zhu W, Silva J C, et al. Intron gain and loss in segmentally duplicated genes in rice [J]. Genome Biology, 2006, 7(5): R41.
- [35] Gilbert W: The exon theory of genes[J]. Cold Spring Harbor Symposium on Quantitative Biology, 1987, 52: 901-905.
- [36] Roy S W, Gilbert W. Complex early genes[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(6): 1986-1991.
- [37] Takahashi H. Regulation of sulfate transport and assimilation in plants[J]. International Review of Cell and Molecular Biology, 2010, 281: 129-159.
- [38] Shoemaker R C, Polzin K, Labate J, et al. Genome duplication in soybean (*Glycine subgenus soja*) [J]. Genetics, 1996, 144(1): 329-338.
- [39] Schmutz J, Cannon S B, Schlueter J, et al. Genome sequence of the palaeopolyploid soybean [J]. Nature, 2010, 463 (7278): 178-183.
- [40] Doyle J J, Luckow M A. The rest of the iceberg. Legume diversity and evolution in a phylogenetic context [J]. Plant Physiology, 2003, 131(3): 900-910.
- [41] Severin A J, Woody J L, Bolon Y T, et al. RNA-Seq Atlas of *Glycine max*: A guide to the soybean transcriptome[J]. BMC Plant Biology, 2010, 10: 160.
- [42] Xu G, Guo C, Shan H, et al. Divergence of duplicate genes in exon-intron structure[J]. Proceedings of the National Academy of Sciences of the United States of America, 2012, 109 (4): 1187-1192.
- [43] Lynch M, Conery J S. The evolutionary fate and consequences of duplicate genes[J]. Science, 2000, 290(5494): 1151-1155.