# Genome-wide analysis of the *GRAS* gene family in physic nut (*Jatropha curcas* L.)

**Z.Y. Wu[1,2], P.Z. Wu[1], Y.P. Chen[1], M.R. Li[1], G.J. Wu[1] and H.W. Jiang[1]**

[1]Key Laboratory of Plant Resources Conservation and Sustainable Utilization,
South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China
[2]University of Chinese Academy of Sciences, Beijing, China

Corresponding author: H.W. Jiang
E-mail: hwjiang@scbg.ac.cn

**ABSTRACT.** GRAS proteins play vital roles in plant growth and development. Physic nut (*Jatropha curcas* L.) was found to have a total of 48 GRAS family members (JcGRAS), 15 more than those found in *Arabidopsis*. The *JcGRAS* genes were divided into 12 subfamilies or 15 ancient monophyletic lineages based on the phylogenetic analysis of GRAS proteins from both flowering and lower plants. The functions of *GRAS* genes in 9 subfamilies have been reported previously for several plants, while the genes in the remaining 3 subfamilies were of unknown function; we named the latter families U1 to U3. No member of U3 subfamily is present in *Arabidopsis* and Poaceae species according to public genome sequence data. In comparison with the number of *GRAS* genes in *Arabidopsis*, more were detected in physic nut, resulting from the retention of many ancient GRAS subfamilies and the formation of tandem repeats during evolution. No evidence of recent duplication among *JcGRAS* genes was observed in physic nut. Based on digital gene expression data, 21 of the 48 genes exhibited differential expression in four tissues analyzed. Two members of subfamily U3 were expressed only in buds and flowers, implying that they

may play specific roles. Our results provide valuable resources for future studies on the functions of GRAS proteins in physic nut.

**Key words:** GRAS; Phylogeny; Expression profile; Physic nut; *Jatropha curcas* L.

## INTRODUCTION

The GRAS proteins are a plant-specific family of putative transcription factors, members of which share a conserved domain at the C-terminal end. The name GRAS is derived from the first three members of the family to be functionally characterized (Pysh et al., 1999). The *Arabidopsis* SCARECROW protein (SCR, Q9M384), the first member isolated, was identified as the result of a screen for mutations affecting root growth and development (Benfey et al., 1993). The *Arabidopsis* GAI (GIBBERELLIN-ACID INSENSITIVE) and RGA (REPRESSOR of *gai1-3*) proteins are involved in a signaling pathway that negatively regulates responses of gibberellin (Peng et al., 1997; Silverstone et al., 1998). A total of 33 and 55 GRAS family members have been identified in *Arabidopsis* and rice, respectively. They usually consist of 400-700-amino acid residues, and commonly include five conserved motifs, LHRI, VHIID, LHRII, PFYRE, and SAW, at the carboxyl terminus, while the variable regions mainly lie towards the amino terminus (Hirsch and Oldroyd, 2009). On the basis of sequence differences and lengths of the amino terminus, the GRAS proteins were initially divided into eight groups, namely LISCL (LILIUM LONGIFLORUM SCARECROW-LIKE), DELLA (The DELLA motif containing protein), HAM (HAIRY MERISTEM), PAT1 (PHYTOCHROME A SIGNAL TRANSDUCTION 1), LS (LATERAL SUPPRESSOR), SHR (SHORT-ROOT), and SCR and SCL3 (SCARECROW-LIKE 3) (Tian et al., 2004). Tong et al. (2009) subsequently identified a new member of the GRAS family, DLT (DWARF AND LOW-TILLERING) in rice (Os06g03710), which is not a member of any of the original eight groups; this DLT protein was shown to play positive roles in brassinosteroid signaling.

The GRAS proteins have diverse and important functions in plant growth and development. DELLA proteins have been shown to be involved in gibberellin signal transduction, acting as negative regulators of plant growth; they include GAI, RGA and RGL1/2/3 in *Arabidopsis* (Peng et al., 1997; Silverstone et al., 1998), OsSLR1 in rice (Sato et al., 2014) and HvSLN1 in barley (Fu et al., 2002). SCR and SHR proteins participate in root radial patterning and regulate root development (Benfey et al., 1993; Helariutta et al., 2000). LS proteins, such as LAS in *Arabidopsis* and MOC1 in rice regulate the formation of axillary meristems and tillers (Greb et al., 2003; Li et al., 2003). The *Arabidopsis* GRAS protein PAT1 has been reported to be involved in an early stage of the phytochrome-A signal transduction cascade (Bolle et al., 2000). LISCL in lily (*Lilium longiflorum*) participates in regulation of microsporogenesis in the anther by binding to a meiosis-associated promoter (Morohashi et al., 2003).

Physic nut (*Jatropha curcas* L.) is a plant which has multiple uses for bio-energy, which belongs to the Euphorbiaceae family (King et al., 2009). Recently, the genome of this species has been sequenced and expressed sequence tags (ESTs) libraries have been constructed by our group and others (Sato et al., 2011; Wang et al., 2011; King et al., 2013); these resources are helpful for the analysis of different gene families in *J. curcas* and their evolution. A total of 58 *WRKY* genes (*JcWRKYs*) were identified in the physic nut genome, and no recent gene duplication was detected when a phylogenetic comparison was made between *JcWRKY* genes and those of *Arabidopsis*, rice and castor bean (Xiong et al., 2013). In the present study, we

identified *JcGRAS* genes from physic nut genome sequences and determined their chromosomal distribution. The exon-intron arrangement and conserved domains and motifs of each gene were also characterized. The study also includes a phylogenetic analysis of GRAS proteins based on the complete set of GRAS proteins from the flowering plants physic nut, castor bean, *Arabidopsis* and rice, the lycophyte *Selaginella moellendorffii*, and the bryophyte *Physcomitrella patens*. Finally, we analyzed the patterns of expression of *JcGRAS* genes in different tissues. This work presents a detailed examination of the *GRAS* gene family at the genome level and provides a basis for further investigation of the functions of *GRAS* genes in physic nut.

## MATERIAL AND METHODS

### Sequence retrieval and identification

Sequences of *Arabidopsis* GRAS proteins were retrieved from the *Arabidopsis* genome database, TAIR 9.0 (http://www.arabidopsis.org/) and used as queries in BLAST searches against the physic nut genome database of the KaZuSa DNA Research Institute (http://www.kazusa.or.jp/jatropha/) (Sato et al., 2011) and our own genome database which is available from DDBJ/EMBL/GenBank under the accession No. AFEW00000000 (the version described in this paper is the first version, AFEW01000000). Sequences were selected for further analysis if the E value was less than 1e$^{-10}$. Next, we corrected errors in the annotation of GRAS coding domain sequences using information from the physic nut EST database available from GenBank (accession number SRX750579-SRX750581) and from our own EST datasets from physic nut and *Jatropha integerri*ma L. (accession numbers SRX750578, SRX757230, SRX757232, and SRX757234). The exon-intron structures of *JcGRAS* genes were determined by comparing the coding sequences and the corresponding genomic sequences in the Gene Structure Display Server (GSDS, http://gsds.cbi.pku.edu.cn/).

### Phylogenetic analysis of the *GRAS* proteins

Sequences of the rice GRAS domain proteins were downloaded from the rice genome annotation database (http://rice.plantbiology.msu.edu/, release 5.0). Castor bean (*Ricinus communis* L.), poplar (*Populus trichocarpa*), grape (*Vitis vinifera*), soybean (*Glycine max*), *Medicago truncatula*, *Physcomitrella patens*, and *Selaginella moellendorffii* GRAS protein sequences were downloaded from GenBank or Phytozome (http://www.phytozome.net). The conserved domains of GRAS proteins were predicted at http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi (Marchler-Bauer et al., 2015). Analysis of conserved motifs was carried out using MEME (http://meme.nbcr.net/meme/) (Bailey et al., 2009). Multiple sequence alignments of the conserved protein sequences were performed using Clustal X 1.83 (Jeanmougin et al., 1998). Unrooted phylogenetic trees were constructed with MEGA 5.0 using the maximum likelihood (ML) and neighbor-joining (NJ) methods and bootstrap tests were carried out on the basis of 1000 iterations (Tamura et al., 2011). Chromosome localization was performed using MapChart 2.1 (Voorrips, 2002).

### Preparation of plant materials

Seeds of the inbred physic nut cultivar GZQX0401 were sterilized with KMnO$_4$ solutions (1/5000) for 30 min, leaved in distilled water for 12 h, and then planted in sand. When cotyledons of

the germinated seeds were fully expanded, seedlings were transferred to a 3:1 mixture of sand and soil in a greenhouse (30-35°C) in Guangzhou (113.3°E, 23.1°N) illuminated with natural sunlight (100-300 µmol·m⁻²·s⁻¹). After emergence of the first true leaf, the trays were irrigated with 1 L Hoagland nutrient solution (pH 6.0) once every two days at dusk. Samples of the roots, stem cortex, and leaves collected at the six-leaf stage and seed samples, were frozen immediately in liquid nitrogen and stored at -80°C prior to digital gene expression analysis. Samples of undifferentiated, male, and female physic nut flowers were collected for RNA isolation. Roots, stems, flowers, and leaves of *J. integerrima* L. for RNA-seq analysis were collected from 1-year-old seedlings propagated from cuttings (Zhang et al., 2014).

## RNA isolation and RT-and qRT-PCR

Total RNA of roots was extracted using the CTAB method with some modifications (Chang et al., 1993). Briefly, the CTAB extraction buffer was spermidine-free, and 1/4th volume 5 M KAc, pH 4.8, was added before the first centrifugation (4°C, 12,000 *g*, 10 min); 1/4th volume 8 M LiCl was used for precipitation and the pellet was dissolved in guanidine thiocyanate solution [2.5 g guanidine thiocyanate in 6.6 mL CBS solution (42 mM sodium citrate and 8.38 g/L N-lauroylsarcosine sodium)] instead of SSTE buffer (1.0 M NaCl, 0.5% SDS, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA).

Subsequently, first-strand cDNA was synthesized from 2 µg total RNA treated with RNase-Free DNase I (Recombinant DNase I (RNase-free), 2270A, Japan), using M-MLV reverse transcriptase (Promega, M1701, USA) according to the manufacturer instructions. Using a physic nut *Actin* gene as an internal control, semi-quantitative RT-PCR and real-time fluorescent quantitative PCR (qRT-PCR) analysis were performed. The specific primer sequences and the PCR conditions are listed in **Table S1**. A 10-µL sample of each PCR product was analyzed by electrophoresis on a 1.5% agarose gel containing ethidium bromide for RT-PCR analysis. All qRT-PCRs were run on a LightCycler®480 Real-Time PCR System (Roche). Each 20-µL reaction volume included 10 µL 2X SYBR Premix Ex Taq, 0.4 µL forward and reverse primer (10 µM), 2 µL cDNA dilution, and 7.2 µL double-distilled water. The thermal profile used for all PCRs was as follows: 10 min at 95°C for DNA polymerase activation and 40 cycles each consisting of 15 s at 95°C and 35 s at 60°C. The expression levels were calculated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). Each PCR was performed in triplicate for three independent biological repeats.

## RESULTS

## Identification of GRAS family members in physic nut

On the basis of the conserved domain sequences present in the GRAS proteins of *Arabidopsis* and rice, a total of 48 GRAS family members were identified from the publicly-available physic nut genome database (Sato et al., 2011) and our own physic nut genome sequence (available from DDBJ/EMBL/GenBank under the accession No. AFEW00000000). We corrected errors in the annotations for several GRAS coding sequences using information from the physic nut EST database available from GenBank and from our own EST library. The predicted GRAS proteins of physic nut (JcGRAS) range from 379- to 840-amino acid (aa) residues in length, with an average of 566 aa, and all of them contain the complete GRAS domain (pfam03514; cl15987) (**Table S2**). To analyze the features of physic nut GRAS protein sequences in more detail, conserved motifs were identified (**Figure S1**) and sequence logos (**Figure S2**) were created for these motifs using

the MEME suite (http://meme.nbcr.net) in both cases, as described previously (Song et al., 2014). Most JcGRAS proteins contain 10 conserved motifs. Motif 10 was not found in the JcGRAS28 and JcGRAS29 sequences, and it was translocated in JcGRAS13, JcGRAS14, JcGRAS27, JcGRAS30, and JcGRAS31. Motifs 3 and 7 were absent from JcGRAS45 and JcGRAS11, respectively (**Figure S1**), and a summary of their features are given in **Table S2**. According to intron-exon analysis, only two *JcGRAS* genes (*JcGRAS27* and *JcGRAS43*) have one intron within the coding domain (**Table S1**).

The positions of 46 of the predicted 48 *JcGRAS* genes were mapped onto the pseudo-chromosomes of the physic nut genetic map previously constructed (King et al., 2013; Wu et al., 2015). The distribution and density of the *GRAS* genes on chromosomes are not uniform (Figure 1). It was found that 10 (20.8%) of the physic nut *GRAS* genes are present as tandem repeats at two chromosomal loci (T1 and T2) (Figure 1). Tandem duplicated genes are defined as gene pairs that are located within 50 kb from each other or are separated by ≤4 non-homologous spacer genes (Cannon et al., 2004). One locus (T1) on chromosome 1 contains four tandem repeated genes (*JcGRAS38*, *JcGRAS40*, *JcGRAS42*, and *JcGRAS43*); the other locus (T2) on chromosome 8 contains six tandem repeated genes (*JcGRAS01-06*) ().
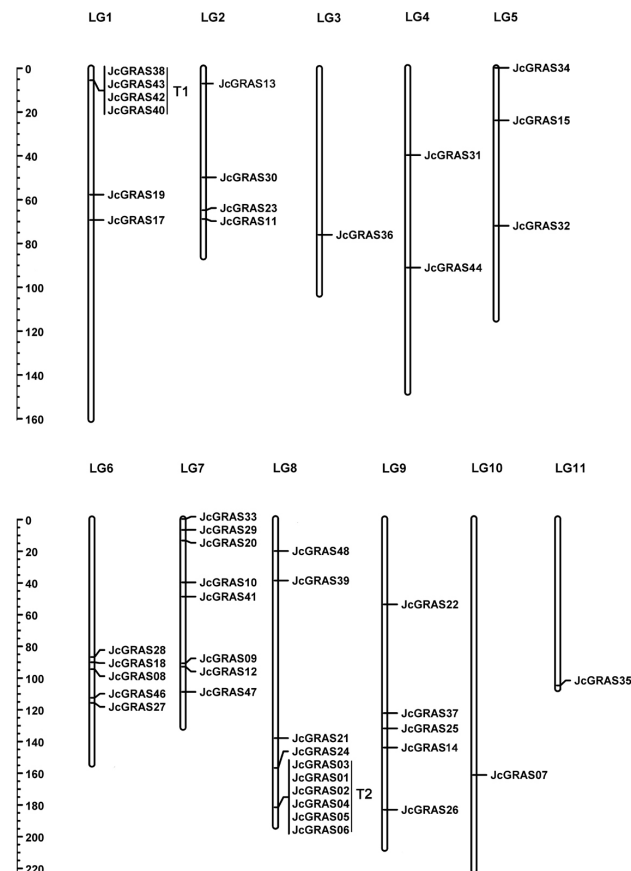


**Figure 1.** Chromosomal localization of physic nut *GRAS* genes. Chromosomal localization of physic nut *GRAS* genes on the linkage map was carried out using MapChart 2.1. Of the 48 predicted *JcGRAS* genes, 46 were mapped to the 11 linkage groups (LGs). The scale is in centimorgans. T, tandem duplication.

## Phylogenetic analysis of the *GRAS* proteins

To examine the phylogenetic relationships between the physic nut GRAS proteins and those of other plants, unrooted trees were constructed from alignments of the conserved GRAS domain sequences using MEGA 5.0 (Figures 2, **Figure S3**, and **Figure S4**).

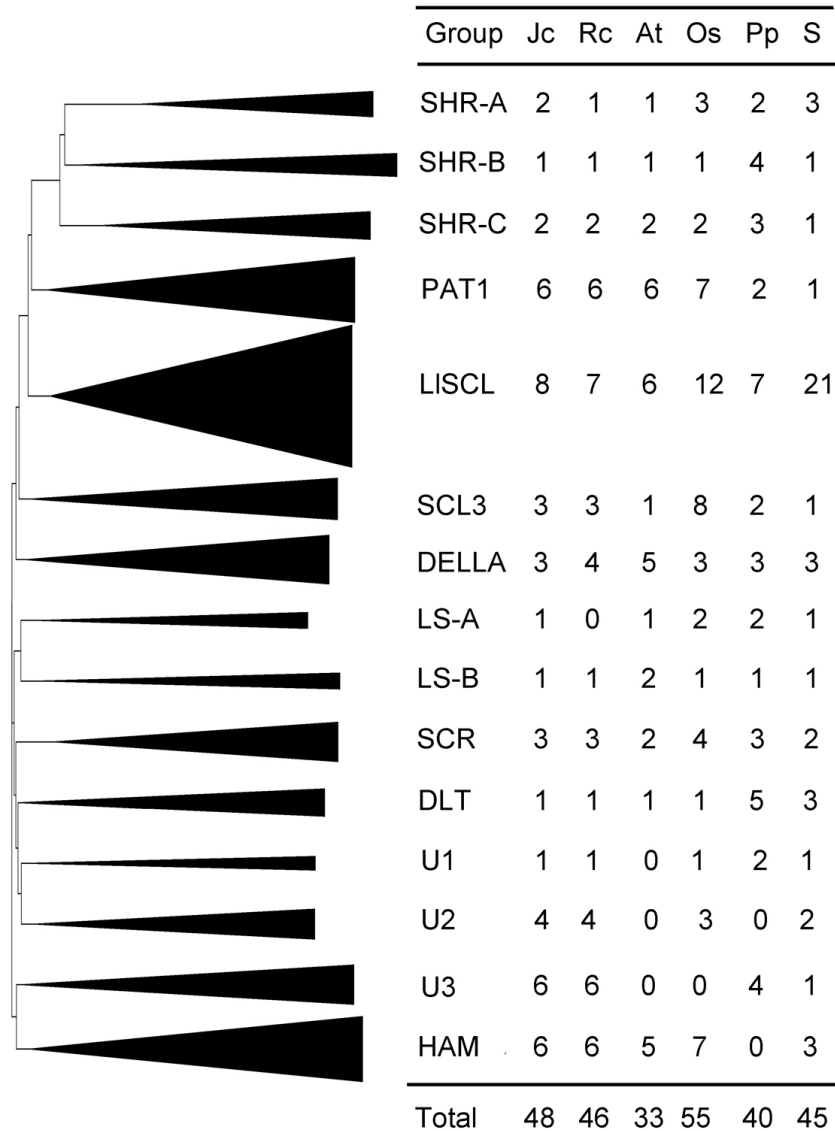| Group | Jc | Rc | At | Os | Pp | S |
|---|---|---|---|---|---|---|
| SHR-A | 2 | 1 | 1 | 3 | 2 | 3 |
| SHR-B | 1 | 1 | 1 | 1 | 4 | 1 |
| SHR-C | 2 | 2 | 2 | 2 | 3 | 1 |
| PAT1 | 6 | 6 | 6 | 7 | 2 | 1 |
| LISCL | 8 | 7 | 6 | 12 | 7 | 21 |
| SCL3 | 3 | 3 | 1 | 8 | 2 | 1 |
| DELLA | 3 | 4 | 5 | 3 | 3 | 3 |
| LS-A | 1 | 0 | 1 | 2 | 2 | 1 |
| LS-B | 1 | 1 | 2 | 1 | 1 | 1 |
| SCR | 3 | 3 | 2 | 4 | 3 | 2 |
| DLT | 1 | 1 | 1 | 1 | 5 | 3 |
| U1 | 1 | 1 | 0 | 1 | 2 | 1 |
| U2 | 4 | 4 | 0 | 3 | 0 | 2 |
| U3 | 6 | 6 | 0 | 0 | 4 | 1 |
| HAM | 6 | 6 | 5 | 7 | 0 | 3 |
| Total | 48 | 46 | 33 | 55 | 40 | 45 |

**Figure 2.** Compressed phylogenetic tree of GRAS proteins from physic nut, castor bean, *Arabidopsis*, rice, *Physcomitrella* and *Selaginella*. Amino acid sequences of the GRAS protein domains were aligned using Clustal X 1.83 and the compressed phylogenetic tree was constructed by MEGA 5.0 using the neighbor-joining method. At, *Arabidopsis*; Os, rice; Rc, castor bean; Jc, physic nut; Pp, *Physcomitrella patens*; S, *Selaginella moellendorffii* (GRAS protein sequence alignments are shown in **Text S1**).

GRAS proteins used for the analysis include 48 members of the family from physic nut (dicotyledon, Euphorbiaceae), 46 from castor bean (dicotyledon, Euphorbiaceae) (http://www.phytozome.net), 33 from *Arabidopsis* (dicotyledon) (Tian et al., 2004), 55 from rice (monocotyledon) (Tian et al., 2004), 40 from *Physcomitrella patens* (moss) (Engstrom, 2011) (http://www.phytozome.net) and 45 from *Selaginella moellendorffii* (lycophyte) (Engstrom, 2011) (http://www.phytozome.net) (**Table S3**). According to the tree topologies and the bootstrap values produced by the Neighbor-Joining algorithm, these GRAS proteins could be grouped into 12 subfamilies or a minimum of 15 discrete monophyletic lineages that existed prior to divergence between the lower plant and flowering plant lineages (Figure 2, **Figure S3**). The subfamilies include the eight groups previously reported by Tian et al. (2004): SHR, PAT1, LISCL, SCL3, DELLA, LS, SCR, and HAM. Subfamilies DLT, and unknown function (U) subfamilies U1, U2, and U3, are well-supported monophyletic groups that are clearly separated from the eight GRAS subfamilies originally recognized by Tian et al. (2004).

There are five *JcGRAS* genes in the SHR subfamily (Figure 2, **Table S3**). The SHR proteins are diversified into three monophyletic lineages, SHR-A (SHR), SHR-B (NSP1), and SHR-C. *Arabidopsis* SHORT-ROOT (SHR, Q9SZF7) acts upstream of SCR and participates in a radial signaling pathway (Helariutta et al., 2000). *Medicago* NSP1 functions as a primary transcriptional regulator mediating all known Nod factor-induced transcriptional responses and thus regulating the rhizobium-legume symbiosis (Smit et al., 2005). The function of the SHR-C proteins is as yet unknown.

Physic nut has six *GRAS* genes in the PAT1 subfamily (Figure 2, **Table S3**). Flowering plants have more *PAT1* genes than *Physcomitrella* and *Selaginella*. In this subfamily, the *Arabidopsis* proteins AtPAT1 (Q9LDL7), AtGRAS-24 (SCL13, Q9M0M5), and AtGRAS-11 (SCL21, Q9S7H5) were reported to be involved in the phytochrome A signal transduction pathway where they act as early signaling components by regulating the red and far-red light-absorbing forms of phytochrome (Bolle et al., 2000; Torres-Galea et al., 2006; Torres-Galea et al., 2013).

There are eight *JcGRAS* genes in the LISCL subfamily (Figure 2, **Table S3**). The LISCL subfamily is the largest subfamily in the *GRAS* gene family. The expansion of the *LISCL* genes probably occurred independently among different plant taxa because genes from within a given plant class are clustered into a group. In physic nut, six tandem repeats of *LISCL* genes are observed on chromosome 8, while there are five on scaffold 28650 in castor bean (Figure 2). *LISCL* was reported to participate in transcriptional regulation during microsporogenesis within the lily anther by enhancing the activity of a meiosis-associated promoter (Morohashi et al., 2003).

Three *JcGRAS* genes are classified into in the SCL3 subfamily (Figure 2, **Table S3**). *Arabidopsis* has only one SCL3 (Q9LPR8) gene, which is a direct target of DELLA protein and functions as a positive regulator of GA signaling to modulate plant growth (Zhang et al., 2011).

DELLA proteins are mainly involved in the gibberellin signal transduction pathway that regulates many aspects of plant growth and development. There are three *JcGRAS* genes in DELLA subfamily, which is one and two fewer than in castor bean and *Arabidopsis*, respectively (Figure 2, **Table S3**). *Arabidopsis* GAI, RGA, and RGL1/2/3, members of this subfamily, play vital roles in plant growth as negative regulators of gibberellin responses (Peng et al., 1997; Silverstone et al., 1998; Wen and Chang, 2002). The rice DELLA protein SLR1 (Os03g49990) has been shown to interact with the GA receptor GID1 (GIBBERELLIN INSENSITIVE DWARF1) and thus participate in the GA signaling pathway (Sato et al., 2014).

Physic nut has two *GRAS* genes in the LS subfamily (Figure 2, **Table S3**). The LS proteins could be classed into two monophyletic lineages, LS-A/LAS and LS-B. The CDS sequence of 28966.m000535 represents the LS-A gene in castor bean, but it contains only a partial *GRAS* gene sequence because of the incompleteness of the genome sequence (Chan et al., 2010). The

LAS (Q9ZWC5) protein regulates the initiation of axillary meristems during the vegetative phase of development in *Arabidopsis* (Greb et al., 2003). Rice MONOCULM 1 (MOC1, Os06g40780) is expressed mainly in the axillary buds and functions to initiate axillary buds and to promote their outgrowth (Li et al., 2003).

Three *JcGRAS* genes are allocated to the SCR subfamily (Figure 2, **Table S3**). The *Arabidopsis* SCR protein encoded by locus AT3G54220 regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root (Benfey et al., 1993; Di Laurenzio et al., 1996).

The HAM gene was first identified in petunia as a component of a novel signaling pathway promoting shoot indeterminacy (Stuurman et al., 2002). There are six *JcGRAS* genes allocated to this subfamily (Figure 2, **Table S3**). *Medicago* NSP2 is a member of the HAM subfamily. It forms a complex with NSP1 and associates with the promoters of the early nodulin gene *ENOD11* to regulate the nodulation signaling pathway (Hirsch et al., 2009). LOST MERISTEMS 1 (LOM1, AtGRAS-14/Q7XJM8) and LOM2 (AtGRAS-21/Q9M000) are involved in the cellular differentiation of central zone descendants in *Arabidopsis* shoot meristems (Schulze et al., 2010).

Physic nut has one GRAS member in the DLT subfamily (Figure 2, **Table S3**). According to the phylogenetic tree of *GRAS* genes constructed by Tian et al. (2004), AtGRAS-8 (Q9CAN3) and DLT (Os06g03710) also clustered into a group. OsBZR1 was reported to bind to the *DLT* promoter through the brassinosteroid (BR)-response element, suggesting that the DLT protein plays vital roles in BR signaling in rice (Tong et al., 2009).

For the other three GRAS subfamilies, no function has previously been reported. In this study, we named these subfamilies U1 to U3 (Figure 2, **Table S3**). Subfamily U1 includes one *JcGRAS* gene, *JcGRAS31*. Subfamily U2 includes four *JcGRAS* genes, similar to those seen in castor bean. No *Arabidopsis* genes were grouped into the U1 and U2 subfamilies. Subfamily U3 contains six *JcGRAS* genes. Four tandem repeats of U3 genes were observed on chromosome 1 in physic nut, while there are four on scaffold 29889 in castor bean. Based on the genome sequences available from GenBank and other databases, no U3 genes were detected in *Arabidopsis*, or in Poaceae species such as rice, maize, and sorghum. However, they are present in many dicotyledons, such as *Vitis vinifera* (*CAN81356*, *CBI22895*, *XP_002270640*), *Populus trichocarpa* (*XP_002329212*, *XP_002316034*, *XP_002316033*, *XP_002316032*, *XP_002316031*, *XP_002316030*, *XP_002302767*, and *XP_002320295*), *Glycine max* (*XP_003546321*, *XP_003534817*, *XP_003536868*, *XP_003520753*, and *XP_003553668*) and *Medicago truncatula* (*XP_003594755*, *XP_003594754*, *XP_003594757*, *XP_003591395*, and *XP_003625975*).

On the basis of the phylogenetic tree (Figure 2), no candidate recent paralogs or inparalogs (paralogs created after the separation of Euphorbiaceae species) (Sonnhammer and Koonin, 2002) were found for physic nut *GRAS* genes. The tandem repeats in the physic nut LISCL and U3 subfamilies probably resulted from gene duplication events that occurred before the divergence of the Euphorbiaceae.

## Expression of *JcGRAS* genes

We analyzed the expression of these *JcGRAS* genes under normal growth conditions in four different tissues: root, stem cortex, leaf, and seeds [early development stage (Seed 1) and filling and maturation stage (Seed 2)] (Jiang et al., 2012). ESTs for 39 *JcGRAS* genes were identified based on EST databases for physic nut and/or its related species *J. integerrima* L. According to the transcriptome expression database generated from next-generation sequencing-based digital

gene expression tags, 45 *JcGRAS* genes (93.8% of the total) were found to be expressed in at least one of the physic nut tissues analyzed (**Table S4**).

ESTs for the five SHR genes were detected in physic nut plants. The expression level of *JcGRAS12* is relatively lower in Seed 2 than in other tissues. *JcGRAS11*, which is closely-related to *NSP1*, is expressed relatively higher in roots than in the other tissues tested (Table 1, **Table S4**). ESTs for all PAT1 subfamily genes were observed. *JcGRAS15* was expressed at the highest level in stems, while *JcGRAS16* was highest in roots and most abundant in Seed 2. The expression of *JcGRAS17* was high in Seed 1 but low in roots. For the LISCL subfamily genes, ESTs were observed in physic nut and in *J. integerrima* L. The expression of several of the LISCL subfamily genes was divergent among different physic nut tissues. *JcGRAS03* and *JcGRAS05* were expressed highly in all tissues investigated.

**Table 1.** Expression of *JcGRAS* genes in different tissues based on digital gene expression tag analysis.

| Gene code | EST | Expression level** | | | | |
|---|---|---|---|---|---|---|
| | | Root | Stem | Leaf | Seed1 | Seed2 |
| JcGRAS01 | + | | | | | |
| JcGRAS02 | + | | | | | |
| JcGRAS03 | + | | | | | |
| JcGRAS05 | + | | | | | |
| JcGRAS06 | + | | | | | |
| JcGRAS07 | + | | | | | |
| JcGRAS08 | + | | | | | |
| JcGRAS09 | + | | | | | |
| JcGRAS11 | + | | | | | |
| JcGRAS12 | + | | | | | |
| JcGRAS15 | + | | | | | |
| JcGRAS16 | + | | | | | |
| JcGRAS17 | + | | | | | |
| JcGRAS18 | + | | | | | |
| JcGRAS19 | + | | | | | |
| JcGRAS20 | + | | | | | |
| JcGRAS21 | + | | | | | |
| JcGRAS22 | + | | | | | |
| JcGRAS23 | + | | | | | |
| JcGRAS26 | + | | | | | |
| JcGRAS27 | + | | | | | |
| JcGRAS28 | + | | | | | |
| JcGRAS32 | - | | | | | |
| JcGRAS34 | + | | | | | |
| JcGRAS35 | + | | | | | |
| JcGRAS36 | + | | | | | |
| JcGRAS37 | + | | | | | |
| JcGRAS44 | + | | | | | |
| JcGRAS45 | * | | | | | |
| JcGRAS47 | + | | | | | |
| JcGRAS48 | + | | | | | |
| JcGRAS14 | + | | | | | |

| -5 | 0 | 4 | 8 | NA*** |
|---|---|---|---|---|

*An EST from the ortholog of this gene was detected in *Jatropha integerrima* L. **Log (base2) of number of transcripts per million (TPM) tags ≥2.0; ***TPM not available.

The expression levels of *JcGRAS06-08* in Seed 1 were much higher than those in Seed 2, suggesting that these genes play roles in the early stages of development of physic nut seeds. The expression level of *JcGRAS02* was much higher in stem cortex and leaves than in roots and seeds. ESTs for the three SCL3 subfamily genes were also observed in physic nut plants. The

*JcGRAS23* gene was highly expressed in all tissues analyzed, while the expression pattern of *JcGRAS24* and *JcGRAS25* was not available from the database used. ESTs for the three *DELLA* genes were detected in physic nut. *JcGRAS35* was expressed at a high level in Seed 1 but at a low level in leaves. The expression level of *JcGRAS36* was much higher in Seed 1 than in Seed 2. The expression pattern of the two LS subfamily genes was different. The *JcGARS21* gene was mainly expressed in Seed 1, while the *JcGRAS22* gene was expressed in all the tissues analyzed. ESTs for two of the three SCR subfamily genes were identified in physic nut plants. The expression levels of *JcGRAS27* and *JcGRAS28* in Seed 1 were much higher than those in Seed 2. The *JcGRAS27* gene was expressed at a very low level in roots, while *JcGRAS28* gene was very low in Seed 2. The expression level of the DLT subfamily gene, *JcGRAS26*, was much higher in roots, stem cortex, and Seed 1 than in leaves and Seed 2. The ESTs for four HAM subfamily genes were detected in physic nut. The *JcGRAS14*, *JcGRAS47*, and *JcGRAS48* genes were highly expressed in all tissues tested, although the expression levels of the *JcGRAS47* and *JcGRAS48* genes were relatively lower in Seed 2. A low level of expression of the U1 subfamily gene (*JcGRAS31*) was observed in roots and Seed 1. An EST for one U2 subfamily gene, but not for U3 subfamily genes, was observed in physic nut plants. The expression levels of the U1, U2, and U3 subfamily genes were relatively low in all the tissues tested, according to the database used in this study.

　　　To assess whether the digital expression data (obtained from two repeats) could be confirmed by an alternative method, the expression levels of genes in subfamilies PAT1 and DELLA were measured by qRT-PCR (Figure 3). The results for selected genes showed general agreement with the abundance of their transcripts determined in the digital gene expression tag profiling experiments, suggesting that the digital expression data were generally accurate. No U3 subfamily genes were identified in *Arabidopsis* and rice, and no convincing expression of U3 genes was observed in any of the tested tissues in physic nut. ESTs representing two U3 genes were detected in *J. integerrima* L. flowers ([Table S4](#)). We next tested the expression of U3 genes (*JcGRAS38 - 43*) in physic nut leaves, seeds, roots, flowers, and buds using the RT-PCR method. No expression of any of these genes was detected in leaves, seeds or roots under the conditions used for this study, but the expression of *JcGRAS40* and *JcGRAS41* was observed in flowers and buds (Figure 4).
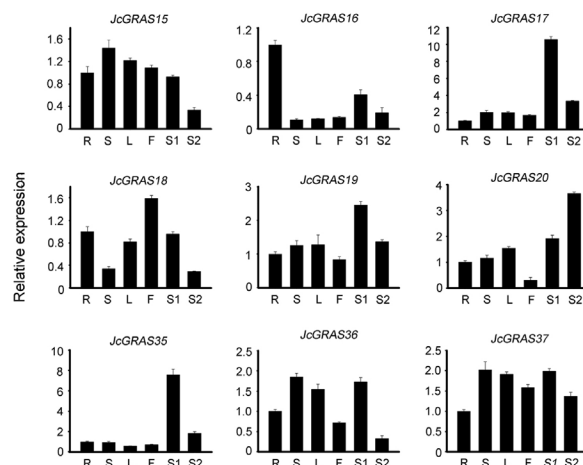


**Figure 3.** Expression analysis of selected *JcGRAS* genes using quantitative RT-PCR. qRT-PCR analysis of *JcGRAS* genes in roots (R), stem cortex (S), leaves (L), flowers (F), and seeds (S1 and S2). Relative expression was normalized to that of the reference gene JcActin as an internal control. Bars show means ± SD of three technical replicates.
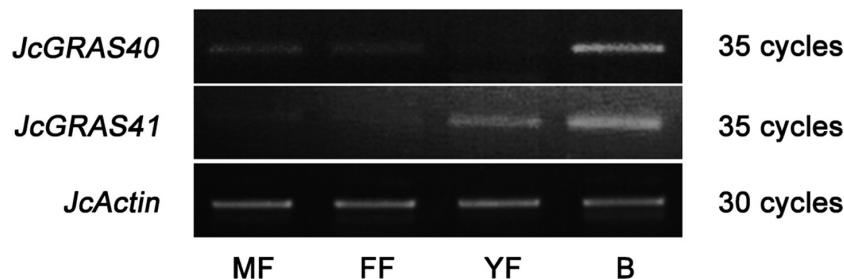
**Figure 4.** Expression analysis of physic nut group U3 genes using semi-quantitative RT-PCR. RT-PCR analysis of expression of group U3 genes, *JcGRAS40* and *JcGRAS41* in male flower (MF), female flower (FF), young (undifferentiated) flower (YF), and bud (B). *JcActin* was used as the internal control to normalize cDNA concentrations.

## DISCUSSION

*GRAS* genes encode an important plant-specific family of putative transcription factors, which have been detected in lycophytes, mosses, and flowering plants (Engstrom, 2011). In this study, a total of 48 *GRAS* genes were identified in physic nut genome sequences; this is more than the number of *GRAS* genes present in *Arabidopsis* (33), castor bean (46), *Physcomitrella* (40) and *Selaginella* (45), but fewer than those in rice (55). On the basis of conserved domains and phylogenetic analysis, the 48 genes were divided into twelve subfamilies. Subfamily DLT was named for the rice gene *DLT* (*Os06g03710*), while subfamilies U1, U2, and U3 contain genes of unknown function. In addition, the GRAS proteins could be classified into a minimum of 15 discrete monophyletic lineages dating to a point before the divergence of the moss and flowering plant lineages. Each monophyletic lineage contains proteins from both flowering plants and lower plants (Figure 2, **Figure S3**, and **Figure S4**). The SHR proteins were divided into three lineages (SHR-A, SHR-B, and SHR-C) and the LS proteins were classified into two lineages (LS-A and LS-B). These results indicated that the *GRAS* genes had undergone gene duplication and probable functional divergence during the early stages of plant evolution.

Although the LISCL proteins were the largest subfamily, their expansion within different plant groups is likely to have occurred independently, because genes from a given plant group were clustered into a clade. There are more *PAT1* genes in physic nut and other flowering plants than in *Physcomitrella patens* and *Selaginella moellendorffii*. This result suggested that expansion of the PAT1 subfamily occurred only during the evolution of flowering plants. No genes of subfamilies U1 and U2 are present in *Arabidopsis*, and no subfamily U3 genes occur in *Arabidopsis* or rice, suggesting that several *GRAS* gene subfamilies have been lost from some plant species during the course of evolution. The number of genes in several GRAS subfamilies differed between physic nut/castor bean and *Arabidopsis*, suggesting that expansion or contraction of these GRAS subfamilies occurred during the evolution of these species after the separation of dicotyledons. Tandem repeats of U3 genes (locus T1) and *LISCL* genes (locus T2) offer one explanation for the expansion of these two subfamilies of genes in both physic nut and castor bean genomes (Figure 1). Tandem repeats of *GRAS* genes were also observed in rice and *Arabidopsis* (Tian et al., 2004), supporting our view that tandem duplication was one reason for the expansion of the *GRAS* gene family. The number of *GRAS* genes in 11 of the 15 groups was the same in physic nut and castor bean, and no duplication of any *GRAS* gene in physic nut alone was identified from the phylogenetic trees. The T1 and T2 tandem repeat regions were present in both the physic nut and

the castor bean genome. These results suggested that no recent gene duplication events have affected *GRAS* genes in physic nut. Instead, the expansion of *GRAS* gene families took place before the separation of physic nut and castor bean. Thus, the larger number of *GRAS* genes in physic nut and castor bean than in *Arabidopsis* has resulted from the retention of more of the ancient GRAS subfamilies and tandem repeats in the physic nut genome during its evolution.

Differential expression levels of *GRAS* genes in the tissues investigated were observed for several subfamilies (Table 1, **Table S4**). LlSCL proteins were involved in transcriptional regulation during microsporogenesis within the lily anther (Morohashi et al., 2003). The expression levels of *JcGRAS06*, *JcGRAS07*, and *JcGRAS08* in Seed 1 were five-fold greater than those in Seed 2 (**Table S4**), which indicates that these genes probably play roles in early seed development. Of the T2 tandemly repeated genes (Figure 1), *JcGRAS03* and *JcGRAS05* were highly expressed in all tissues tested, while *JcGRAS04* was highly expressed in stem cortex and leaves. The SHR family gene *JcGRAS12* was highly expressed in roots, leaves, stem, cortex, and Seed 1 (**Table S4**). *JcGRAS11*, allocated to the SHR-B (NSP1) group, was highly expressed in roots compared to other tissues; this may be related to its function as a transcriptional regulator in root development (Helariutta et al., 2000; Smit et al., 2005). *JcGRAS23* in subfamily SCL3 and *JcGRAS14*, *JcGRAS47*, and *JcGRAS48* in subfamily HAM were highly expressed in most or all of the tissues tested, but this was not the case for the other genes in these families (**Table S4**). For the T1 tandem-repeat genes, *JcGRAS40* was the only gene that was detected, and was expressed in buds and flowers (Figure 4).These results suggested that sub functionalization of these genes occurred during the evolutionary process in physic nut.

Genes of the U1 and U2 subfamilies were found in physic nut, castor bean, and rice. The expression of several U1 and U2 genes was observed in physic nut plants, suggesting that they may play specific roles in these plants. Of the six U3 genes, the expression of two genes were detected in flowers and buds (Figure 4), indicating that they may play roles in bud and flower differentiation and/or development in physic nut. In addition, some of the U3 subfamily genes may be expressed in other tissues that were not tested in this study. Alternatively, some or all of them may be pseudogenes. In conclusion, our results provide an expanded insight into the origins and expression of the *GRAS* gene subfamilies, thereby enabling future studies on the function of these proteins in physic nut.

## ACKNOWLEDGMENTS

## Supplementary material

## REFERENCES

Bailey TL, Boden M, Buske FA, Frith M, et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acid. Res.* 37: W202-208.

Benfey PN, Linstead PJ, Roberts K, Schiefelbein JW, et al. (1993). Root development in *Arabidopsis*: four mutants with dramatically altered root morphogenesis. *Development* 119: 57-70.

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

Cannon SB, Mitra A, Baumgarten A, Young ND, et al. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* 4: 10.

Chan AP, Crabtree J, Zhao Q, Lorenzi H, et al. (2010) Draft genome sequence of the oilseed species *Ricinus communis. Nat. Biotechnol.* 28: 951-956.

Chang S, Puryear J and Cairney J (1993). A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* 11: 113-116.

Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, et al. (1996). The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell* 86: 423-433.

Fu X, Richards DE, Ait-Ali T, Hynes LW, et al. (2002). Gibberellin-mediated proteasome-dependent degradation of the barley DELLA protein SLN1 repressor. *Plant Cell* 14: 3191-3200.

Helariutta Y, Fukaki H, Wysocka-Diller J, Nakajima K, et al. (2000). The *SHORT-ROOT* gene controls radial patterning of the *Arabidopsis* root through radial signaling. *Cell* 101: 555-567.

Hirsch S and Oldroyd GED (2009). GRAS-domain transcription factors that regulate plant development. *Plant Signal. Behav.* 4: 698-700.

Hirsch S, Kim JY, Munoz A, Heckmann AB, et al. (2009). GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in *Medicago truncatula*. *Plant Cell* 21: 545-557.

Jeanmougin F, Thompson JD, Gouy M, Higgins DG, et al. (1998) Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* 23: 403-405.

Jiang HW, Wu PZ, Zhang S, Song C, et al. (2012). Global analysis of gene expression profiles in developing physic nut (*Jatropha curcas* L.) seeds. *PLoS One* 7: e36522.

King AJ, He W, Cuevas JA, Freudenberger M, et al. (2009). Potential of *Jatropha curcas* as a source of renewable oil and animal feed. *J. Exp. Bot.* 60: 2897-2905.

King AJ, Montes LR, Clarke JG, Affleck J, et al. (2013). Linkage mapping in the oilseed crop *Jatropha curcas* L. reveals a locus controlling the biosynthesis of phorbol esters which cause seed toxicity. *Plant Biotechnol. J.* 11: 986-996.

Li XY, Qian Q, Fu ZM, Wang GS, et al. (2003). Control of tillering in rice. *Nature* 422: 618-621.

Livak KJ and Schmittgen TD (2001) Analysis of relative expression data using real-time quantitative PCR and $2^{-\Delta\Delta CT}$ method. *Method* 25: 402-408.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, et al. (2015) CDD: NCBI's conserved domain databse. *Nucleic Acid. Res.* 43: D222-226.

Morohashi K, Minami M, Takase H, Hotta Y, et al. (2003). Isolation and characterization of a novel *GRAS* gene that regulates meiosis-associated gene expression. *J. Biol. Chem.* 278: 20865-20873.

Peng JR, Carol P, Richards DE, King KE, et al. (1997). The *Arabidopsis GAI* gene defines a signaling pathway that negatively regulates gibberellin responses. *Genes Dev.* 11: 3194-3205.

Pysh LD, Wysocka-Diller JW, Camilleri C, Bouchez D, et al. (1999). The *GRAS* gene family in *Arabidopsis*: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. *Plant J.* 18: 111-119.

Sato S, Hirakawa H, Isobe S, Fukai E, et al. (2011). Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. *DNA Res.* 18: 65-76.

Sato T, Mjyanoiri Y, Takeda M, Naoe Y, et al. (2014). Expression and purification of a GRAS domain of SLR1, the rice DELLA protein. *Protein Expr. Purif.* 95: 248-258.

Schulze S, Schafer BN, Parizotto EA, Voinnet O, et al. (2010). *LOST MERISTEMS* genes regulate cell differentiation of central zone descendants in *Arabidopsis* shoot meristems. *Plant J.* 64: 668-678.

Silverstone AL, Ciampaglio CN and Sun TP (1998). The *Arabidopsis RGA* gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant Cell* 10: 155-169.

Smit P, Raedts J, Portyanko V, Debelle F, et al. (2005). NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. *Science* 308: 1789-1791.

Song XM, Liu TK, Duan WK, Ma QH, et al. (2014). Genome-wide analysis of the GRAS family in Chinese cabbage (*Brassica rapa ssp. pekinensis*). *Genomics* 103: 135-146.

Sonnhammer EL and Koonin EV (2002). Orthology, paralogy and proposed classification for paralog subtypes. *Trends Genet.* 18: 619-620.

Stuurman J, Jaggi F and Kuhlemeier C (2002). Shoot meristem maintenance is controlled by a *GRAS*-gene mediated signal from differentiating cells. *Genes Dev.* 16: 2213-2218.

Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.

Tian CG, Wan P, Sun SH, Li JY, et al. (2004). Genome-wide analysis of the *GRAS* gene family in rice and *Arabidopsis*. *Plant Mol. Biol.* 54: 519-532.

---

Tong H, Jin Y, Liu W, Li F, et al. (2009). DWARF AND LOW-TILLERING, a new member of the GRAS family, plays positive roles in brassinosteroid signaling in rice. *Plant J.* 58: 803-816.

Torres-Galea P, Huang LF, Chua NH and Bolle C (2006). The GRAS protein SCL13 is a positive regulator of phytochrome-dependent red light signaling, but can also modulate phytochrome A responses. *Mol. Genet. Genomics* 276: 13-30.

Torres-Galea P, Hirtreiter B and Bolle C (2013). Two GRAS proteins, SCARECROW-LIKE 21 and PHYTOCHROME A SIGNAL TRANSDUCTION 1, function cooperatively in phytochrome A signal transduction. *Plant Physiol.* 161: 291-304.

Wang CM, Liu P, Yi C, Gu K, et al. (2011). A first generation microsatellite- and SNP-based linkage map of *Jatropha*. *PLoS One* 6: e23632.

Wen CK and Chang C (2002). *Arabidopsis RGL1* encodes a negative regulator of gibberellin responses. *Plant Cell* 14: 87-100.

Wu P, Zhou C, Cheng S, Wu Z, et al. (2015). Integrated genome sequence and linkage map of physic nut (*Jatropha curcas* L.), a biodiesel plant. *Plant J.* 81: 810-821.

Xiong W, Xu X, Zhang L, Wu P, et al. (2013). Genome-wide analysis of the WRKY gene family in physic nut (*Jatropha curcas* L.). *Gene* 524: 124-132.

Zhang L, Zhang C, Wu P, Chen Y, et al. (2014). Global analysis of gene expression profiles in physic nut (*Jatropha curcas* L.) seedlings exposed to salt stress. *PLoS One* 9: e97878.

Zhang ZL, Ogawa M, Fleet CM, Zentella R, et al. (2011). SCARECROW-LIKE 3 promotes gibberellin signaling by antagonizing master growth repressor DELLA in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 108: 2160-2165.