

大豆 *XHS* 基因家族生物信息学分析

陈红梅

(安庆师范学院 生命科学学院,安徽 安庆 246011)

摘 要:利用生物信息学方法,通过全基因组数据扫描对大豆 *XHS* 家族基因进行鉴定和分类,并对其基因结构、染色体定位和基因组织表达方式进行分析。结果表明:大豆基因组中含有 21 个 *XHS* 家族基因,分布于大豆的 12 条染色体上,系统发育分析将 *XHS* 家族基因分成 4 个亚类,有 EST 数据的大豆 *XHS* 家族基因有 10 个,大多在叶片和胚中表达。研究结果促进了大豆 *XHS* 家族基因的功能研究和利用。

关键词:大豆;*XHS* 基因家族; 生物信息学

中图分类号:S565. 1 **文献标识码:**A **DOI:**10. 11861/j. issn. 1000-9841. 2015. 03. 0384

Primary Bioinformatics Analysis of the *XHS* Family Genes in Soybean

CHEN Hong-mei

(School of Life Science, Anqing Normal University, Anqing 246011, China)

Abstract: Based on soybean genome database and bioinformatics method, soybean *XHS* family genes were identified and their sequences, gene structure and chromosome location were analyzed. Besides of those expression patterns of *XHS* genes were also analyzed. The results showed that a total of 21 *XHS* genes were systematically identified from soybean and classified into four types. They were located on 12 chromosomes. 10 *XHS* genes were found EST evidence and the genes expression were rich in sites, most of them were expressed in leaf and somatic embryo. The results facilitate functional analysis and utilization of *XHS* genes in soybean genetic improvement.

Keywords: Soybean; *XHS* gene family; Bioinformatics

大豆蛋白质可以根据其结构的相似或序列的相似聚集成蛋白家族,通过对蛋白家族中域的研究,可以预测其生物学功能。在植物中将含有 XS、XH 和 Zf-XS 三种结构域的蛋白,称为 *XHS* 家族蛋白,通过分析拟南芥 *XHS* 家族蛋白和水稻 *XHS* 家族蛋白,发现 *XHS* 基因是植物中一个非常古老的蛋白家族^[1],只存在于植物体内。在 XS 结构域中,包含一个高度保守的天冬氨酸,可能是酶的活性中心,XH 结构域中含有一个极度保守的谷氨酸,预测 XH 结构域可以和 XS 结构域互作^[2]。Zf-XS 结构域是典型的锌指结构,含有高度保守的组氨酸和半胱氨酸,还有多个保守的精氨酸和亮氨酸,可能有与核酸结合的功能^[1]。Rose 等^[3]通过研究指出 *XHS* 蛋白是植物中含有 coiled-coil 结构域的一类,可能与植物的信号转导相关。*XHS* 家族中有功能报道的是拟南芥 SGS3,它不是一个典型的 *XHS* 蛋白,而是通过与其它蛋白形成复合物介导了 PTGS^[4]。SGS3 的 XS 结构域结合有 RNA 的活性,是一个双链 RNA 结合蛋白^[5]。最早克隆到的 *XHS* 基因是水稻中的 X 基因,并推测可能是一个转录因子,但尚未有有力证据^[6],在水稻逆境芯片中,*XHS* 家族绝大多数基因都会被逆境诱导^[7-8]。通过 Real-time PCR 对水稻中 10 个 *XHS* 基因作苗期逆境表达分析,包括干旱、盐、冷等,发现除 1 个基因外,其余均至少受一种胁迫诱导,证明 *XHS* 基因在水稻逆境中起着较

为重要的作用^[1]。

大豆(*Glycine max*)为豆科(Leguminosae)、蝶形花亚科(Papilionatae)植物,在我国有着悠久的栽培历史,是一种重要的粮食作物,同时也是一种优质的植物蛋白质资源。随着生物信息学的飞速发展和大豆基因组测序的完成^[9],大豆分子生物学研究也越来越深入,但是大豆中关于 *XHS* 的信息并不多。为了更好地了解大豆 *XHS* 基因家族的结构与功能,对大豆 *XHS* 基因家族进行了生物信息学分析,首次在全基因组水平进一步挖掘大豆基因组信息,研究大豆 *XHS* 家族基因的分类、染色体定位、组织器官表达等,以期为进一步认识该家族成员的功能提供理论依据。

1 材料与方法

1.1 材料

大豆全基因组数据和相关编码区序列数据信息来自于 Phytozome 网站的大豆相关数据库(<http://www.phytozome.org/soybean.php>)和 NCBI 数据库(<http://www.ncbi.nlm.nih.gov/>)。

1.2 大豆 *XHS* 基因家族成员鉴定和基因结构分析

利用水稻和拟南芥中已鉴定的所有 *XHS* 蛋白序列,使用 SMART 数据库工具^[10]获得多种相关的典型蛋白结构域。以两种模式植物典型 *XHS* 结构域序列 tBLASTN($P = 0.001$)在大豆基因组数据库

中搜索,并手工剪切无完整读码框的序列,获得大豆相关同源序列。利用 Pfam 数据库工具^[11]·对大豆相关同源序列结构域作进一步分析,去除无典型 XHS 结构域序列,得到大豆中所有 XHS 基因家族成员。利用 GSDS 在线工具 (<http://gsds.cbi.pku.edu.cn/>)^[10]制作大豆 *XHS* 基因外显子-内含子结构图。

1.3 大豆 XHS 蛋白家族基因定位

利用 MapDraw V2.1 软件^[12],根据大豆基因组数据库提供的基因相关定位信息,确定相关 *XHS* 基因家族成员在各染色体上准确的位置,并按照相关基因和染色体的长度比例绘制基因染色体分布图。

1.4 大豆 XHS 基因家族的的系统发育分析

利用 BioEdit V7.0.1 软件对大豆 XHS 蛋白序列进行多序列比对,参数为默认值,联配结果使用 MEGA V5.1^[13],采用邻接法(Neighbor-Joining Method)生成 *XHS* 基因的系统进化树,校正参数 bootstrap 值设置为重复 1 000 次,生成最终的系统进化树。

1.5 大豆 XHS 蛋白家族基因表达分析

在 soybase 表达数据库中筛选,寻找大豆 XHS 蛋白家族基因在不同组织器官中的相关表达数据。对检索到的大豆 XHS 序列按组织器官的来源进行分类,获得大豆 XHS 蛋白家族基因的表达信息。

2 结果与分析

2.1 大豆 XHS 蛋白家族基因

在大豆基因组中检索,将重复序列去除后,在大豆中共发现了同源序列 21 条,这 21 个蛋白构成的家族命名为 GmXHS 家族,分别命名为 GmXHS1-21。同时含有 3 个完整结构域的分别为:*GmXHS2*、*GmXHS4*、*GmXHS7*、*GmXHS8*、*GmXHS10*、*GmXHS11*、*GmXHS12*、*GmXHS13*、*GmXHS18*、*GmXHS20*。

利用生物信息学方法,对大豆 XHS 家族基因进行了分子量、蛋白质长度、等电点、DNA 长度等基本属性分析见表 1。大豆中 21 个 *XHS* 基因中 *GmXHS2* 所编码的蛋白分子量最大,为 133.035 kDa, *GmXHS17* 所编码的蛋白分子量最小,为 13.913 kDa。21 个 XHS 基因所编码蛋白的理论等电点为 5.21 (*GmXHS3*) ~ 9.53 (*GmXHS21*)。21 个 *XHS* 基因所编码的氨基酸长度差异很大,氨基酸长度从 122 ~ 1 154 个不等,DNA 长度相差也很大,从 1 654 ~ 9 891 bp 不等。

2.2 大豆 XHS 基因家族染色体定位

基因复制和基因分化在一定程度上会影响基因家族的形成。基因在染色体上的分布在一定程度上也影响了基因的功能。根据大豆基因组数据库,绘制大豆 *GmXHS* 基因在染色体上的分布图。由图 1 可知,21 个 *GmXHS* 分布较为分散,分别分布在大豆 12 个染色体上。4、10、11、13、14、20 号各分布有 1 个 *GmXHS* 基因,2、7、16 和 18 号染色体上分

布了 2 个 *GmXHS* 基因,8 号染色体上分布了 3 个 *GmXHS* 基因,15 号染色体上分布了 4 个 *GmXHS* 基因。其中,第 7、15 和 18 条染色体上分布的基因是以基因簇的形式存在的,提示这些基因可能是在基因组的串联复制中形成,而串联复制也可能在大豆 *XHS* 基因家族的形成中起到相对重要的作用,使大豆 *XHS* 基因家族基因数量得到扩增。

表 1 大豆 XHS 家族基因的基本信息
Table 1 Basic information of XHS family genes in soybean

| 基因 Gene | 蛋白质长度 Protein length/AA | 分子量 Molecular weight/kDa | 染色体位置 Chromosome location | 等电点 pI | DNA 长度 DNA length/bp |
|----------------|-------------------------------|--------------------------------|---------------------------------|-----------|----------------------------|
| <i>GmXHS1</i> | 410 | 47.811 | 2 | 6.24 | 3465 |
| <i>GmXHS2</i> | 1153 | 133.035 | 2 | 8.63 | 8633 |
| <i>GmXHS3</i> | 191 | 22.475 | 4 | 5.21 | 1654 |
| <i>GmXHS4</i> | 629 | 73.216 | 7 | 5.58 | 5503 |
| <i>GmXHS5</i> | 236 | 27.555 | 7 | 5.5 | 1731 |
| <i>GmXHS6</i> | 611 | 69.983 | 8 | 5.77 | 4075 |
| <i>GmXHS7</i> | 629 | 73.251 | 8 | 5.56 | 5058 |
| <i>GmXHS8</i> | 641 | 74.500 | 8 | 6.22 | 6082 |
| <i>GmXHS9</i> | 460 | 51.161 | 10 | 9.22 | 9891 |
| <i>GmXHS10</i> | 683 | 78.940 | 11 | 5.63 | 6862 |
| <i>GmXHS11</i> | 679 | 79.692 | 13 | 8.81 | 4796 |
| <i>GmXHS12</i> | 1154 | 132.641 | 14 | 8.48 | 9417 |
| <i>GmXHS13</i> | 670 | 78.447 | 15 | 8.82 | 6116 |
| <i>GmXHS14</i> | 609 | 69.981 | 15 | 5.61 | 4187 |
| <i>GmXHS15</i> | 538 | 62.879 | 15 | 5.41 | 3409 |
| <i>GmXHS16</i> | 543 | 63.500 | 15 | 5.55 | 4265 |
| <i>GmXHS17</i> | 122 | 13.913 | 16 | 6.71 | 2144 |
| <i>GmXHS18</i> | 638 | 74.275 | 16 | 5.28 | 4085 |
| <i>GmXHS19</i> | 356 | 42.024 | 18 | 6.95 | 3601 |
| <i>GmXHS20</i> | 631 | 73.640 | 18 | 5.77 | 4914 |
| <i>GmXHS21</i> | 475 | 52.920 | 20 | 9.53 | 8936 |

2.3 大豆 XHS 家族基因的系统进化分析

为进一步了解大豆中 XHS 蛋白系统进化关系,使用大豆 *XHS* 基因编码蛋白序列构建系统发育树(图 2),大豆 21 个 *XHS* 基因分为 4 类,A 类含有 9 个 *XHS* 基因,分别为 *GmXHS18*、*GmXHS20*、*GmXHS8*、*GmXHS2*、*GmXHS12*、*GmXHS4*、*GmXHS7*、*GmXHS9*、*GmXHS21*;B 类含有 7 个 *XHS* 基因,包括 *GmXHS6*、*GmXHS14*、*GmXHS10*、*GmXHS1*、*GmXHS19*、*GmXHS11*、*GmXHS13*;C 类含有 2 个 *XHS* 基因,包括 *GmXHS3*、*GmXHS5*;D 类含有 3 个 *XHS* 基因,包括 *GmXHS17*、*GmXHS15*、*GmXHS16*。

2.4 大豆 XHS 基因家族的基因结构

大豆 *GmXHS* 基因结构分析如图 3 所示。大豆中的 21 个 *XHS* 基因均含有内含子和外显子。内含子的数目为 1 ~ 9 个不等,外显子数目为 2 ~ 10 个,其中 *GmXHS2* 和 *GmXHS12* 内含子和外显子数目最多,*GmXHS17* 内含子和外显子数目最少。大豆 XHS 家族具有较多内含子,其中含有 1 ~ 3 个内含子的基因占 23.8%,含有 3 个以上内含子的基因占 76.2%。大豆 XHS 蛋白基因内含子和外显子结构的多样化,可能对基因功能的分化具有一定的作用。

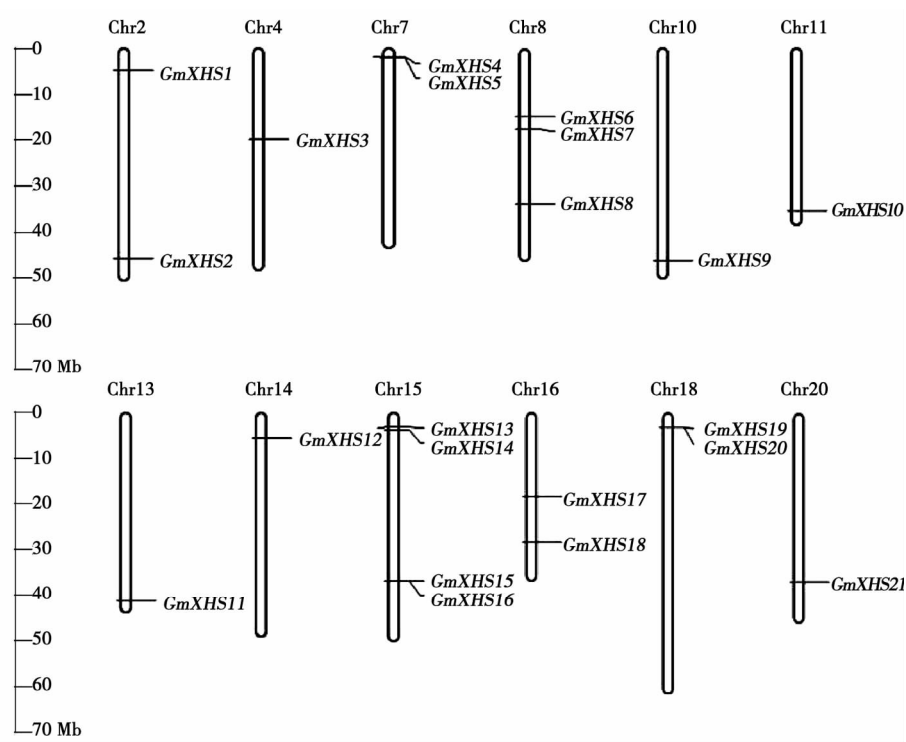


图 1 大豆 *XHS* 基因在染色体上的定位

Fig. 1 Chromosomal location for soybean *XHS* genes

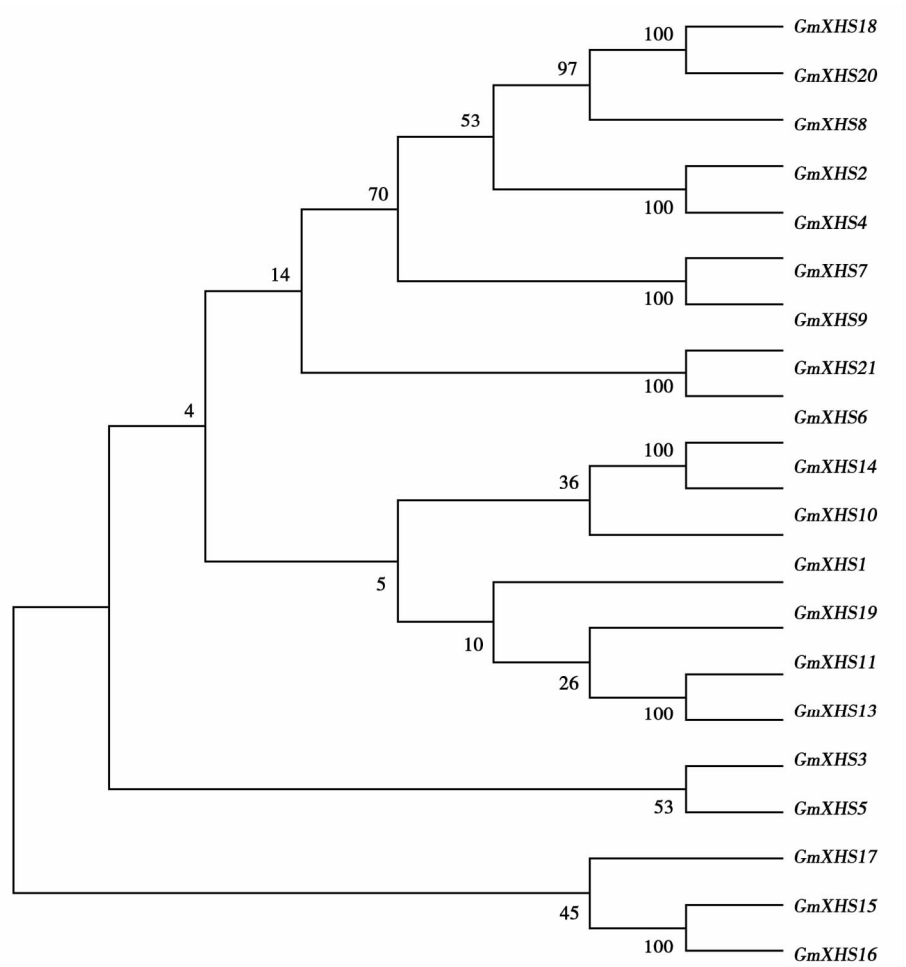
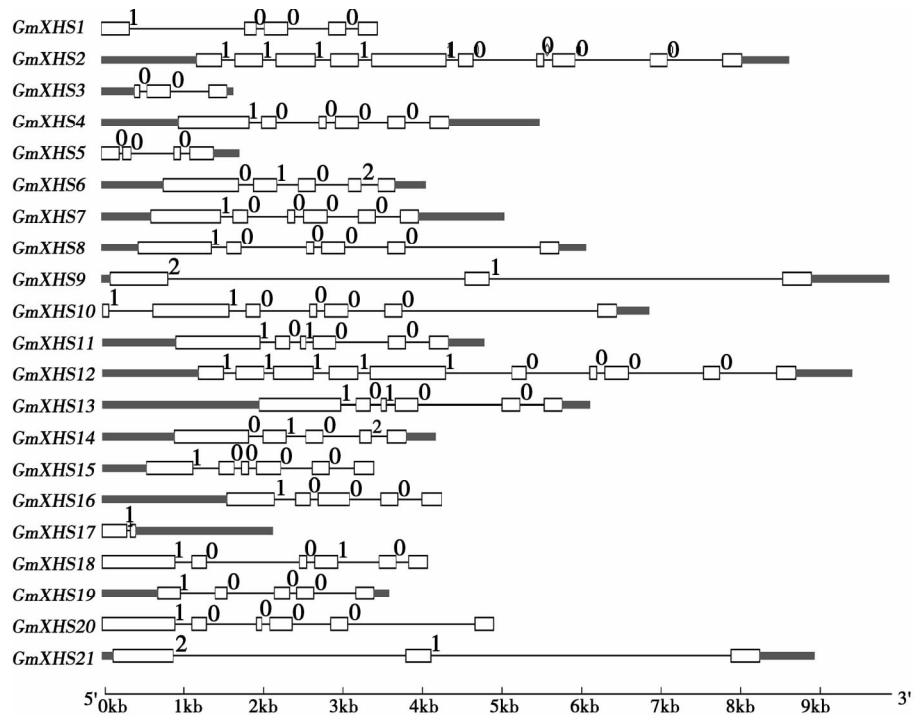


图 2 大豆中 *XHS* 蛋白系统发育树

Fig. 2 The phylogenetic tree of *XHS* proteins in soybean



内含子用直线表示;外显子用白色方框表示;非翻译区用灰色方框表示。
The introns were indicated by lines, exons were white rectangles and untranslated region were gray

图 3 大豆 *XHS* 家族基因内含子外显子结构

Fig. 3 Intron-exon structures of *XHS* genes in soybean rectangles

2.5 大豆 *XHS* 基因家族的表达部位分析

大豆 *XHS* 基因 EST 表达分析表明(表 2),在大豆 21 个 *XHS* 基因中只有 10 个找到了 EST 证据。其中 *XHS3* 基因获得的表达信息较丰富,主要在花、叶片、茎、种皮、胚中表达;*XHS5* 和 *XHS7* 均在子叶、叶片、根、胚、果荚中表达; *XHS8* 只在种皮和胚中表达, *XHS9* 只在花和叶片中表达, *XHS21* 只在叶片和

根中表达,这 3 个基因在组织中表达信息相对较少;从表 2 还可看出,除了 *XHS8*,其余 8 个均在叶片中表达,除了 *XHS9* 和 *XHS21* 之外,其余均在胚中表达。未找到 EST 表达证据的基因可能是因为大豆数据库不全,或者是因为进化过程中部分 *XHS* 基因沉默。

表 2 大豆 *XHS* 家庭基因在不同组织中的 EST 表达分析

Table 2 EST expression analysis of soybean *XHS* genes in different tissues

| 组织 | 子叶 | 花 | 叶片 | 茎 | 种皮 | 分生组织 | 叶芽 | 根 | 胚 | 果荚 |
|--------------|-----------|--------|------|------|-----------|----------|----------------|------|----------------|-----|
| Tissue | Cotyledon | Flower | Leaf | Stem | Seed coat | Meristem | Vegetative bud | Root | Somatic embryo | Pod |
| <i>XHS3</i> | | 59 | 43 | 47 | 90 | | | | 148 | |
| <i>XHS4</i> | | | 65 | | | | | 21 | 148 | 456 |
| <i>XHS5</i> | 32 | | 65 | | | | | 21 | 148 | 456 |
| <i>XHS7</i> | 32 | | 65 | | | | | 21 | 148 | 456 |
| <i>XHS8</i> | | | | | 90 | | | | 74 | |
| <i>XHS9</i> | | 59 | 21 | | | | | | | |
| <i>XHS10</i> | | 59 | 43 | 47 | 90 | | | | 148 | |
| <i>XHS17</i> | | | 65 | | | | | 21 | 148 | 456 |
| <i>XHS20</i> | | 59 | 43 | 47 | 90 | | | | 148 | |
| <i>XHS21</i> | | | 21 | | | | | 21 | | |

表达值单位是每百万转录本中基因的 EST 数。
Expression unit is EST transcripts per million.

3 结论与讨论

随着基因组研究领域的不断深入,通过对植物家族基因的分析,探寻潜在的基因功能,为各个基因家族的功能分析提供理论基础,已成为当今基因研究的热点。*XHS* 基因只存在于植物中,在其它物种中没有发现同源蛋白。在已测序的模式物种中,拟南芥中发现有 14 个 *XHS* 家族基因,水稻中有 11 个。目前对大豆 *XHS* 蛋白基因的研究相对较少,本研究以生物信息学研究方法,通过对大豆 *XHS* 全基因组扫描分析鉴定,共获得 21 个 *XHS* 家族基因,分属于 4 类,分别分布于 12 条染色体上。大豆相关数据库中 *XHS* 基因,数目多于拟南芥、水稻,约为拟南芥和水稻的 2 倍,这反映出大豆作为古四倍体的特征^[14-15]。目前推测大豆基因组的复制发生了 2 次或以上,一次可能发生在 5 900 万年前,一次可能发生在 1 300 万年前,最终导致了大豆基因组的高度重复,而对于拟南芥和水稻仅仅发生过一次全基因组复制。也可能因为在植物基因组发生复制后的进化过程中,大豆 *XHS* 家族基因大多数被保留下来。大豆基因组的 2 次复制又表现出基因的多样化和基因缺失,很多染色体进行重排,根据大豆的全基因组测序,大约有 75% 的基因是多拷贝形式^[9, 16]。对大豆 *XHS* 基因构建系统发育树,将大豆 *XHS* 基因分为 4 类。EST 表达分析发现大豆 *XHS* 基因的表达大多具有特异性。

拟南芥中对 *SGS3* 的结构和功能研究得较详细,*SGS3* 通过转录后基因沉默机制介导了植物的抗病毒能力^[2, 17],与其它蛋白共同调控了 siRNA^[18-19]。而对于大豆中 *XHS* 蛋白是否是和 *SGS3* 有相似的功能,或是可能的转录因子,尚需进一步试验验证。水稻苗期逆境表达研究中表明 *XHS* 基因起了较为重要的作用^[1],而对于大豆 *XHS* 基因如何参与逆境反应也将成为今后研究的重点。以大豆全基因组序列为背景,对大豆 *XHS* 基因家族进行结构和表达分析,将有助于揭示不同物种中 *XHS* 基因的起源和生物进化关系,也为进一步深入了解大豆 *XHS* 基因功能奠定基础。

参考文献

[1] Qin Y H, Ye H Y, Tang N, et al. Systematic identification of X1-homologous genes reveals a family involved in stress responses in rice[J]. Plant Molecular Biology, 2009, 71(4-5): 483-496.

[2] Bateman A. The SGS3 protein involved in PTGS finds a family[J]. BMC Bioinformatics, 2002, 3: 21.

[3] Rose A, Schraegle S J, Stahlberg E A, et al. Coiled-coil protein composition of 22 proteomes-differences and common themes in subcellular infrastructure and traffic control[J]. BMC Evolution-

ary Biology, 2005, 5: 66.

[4] Mourrain P, Beclin C, Elmayan T, et al. *Arabidopsis* SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance[J]. Cell, 2000, 101(5): 533-542.

[5] Fukunaga R, Doudna J N A. dsRNA with 5' overhangs contributes to endogenous and antiviral RNA silencing pathways in plants[J]. Embo Journal, 2009, 28(5): 545-555.

[6] Chen M, Bennetzen J L. Sequence composition and organization in the Sh2/A1-homologous region of rice[J]. Plant Molecular Biology, 1996, 32(6): 999-1001.

[7] Zhou J X, Wang X F, Jiao Y L, et al. Global genome expression analysis of rice in response to drought and high-salinity stresses in shoot, flag leaf, and panicle[J]. Plant Molecular Biology, 2007, 63(5): 591-608.

[8] Huang Y, Zhang L D, Zhang J W, et al. Heterosis and polymorphisms of gene expression in an elite rice hybrid as revealed by a microarray analysis of 9198 unique ESTs[J]. Plant Molecular Biology, 2006, 62(4-5): 579-591.

[9] Schmutz J, Cannon S B, Schlueter J, et al. Genome sequence of the palaeopolyploid soybean [J]. Nature, 2010, 465(7294): 120.

[10] 郭安源, 朱其惠, 陈新, 等. GSDS: 基因结构显示系统[J]. 遗传, 2007, 29(8): 1023-1026. (Guo A Y, Zhu Q H, Chen X, et al. GSDS: a gene structure display server[J]. Hereditas, 2007, 29(8): 1023-1026.)

[11] Finn R D, Mistry J, Schuster-Bockler B, et al. Pfam: clans, web tools and services [J]. Nucleic Acids Research, 2006, 34: 247-251.

[12] 刘仁虎, 孟金陵. MapDraw, 在 Excel 中绘制遗传连锁图的宏[J]. 遗传, 2003, 25(3): 317-321. (Liu R H, Meng J L. MapDraw: a microsoft excel macro for drawing genetic linkage maps based on given genetic linkage data[J]. Hereditas, 2003, 25(3): 317-321.)

[13] Tamura K, Peterson D, Peterson N, et al. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods[J]. Molecular Biology and Evolution, 2011, 28(10): 2731-2739.

[14] Mukherjee K, Brocchieri L. Burglin T R. A comprehensive classification and evolutionary analysis of plant homeobox genes[J]. Molecular Biology and Evolution, 2009, 26(12): 2775-2794.

[15] Hay A, Tsiantis M. KNOX genes: versatile regulators of plant development and diversity[J]. Development, 2010, 137(19): 3153-3165.

[16] Severin A J, Cannon S B, Graham M M, et al. Changes in twelve homoeologous genomic regions in soybean following three rounds of polyploidy[J]. Plant Cell, 2011, 23(9): 3129-3136.

[17] Beclin C, Boutet S, Waterhouse P, et al. A branched pathway for transgene-induced RNA silencing in plants[J]. Current Biology, 2002, 12(8): 684-688.

[18] Peragine A, Yoshikawa M, Wu G, et al. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis* [J]. Genes & Development, 2004, 18(19): 2368-2379.

[19] Boutet S, Vazquez F, Liu J, et al. *Arabidopsis* HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance[J]. Current Biology, 2003, 13(10): 843-848.