GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Homology-based analysis of the *GRAS* gene family in tobacco

Y.Q. Chen[1,2]*, S.S. Tai[3*], D.W. Wang[1], A.M. Ding[1], T.T. Sun[1,2], W.F. Wang[1] and Y.H. Sun[1]

[1]Key Laboratory for Tobacco Gene Resources,
Tobacco Research Institute of Chinese Academy of Agricultural Sciences,
Qingdao, China
[2]Graduate School of Chinese Academy of Agricultural Sciences, Beijing, China
[3]BGI-Shenzhen, Shenzhen, China

*These authors contributed equally to this study.
Corresponding author: Y.H. Sun
E-mail: yhsun@163.com

Genet. Mol. Res. 14 (4): 15188-15200 (2015)
Received August 20, 2015
Accepted October 19, 2015
Published November 25, 2015
DOI http://dx.doi.org/10.4238/2015.November.25.7

ABSTRACT. Members of the *GRAS* gene family are important transcriptional regulators. In this study, 21 *GRAS* genes were identified from tobacco, and were classified into eight subgroups according to the classification of *Arabidopsis thaliana*. Here, we provide a preliminary overview of this gene family in tobacco, describing the gene structure, gene expression, protein motif organization, phylogenetic analysis, and comparative analysis in tobacco, *Arabidopsis*, and rice. Using the sequences of 21 *GRAS* genes in *Arabidopsis* to search against the American tobacco genome database, 21 homologous *GRAS* genes in tobacco were identified. Sequence analysis indicates that these GRAS proteins have five conserved domains, which is consistent with their counterparts in other plants. Phylogenetic analyses divided the *GRAS* gene family into eight subgroups, each of which has distinct conserved domains and biological functions. Furthermore, the expression pattern of these 21 *GRAS* genes reveals that most are expressed in all six

tissues studied; however, some have tissue specificity. Taken together, this comprehensive analysis will provide a rich resource to assist in the study of GRAS protein functions in tobacco.

Key words: *GRAS* gene family; Gene structure; Protein motif analysis; Phylogenetic analysis; Gene expression pattern analysis; Tobacco

INTRODUCTION

Belonging to a family of plant-specific transcription factors, the name GRAS originates from the first three proteins of this family that were functionally characterized in *Arabidopsis*, *GIBBERELLIC ACID INSENSITIVE* (*GAI*), *REPRESSOR OF GAI* (*RGA*), and *SCARECROW* (*SCR*) (Di Laurenzio et al., 1996; Peng et al., 1997; Silverstone et al., 1998). GRAS proteins typically consist of 400-770 amino acid residues (Bolle, 2004). The N-terminal region of GRAS proteins is variable, whereas the C-terminal sequence is highly conserved, and contains five motifs; leucine heptad repeat I (LHRI), VHIID, leucine heptad repeat II (LHRII), PFYRE, and SAW (Pysh et al., 1999; Bolle, 2004). It is noted that the conserved domains in the C-terminus are related to transcriptional regulation; however, the N-terminus targets the downstream proteins and determines the specificity of function (Sun et al., 2011). The GRAS protein family can be divided into eight subfamilies that have been named either after their members or after a common motif: DELLA, HAM, LISCL, PAT, LS, SCR, SHR, and SCL3 subfamilies (Tian et al., 2004). Consistent with other transcriptional regulators, most GRAS proteins are nuclear-localized; however, *PAT1* and *SCL13* are localized to the cytoplasm, although *SCL13* can also be detected in the nucleus (Bolle et al., 2000; Torres-Galea et al., 2006).

The *GRAS* gene family has been studied intensely for more than a decade, and new members have been gradually identified in recent years. To date, at least 33, 60, and 48 *GRAS* genes have been identified, respectively, in the whole genome of *Arabidopsis*, rice, and Chinese cabbage, many of which have been functionally characterized (Tian et al., 2004; Itoh et al., 2005; Lee et al., 2008; Tong et al., 2009; Song et al., 2014). Furthermore, two GRAS proteins have already been described in the model legumes *Medicago truncatula* and *Lotus japonicus* (Kalo et al., 2005; Smit et al., 2005; Heckmann et al., 2006). The functions of these GRAS proteins are related to diverse processes involved in plant growth and development. These range from gibberellic acid signaling (Peng et al., 1997; Silverstone et al., 1998; Peng et al., 1999; Ikeda et al., 2001), shoot meristem maintenance (Stuurman et al., 2002), root radial patterning (Di Laurenzio et al., 1996; Helariutta et al., 2000), light signal transduction (Torres-Galea et al., 2006), male gametogenesis (Morohashi et al., 2003), axillary meristem formation (Schumacher et al., 1999; Greb et al., 2003; Li et al., 2003), and nodule morphogenesis (Kalo et al., 2005; Smit et al., 2005; Heckmann et al., 2006).

Nicotiana tabacum is one of the most widely cultivated non-food crops worldwide and is grown in about 120 countries. Given its significant economic value and distinct biological properties, tobacco has been studied as a model plant organism for fundamental biological processes (Zhang et al., 2011). However, the GRAS proteins of tobacco are poorly characterized despite the important roles of these proteins in plant growth regulation. So far, only *NtGRAS1* (GenBank No. DQ449940.1) and *NtLS* (GenBank No. AF098674.1) have been characterized in tobacco (Czikkel and Maxweel, 2007; Tai SS, Chen YQ, Wang L, Wang WF, et al., unpublished results). Thus, the investigation of *GRAS* genes in tobacco is timely. In this

study, we report attempts to isolate *GRAS* genes from tobacco and provide an initial evaluation of the tobacco *GRAS* genes. Using the homology-based method, we isolated 21 homologous *GRAS* genes in tobacco. We aligned the GRAS proteins of tobacco with those of other plants, performed a phylogenetic analysis, and calculated the value of nonsynonymous (K_A)/synonymous (K_S) substitution rate. We also analyzed the expression patterns of these *GRAS* genes in tobacco using semi-quantitative reverse transcription polymerase chain reaction (RT-PCR). This study provides useful resources for further investigation on the structure and function of GRAS proteins in the regulation of tobacco growth.

MATERIAL AND METHODS

Plant materials

The cultivated variety of tobacco, K326, was planted and cultivated in a greenhouse under natural conditions. During the vegetative period, shoot tips (ST), lateral roots (LR), stems (S), and leaves (L) were collected from K326 plants; in the reproductive stage, flower buds (FB), and blooming flowers (F) were sampled from the plants. The samples were immediately frozen in liquid nitrogen, then total RNA was extracted from each organ and stored at -70°C after purity and integrity were determined.

Identification of *GRAS* genes

To isolate *GRAS* genes from tobacco, the sequences of 21 *GRAS* genes from *Arabidopsis* were downloaded from The *Arabidopsis* Information Resource (TAIR, http://www.arabidopsis.org/) (Tian et al., 2004; Lee et al., 2008). With the sequences of GRAS proteins from *Arabidopsis* as queries, the tBLASTn program was used to search against the American tobacco genome database (Sierro et al., 2014). Default parameters were used in the tBLASTn analyses, and false hits were removed by manual inspection. The ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) was used for open reading frame (ORF) prediction of tobacco *GRAS* genes. To perform comparative analysis of genomic organization of *GRAS* genes among tobacco, *Arabidopsis,* and rice, the sequences of homologous genes in rice were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/).

Sequence alignment, phylogenetic analysis and K_A/K_S

Alignment of 23 GRAS proteins in tobacco (NtGRAS1, NtLS, and 21 newly identified GRAS proteins) was performed to analyze the organization of protein motif, using the ClustalX2.0 program and was subsequently adjusted manually (Thompson et al., 1997). The sequence information of 33 previously reported GRAS proteins of *Arabidopsis*, 7 of tomato, 5 of rice, and 9 of other plants was retrieved from TAIR and NCBI. Including 23 GRAS proteins of tobacco, sequence alignment of 77 GRAS proteins was made using the ClustalX2.0 program. The phylogenetic tree was constructed by the neighbor-joining (NJ) method using a human (*Homo sapiens*) STAT protein (HsSRC; NP004374) as an outgroup (Lee et al., 2008). Bootstrap analysis was conducted to measure node robustness using 2000 replicates (Felsenstein, 1985). The ratio of K_A/K_S was calculated on the PAL2NAL web server (http://coot.embl.de/pal2nal/), which uses the codeml program of PAML to estimate K_A and K_S (Yang, 1997; Yang and Nielsen, 2000).

Semi-quantitative RT-PCR analysis

Semi-quantitative RT-PCR was performed to uncover the expression patterns of 21 new full-length *GRAS* genes in different organs of tobacco. Total RNA from ST, LR, S, L, F, and FB from tobacco were isolated using the GeneJET Plant RNA Purification Mini Kit (Thermo Fisher Scientific Inc., USA). Residual genomic DNA contamination was removed with DNase I and an RNase inhibitor (TaKaRa Biotechnology Co., Ltd., Dalian, China). Extracted RNAs were used as templates for synthesis of first-strand cDNAs using PrimeScript™ RT-PCR Kit (Perfect Real Time, TaKaRa Biotechnology Co., Ltd., Dalian) according to the product manual. Gene-specific primers for RT-PCR were designed according to the sequences of 21 *NtGRASLs* (Table 1). The specific primer pair (ATTAGGCCTGATGGGACGAA/GGCAACGTCCAAAGCATCAT) of the tobacco *NtL25* gene (ribosomal protein-coding gene, GenBank No. L18908) was used as an internal control. Reactions were performed with *Taq*HS Polymerase (TaKaRa Biotechnology Co., Ltd., Dalian) on a TaKaRa PCR Thermal Cycler Dice™, with the following profile: 94°C for 3 min, 35 cycles of 10 s at 94°C, 30 s at 55°C, and 1 min at 72°C.

Table 1. Sequences of primers used for RT-PCR analysis.

Gene name	Forward primers	Reverse primers
NtGRASL1	TTTCTTGTCCCAAGCAGTCCAT	ACAACGCTCCGTAATACCCACA
NtGRASL2	CGATGTTGAGATTGAGCACCTA	CCTGGGAAAGAACTGTACTGTATT
NtGRASL3	AGAATCTTCATGTTAGCGTGAGA	AAACCTGTAATTTGGGCTGTAC
NtGRASL4	GTGAGAAGCGGCACAAAGCAAT	CTGGTGACTTCAGGAGCGAAAA
NtGRASL5	CCAGCGTCTTGGTGCTTACATT	CTCATCCCTTGGGTTATTCACG
NtGRASL6	TGAAACTCCTCATCTCCGACTC	AGCCCTCACATCATCACAAACT
NtGRASL7	TTACCACCTATGCTTGATCTTTC	ACAACATCTTCCTCATCCCTCT
NtGRASL8	ATCGCAAACTCAGCCGTTCTAC	TATTGTTCTCGCTCACAACTACTC
NtGRASL9	CTCTTCACCGCTCTTCAATCGT	CTGCTCCACTGTATGCCTTTCT
NtGRASL10	TGGGTAGTGGCAATTTCTGGTT	CCTGTCATTTGGATTAGGTGGC
NtGRASL11	CGGGTCGGGTTCTATTTCTTAG	TACTCGCCCAATGTCACTATGC
NtGRASL12	GCATTTAAGGCTTTCTACCAGA	AACCAACCATTCATTCCCATTT
NtGRASL13	CTGAAGTTGGTCCGAGGAGAAG	AGTGAATCATAAAGGGCTGAGTAG
NtGRASL14	AAGAGTTAGCCGATAAAGCCGAGTG	ATCGCATCCTTCCACGAACCAC
NtGRASL15	TGGCTGTGGTGCTCAATGGTCC	CGGAGTGGTGGGTTGAAGGAGA
NtGRASL16	CTTCTTGAGTCGTTAGATGCCGTAA	AAACAGTGCCCTCCAATGCTG
NtGRASL17	GAACAAGAAGCGAATCATAACG	TCTCCCTAAATACACCTCAGACAT
NtGRASL18	CCTGCCTTTACCGACCTTCTAC	TACACCCTCACAGGACACCAAA
NtGRASL19	TCTGTAACTGCTCCACAAACTCC	AAAGCCTCTAAGATTGCACCAT
NtGRASL20	TTGCGATTAGTTCAGTTATGC	TAGAATCTTGTTCTGTTACCACC
NtGRASL21	CGCCATATCTACTTCTGATGTTC	AGCCATTCTAAACCGTGACTT

RESULTS

Isolation of tobacco *GRAS* genes

To identify *GRAS* genes from tobacco, we used the sequences of 21 *GRAS* genes in *Arabidopsis* to search against the American tobacco genome database. As a result, we isolated 21 homologous *GRAS* genes in tobacco, that is to say, each *GRAS* gene in *Arabidopsis* had a single orthologous gene in tobacco. Isolated tobacco *GRAS* genes all possessed complete ORFs, and encoded proteins all contained 407-755 amino acids (Table 2). Comparison of genomic organization showed that most tobacco *GRAS* genes have no

introns, expect for *NtGRASL2*, *3*, *4*, and *19*. The orthologous genes among three plants had the same genomic organization, with or without introns. However, several combinations had different conclusions, such as *NtGRASL2*/*AtGRAS-19*/*OsGRAS-19*, *NtGRASL3*/*At-GRAS-24*/*OsGRAS-39*, *NtGRASL20*/*AtGRAS-5*/*OsGRAS-6*, and *NtGRASL21*/*PAT1*.

Table 2. Comparative analysis of the genomic organization of the *GRAS* genes in tobacco, *Arabidopsis*, and rice.

Tobacco			*Arabidopsis*			Rice		
Name	Intron	Length (AA)	Name	Protein ID	Intron	Name	Accession No.	Intron
NtLS	N	407	*AtLAS*	At1g55580.1	N	*MOC1*	AP003944.1	N
NtGRAS1	N	644	*AtGRAS-12*	At2g29060.1	N	*OsGRAS-17*	AC097277.4	N
NtGRASL1	N	640	*AtGRAS-30*	At5g52510.1	N	*OsGRAS-22*	AL606688.2	N
NtGRASL2	Y	546	*AtGRAS-19*	At3g50650.1	N	*OsGRAS-19*	AAAA01008118	N
NtGRASL3	Y	544	*AtGRAS-24*	At4g17230.1	Y	*OsGRAS-39*	AC113335.1	N
NtGRASL4	Y	583	*SCR*	At3g54220.1	Y	*OsGRAS-41*	AC123519.1	Y
NtGRASL5	N	582	*RGL1*	At1g66350.1	N	*OsGAI/SLR1*	AC087797.5	N
NtGRASL6	N	514	*SHR*	At4g37650.1	N	*OsGRAS-38*	AP005437.3	N
NtGRASL7	N	453	*AtGRAS-18*	At3g49950.1	N	*OsGRAS-37*	AP005149.4	N
NtGRASL8	N	529	*AtGRAS-16*	At3g13840.1	N	*OsGRAS-14*	AC135559.4	N
NtGRASL9	N	436	*AtGRAS-28*	At5g41920.1	N	*OsGRAS-35*	AP003866.2	N
NtGRASL10	N	674	*AtGRAS-8*	At1g63100.1	N	*OsGRAS-32*	AP001168.1	N
NtGRASL11	N	602	*AtGRAS-32*	At5g66770.1	N	*OsGRAS-19*	AAAA01008118	N
NtGRASL12	N	518	*RGL3*	At5g17490.1	N	*OsGAI/SLR1*	AC087797.5	N
NtGRASL13	N	508	*AtGRAS-23*	At4g08250.1	N	*OsGRAS-13*	AC135206.3	N
NtGRASL14	N	542	*AtGRAS-25*	At4g36710.1	N	-	-	-
NtGRASL15	N	712	*AtGRAS-21*	At3g60630.1	N	*OsGRAS-21*	AL606647.2	N
NtGRASL16	N	755	*AtGRAS-22*	At4g00150.1	N	*OsGRAS-8*	AP004883.3	N
NtGRASL17	N	568	*GAI*	At1g14920.1	N	*OsGAI/SLR1*	AC087797.5	N
NtGRASL18	N	563	*RGA*	At2g01570.1	N	*OsGAI/SLR1*	AC087797.5	N
NtGRASL19	Y	566	*AtGRAS-4*	At1g21450.1	Y	*OsGRAS-3*	AP003561.2	Y
NtGRASL20	N	466	*AtGRAS-5*	At1g50420.1	Y	*OsGRAS-6*	AP003259.4	N
NtGRASL21	N	544	*PAT1*	At5g48150.1	Y	-	-	-

Names of *GRAS* genes in *Arabidopsis* and rice are adopted from a previous report (Tian et al., 2004). The tobacco *GRAS* genes are named with the prefix *NtGRASL*. "N" indicates no introns in ORF. "Y" indicates introns are in ORF. "-" means no close genes are found. The lengths of the predicted proteins are given as the number of amino acid residues.

Features of tobacco GRAS proteins

Alignment of 23 GRAS proteins in tobacco showed that the order of conserved domains on the C-termini was similar, and the sequence variation was largely caused by different lengths of the N-terminus (Figure 1).

The VHIID sequence is present in all NtGRASLs with the conserved pattern of proline-asparagine-histidine-aspartic acid-glutamine (P-N-H-D-Q), except for the substitution of a glutamine residue (Q) with a histidine residue (H) in NtGRASL14. The LHRI and LHRII domains, which are characterized by several leucine repeats can be divided into two units, A and B units. A conserved LXXLL pattern (Leu-Xaa-Xaa-Leu-Leu; Xaa is for any amino acid) occurred in A unit of LHRI motif in half of the NtGRASLs, such as NtGRASL3 (LKELL), NtGRASL4 (LKQLL), NtGRASL5 (LKQLL), NtGRASL9 (LLGLL), NtGRASL10 (LISLL), NtGRASL12 (LHHLL), NtGRASL13 (LVHLL), NtGRASL20 (LIHLL), and NtGRAS1 (LRGLL). The significance of this structure in plants remains unknown. Unit B of LHRI contains a putative nuclear localization signal (NLS), which is characterized by two basic residues (arginine or lysine) separated by 11 amino acids (Tian et al., 2004). The NLS sequence was

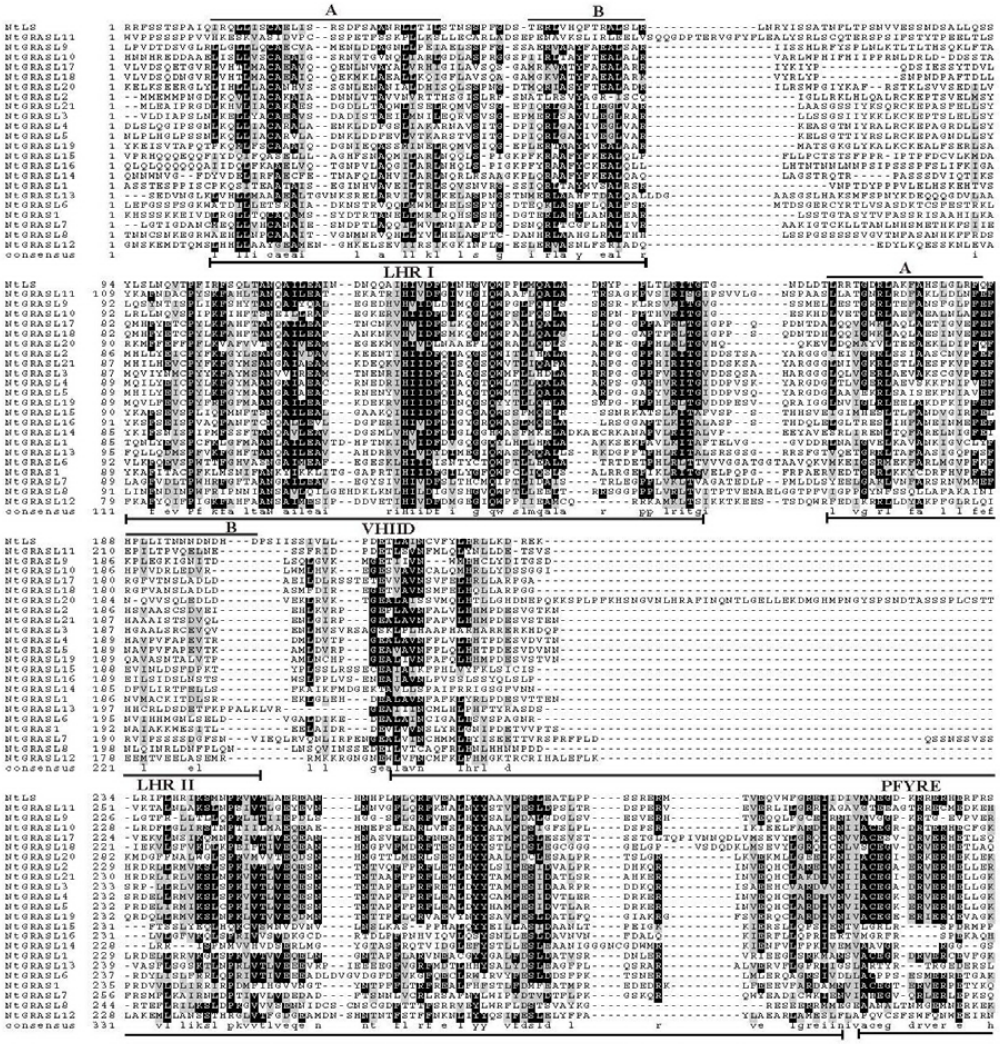

Figure 1. Alignment of GRAS proteins in tobacco. The GRAS proteins contain five conserved motifs, LHRI, VHIID, LHRII, PFYRE, and SAW. Identical amino acid residues are displayed in black, and similar residues are displayed in gray. Conserved residues are highlighted in capital letters on the bottom line, relatively conserved residues are shown in lowercase.

found in NtGRASL17 and NtGRASL18, the consensus of which was KVATYFAEALARR. Other GRAS proteins had few basic residues in the putative NLS region. In unit A of the LHRII motif, a few GRAS proteins were found to contain regularly spaced LHRs (LX6LX6L, where L is for leucine and X is for any amino acid) followed by several irregularly spaced leucine repeats, such as NtLS (LRRTGDRLAKFAHSL), NtGRASL11 (LLATGNRLRDFAKLL), NtGRASL10 (LVETGDRLAEFAEAL), and NtGRASL20 (LDQHAYVLTEEAEKL). Nonetheless, the LXXLL motif was rare in the latter part of LHRII. The LHRI-VHIID-LHRII sequence is involved in the binding of GRAS proteins to target nucleic acids or proteins (Hirsch et al., 2009; Hou et al., 2010).

The PFYRE motif can be divided into three units: proline (P), aromatic phenylalanine and tyrosine (FY), and arginine and glutamic acid residues. An aspartic acid/glutamic acid residue that exists in the FY unit was conserved in almost all tobacco GRAS proteins. In addition, the sequences of the PFYRE motif were largely collinear and showed a high degree of similarity among tobacco GRAS proteins.

The SAW motif was found to be composed of two units, the RVER and W-W-W. RVER patterns were not conserved in all GRAS proteins. For example, NtGRASL14, 15 and 16 lacked this motif. The W-W-W unit included three subunits: W-G, L-W, and S-W. The three W residues were conserved in nearly all GRAS proteins. Although the roles of the PFYRE and SAW motifs are currently unknown, the absolute conservation of the residues in these motifs implies that these motifs are important for the functional and structural integrity of GRAS proteins.

Phylogenetic analysis of the *GRAS* gene family and calculation of K_A/K_S

Proteins with high sequence identity generally have similar functions across different species (Song et al., 2014). Thus, the phylogenetic analysis of GRAS proteins from different plants will provide a solid foundation for functional studies of NtGRASLs in future.

The NJ method was used to construct a phylogenetic tree based on an alignment among 77 GRAS proteins (Figure 2). Eight distinct subfamilies HAM, LS, SHR, PAT, SCL3, DELLA, LISCL, and SCR were defined in this phylogenetic tree. The amino acid sequence of GRAS proteins was highly homologous within each subgroup, which suggests that each GRAS subfamily may share a common ancestral protein or similar functions. It is worth noting that NtGRASL12 formed an independent branch, suggesting that it is a unique *GRAS* gene in tobacco. AtGRAS-8 and NtGRASL10 were close to each other, and did not belong to any existing subgroups. The same was observed for AtGRAS-23/NtGRASL13.

The DELLA subfamily included two GRAS members from tobacco, NtGRASL17, and NtGRASL18. Their putative sequences contain the distinguished DELLA and TVHY(K)NP domains at the N-termini, which were reported to regulate GA responses in plants (Dill et al., 2001; Fu et al., 2002). This result was different from that observed in another solanaceae plant, tomato, which had a single *DELLA* gene (Bassel et al., 2004). The SCL3 subfamily included two members, NtGRASL20 and AtGRAS-5. These two proteins probably had similar biological functions; however, direct evidence remains forthcoming. The SCR subfamily, which plays a key role in radial pattern formation for roots and shoots, contained only one tobacco GRAS protein, NtGRASL9. As the first isolated GRAS member from tobacco, NtGRAS1 belongs to the LISCL subfamily. It was also the first functionally characterized member of this subfamily, and might be an important transcriptional regulator involved in the plant stress response (Czikkel et al., 2007). The LS subfamily consists of nine GRAS

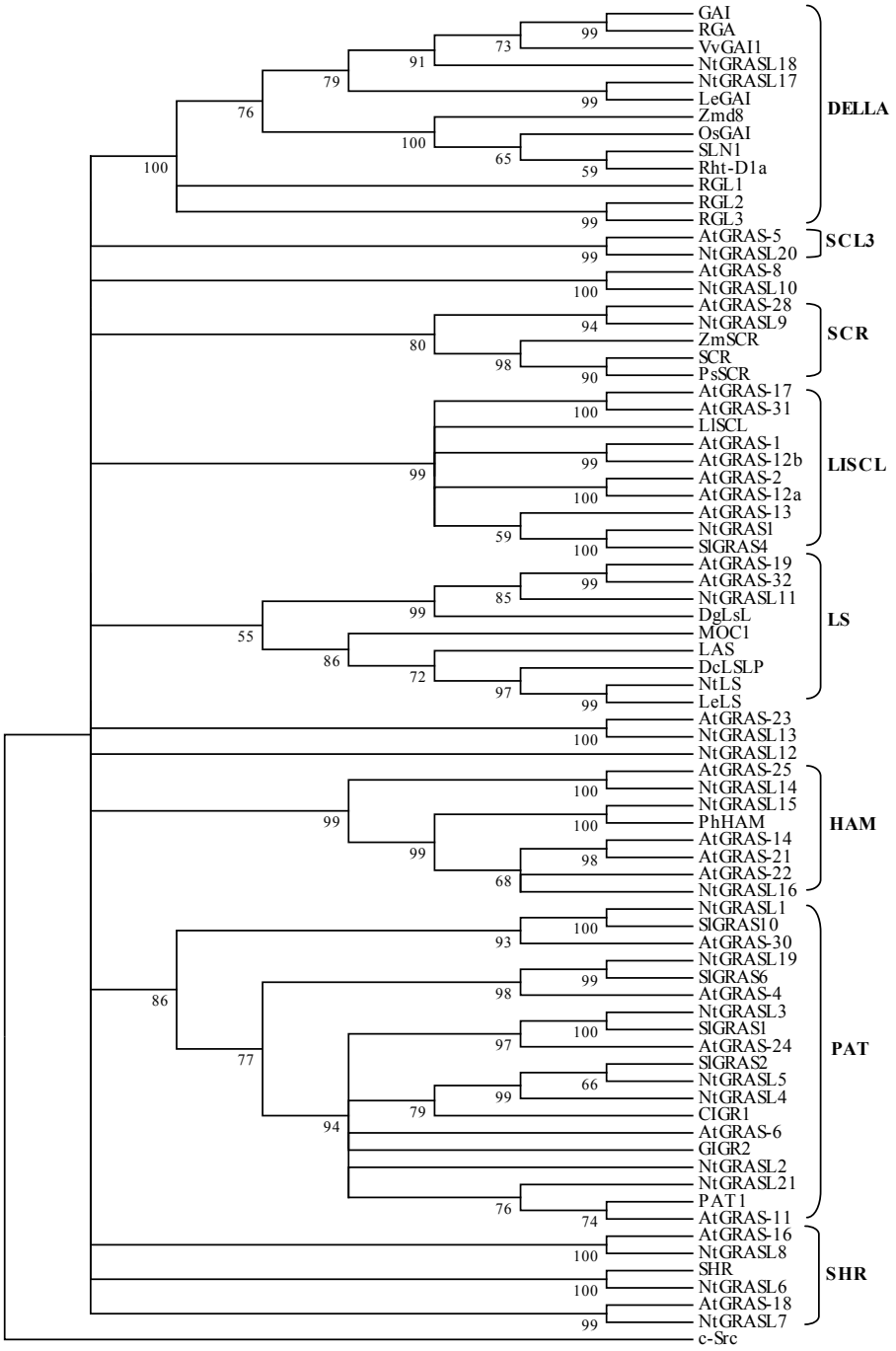

Figure 2. Phylogenetic tree of GRAS proteins in tobacco, *Arabidopsis*, and other plants. The neighbor-joining phylogenetic tree and the combined P values are shown. Human STST protein was used as an outgroup.

proteins, of which NtLS and NtGRASL11 are from tobacco. *LAS*, *LeLs*, *DcLSLP*, *DgLsL*, and *MOC1* were orthologues, and have been reported to regulate shoot branching or tillering (Schumacher et al., 1999; Greb et al., 2003; Li et al., 2003; Yang et al., 2005). *NtLS*, which was closely related to *LeLS*, is thought to be involved in the same process. NtGRASL14, 15, and 16 belong to the HAM subfamily and lack the RVER domain at their C-terminus. The first functionally characterized member of this subfamily, *PhHAM,* has been associated with shoot apical meristem maintenance in petunia (Stuurman et al., 2002). The PAT subfamily was the largest branch in our phylogenetic tree, containing 19 members, seven of which are tobacco GRAS proteins. Among them, PAT1 was found in *Arabidopsis* during phytochrome A signal transduction (Bolle et al., 2000); CIGR1 and CIGR2 are thought to be involved in the elicitor-induced defense response in rice (Day et al., 2003); tomato GRAS proteins, SlGRAS1, SlGRAS2, SlGRAS6, and SlGRAS10 were also classified as members of this subgroup; however, SlGRAS4 belongs to the LISCL subfamily. The SHR subfamily included two members, SHR and NtGRASL6, whereas, the branches containing AtGRAS-16/NtGRASL8 and AtGRAS-18/NtGRASL7 were classified as part of this subfamily as they appeared to be closely related to SHR. Among them, only SHR has been studied in detail. Taken together, all 21 GRAS proteins from tobacco are distributed across eight subfamilies in a decentralized fashion.

Estimation of the K_A/K_S ratio is crucial to determine whether any selection pressure exists during gene evolutionary history. Based on phylogenetic analysis, we measured the K_A and K_S mutation rates of 13 pairs of putative GRAS orthologous genes between tobacco and *Arabidopsis* (Table 3). The results show that K_A was lower than K_S ($K_A/K_S << 1$) in 13 pairs of GRAS orthologs, which indicates that negative selection acted on amino acid mutations. Among them, there was a significant trend for purifying selection in *NtGRASL3* and *AtGRAS-24*. Pairs of tobacco-*Arabidopsis* orthologs might correspond to well-conserved functions. However, only *LAS* has been functionally characterized among these genes.

Table 3. Calculation of K_A and K_S of 13 pairs of *GRAS* orthologs between tobacco and *Arabidopsis*.

Gene pairs	K_A	K_S	K_A/K_S
NtGRASL20 vs AtGRAS-5	0.2058	4.5693	0.0450
NtGRASL10 vs AtGRAS-8	0.3080	4.2265	0.0729
NtGRASL9 vs AtGRAS-28	0.2646	2.4310	0.1089
NtLS vs LAS	0.3548	5.5100	0.0644
NtGRASL13 vs AtGRAS-23	0.4215	3.9809	0.1059
NtGRASL14 vs AtGRAS-25	0.3651	5.3444	0.0683
NtGRASL16 vs AtGRAS-22	0.3838	8.0708	0.0476
NtGRASL1 vs AtGRAS-30	0.4207	11.2680	0.0373
NtGRASL19 vs AtGRAS-4	0.3287	3.3428	0.0983
NtGRASL3 vs AtGRAS-24	0.2595	46.7970	0.0055
NtGRASL8 vs AtGRAS-16	0.4015	2.7267	0.1473
NtGRASL6 vs SHR	0.2647	8.6638	0.0305
NtGRASL7 vs AtGRAS-18	0.2816	3.7121	0.0758

K_S = synonymous substitution rate; K_A = nonsynonymous substitution rate.

Expression analysis of *GRAS* genes in tobacco

In this study, we analyzed different expression levels of tobacco GRAS proteins in six tissues with semi-quantitative RT-PCR (Figure 3). The transcript levels of 21 tobacco GRAS

genes were obtained from at least one tissue. Interestingly, most GRAS genes were expressed at relatively higher levels in the tissues, for example, *NtGRASL2*, *3*, *4*, *5*, *6*, *7*, *15*, *16*, *17*, *18*, *19*, and *20* were found to be expressed in all tissues. Nevertheless, several NtGRAS proteins exhibited tissue-specific expression. For example, NtGRASL10 is expressed only in the ST, L and LR; NtGRASL11 is only expressed in the ST, FB, and LR; NtGRASL21 was not detected in the LR; NtGRASL14 was not expressed in the L and LR; no expression of NtGRASL8 was observed in the shoot tip and leaf; NtGRASL9, however, was found in neither the F or L. The transcript level of NtGRASL1 in L was obviously lower than that in other tissues. NtGRASL12 and NtGRASL13 were expressed at relatively lower levels in all tissues.

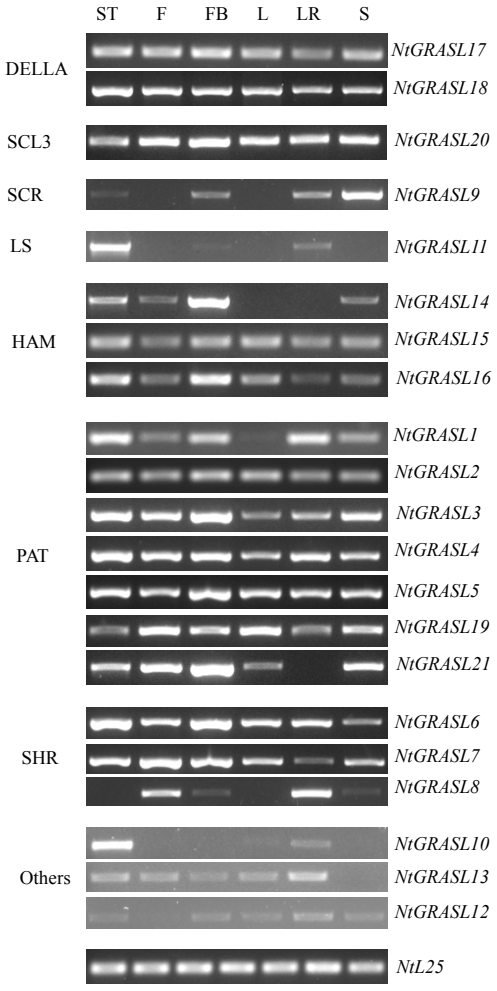

Figure 3. Expression patterns of 21 tobacco *GRAS* genes. The names of GRAS subfamilies are shown on the left hand side. "Others" indicates the genes that cannot be classified in the existing subfamily. ST, F, FB, L, LR, and S represents shoot tip, flower, flower bud, lateral root, and stem, respectively. The names of the tobacco *GRAS* genes are shown on the right hand side. *NtL25* was used as an internal control.

DISCUSSION

Isolating target genes in a plant using homologous sequences identified from model plants is an important and successful strategy. *GRAS* gene products play important roles in the physiology and development of higher plants. As a species, *N. tabacum* evolved through the interspecific hybridization of the ancestors of *Nicotiana sylvestris* and *Nicotiana tomentosiformis* about 2000 years ago. Thus, it can be speculated that abundant *GRAS* genes exist in *N. tabacum*. In this study, because of the lack of information on *GRAS* genes in tobacco species, we used a homology-based cloning method to investigate this further. Of 33 *Arabidopsis GRAS* genes, 21 were chosen to search against the American tobacco genome database, as they had been somewhat functionally characterized. We identified 21 *GRAS* genes in tobacco; however, more *GRAS* genes may be present within the whole-genome of tobacco. Here, we provide a preliminary overview of the *GRAS* genes in tobacco, including their sequence features, phylogenetic analysis, and expression profiles.

Studies on the sequence of 23 tobacco *GRAS* genes and their homologous genes in *Arabidopsis* and rice showed that the homologous genes had very similar genomic organizations, with or without introns. Identification of orthologous genes will facilitate the future study of their functions, as evidenced by *NtLS/Ls/LAS/MOC1/DgLsL*, *SCR/PsSCR/ZmSCR*, and *GAI/RGA/SLR1/ZmD8/Rht1/SLN*. Thus, according to the previous reports about *LS*, *LAS*, *MOC1*, and *DgLs* (Schumacher et al., 1999; Greb et al., 2003; Li et al., 2003; Yang et al., 2005), *NtLS* is thought to be related to axillary meristem formation.

In the analysis of protein structure, tobacco GRAS proteins were also found to contain highly conserved motifs at their C-termini, whilst conservation of their N-terminal sequences is much lower. A conserved LXXLL pattern was found in the A unit of the LHRI motif in several NtGRASLs, which is different from previous reports suggesting that LXXLL occurred in the latter part of the LHRII motif (Sun et al., 2012). The LXXLL pattern has been demonstrated to mediate the binding of steroid receptor co-activator complexes to cognate nuclear receptors in mammals (Heery et al., 1997). Maybe the significance of this structure in plants is unique, but it remains unknown. Phylogenetic analysis has helped to reveal the evolutionary consequences of duplicated *GRAS* genes, such as occurs in *GAI/RGA/RGL1/RGL2/RGL3*. Twenty-one NtGRASLs that are distributed among eight GRAS subfamilies of confirmed that different genes might play different roles.

RT-PCR analysis revealed that most tobacco *GRAS* genes are expressed at relatively higher levels in some tissues, which indicates that these genes have important and extensive functions during tobacco growth and development. While several *NtGRASLs* exhibit tissue-specific expression, which suggests that various *NtGRASLs* may be involved in the diversification of morphological features. *NtGRASL11* was notable for its expression profile being consistent with that of *NtLS* (Tai SS, Chen YQ, Wang L, Wang WF, et al., unpublished results), in the ST, FB, and LR. As *NtGRASL11* also belong to the *LS* subfamily like *NtLS* (Figure 2), we inferred that the expression patterns of *LS* subfamily members had similarities in tobacco, as well as the biological functions. It remains to be determined whether *NtLS* and *NtGRASL11* are both involved in axillary meristem initiation.

In this study, we describe 21 *GRAS* genes in *N. tabacum*; however, more potential *GRAS* genes of tobacco have yet to be identified. The comparative and phylogenetic analysis of these *GRAS* genes will act as a first step towards their functional characterization.

ACKNOWLEDGMENTS

Researtch supported by the National High Technology Research and Development Program ("863" Program) of China (#2012AA021801). The authors thank all the staff in the Department of Biotechnology of TRIC for their help in providing plant materials, giving technical assistance in bioinformatics, and providing advice on the manuscript.

REFERENCES

Bassel GW, Zielinska E, Mullen RT and Bewley JD (2004). Down-regulation of DELLA genes is not essential for germination of tomato, soybean, and *Arabidopsis* seeds. *Plant Physiol*. 136: 2782-2789.

Bolle C (2004). The role of GRAS proteins in plant signal transduction and development. *Planta* 218: 683-692.

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.

Czikkel BE and Maxwell DP (2007). *NtGRAS1*, a novel stress-induced member of the *GRAS* family in tobacco, localizes to the nucleus. *J. Plant Physiol*. 164: 1220-1230.

Day RB, Shibuya N and Minami E (2003). Identification and characterization of two new members of the *GRAS* gene family in rice responsive to N-acetylchitooligosaccharide elicitor. *Biochim. Biophys. Acta*. 1625: 261-268.

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.

Dill A, Jung HS and Sun TP (2001). The DELLA motif is essential for gibberellin-induced degradation of *RGA*. *Proc. Natl. Acad. Sci. U. S. A*. 98: 14162-14167.

Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.

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.

Greb T, Clarenz O, Schafer E, Muller D, et al. (2003). Molecular analysis of the *LATERAL SUPPRESSOR* gene in *Arabidopsis* reveals a conserved control mechanism for axillary meristem formation. *Genes Dev*. 17: 1175-1187.

Heckmann AB, Lombardo F, Miwa H, Perry JA, et al. (2006). *Lotus japonicus* nodulation requires two GRAS-domain regulators, one of which is functionally conserved in a non-legume. *Plant Physiol*. 142: 1739-1750.

Heery DM, Kalkhoven E, Hoare S and Parker MG (1997). A signature motif in transcriptional co-activators mediates binding to nuclear receptor. *Nature* 387: 733-736.

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

Hou X, Lee LY, Xia K, Yan Y, et al. (2010). DELLAs modulate jasmonate signaling via competitive bingding binding to JAZs. *Dev. Cell* 19: 884-894.

Ikeda A, Ueguchi-Tanaka M, Sonoda Y, Kitano H, et al. (2001). Slender rice, a constitutive gibberellin response mutant, is caused by a null mutation of the *SLR1* gene, an ortholog of the height-regulating gene *GAI/RGA/RHT/D8*. *Plant Cell* 13: 999-1010.

Itoh H, Shimada A, Ueguchi-Tanaka M, Kamiya N, et al. (2005). Overexpression of a GRAS protein lacking the DELLA domain confers altered gibberellin responses in rice. *Plant J*. 44: 669-679.

Kalo P, Gleason C, Edwards A, Marsh J, et al. (2005). Nodulation sigaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulator. *Science* 308: 1786-1789.

Lee MH, Kim B, Song SK, Heo JO, et al. (2008). Large-scale analysis of the *GRAS* gene family in *Arabidopsis thaliana*. *Plant Mol. Biol*. 67: 659-670.

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

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 J, 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.

Peng J, Richards DE, Hartley NM, Murphy GP, et al. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* 400: 256-261.

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.

Schumacher K, Schmitt T, Rossberg M, Schmitz G, et al. (1999). The *Lateral suppressor* (*Ls*) gene of tomato encodes a new member of the VHIID protein family. *Plant Biol.* 96: 290-295.

Sierro N, Battey JN, Ouadi S, Bakaher N, et al. (2014). The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Comm.* 5: 3833-4833.

Silverstone AL, Ciampaglio CN and Sun T (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 in 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* gene family in Chinese cabbage. *Genomics* 103: 135-146.

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.

Sun X, Xue B, Jones WT, Rikkerink E, et al. (2011). 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: 205-223.

Sun X, Jones WT and Rikkerink EH (2012). GRAS proteins: the versatile roles of intrinsically disordered proteins in plant signalling. *Biochem. J.* 442: 1-12.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, et al. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 25: 4876-4882.

Tian C, Wan P, Sun S, Li J, 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.

Yang DH, Yun PY, Park SY, Plaha P, et al. (2005). Cloning, characterization and expression of a *Lateral suppressor-like* gene from chrysanthemum (*Dendranthema grandiflorum Kitamura*). *Plant Physiol. Biochem*. 43: 1044-1051.

Yang Z (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci*. 13: 555-556.

Yang Z and Nielsen R (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol*. 17: 32-43.

Zhang J, Zhang Y, Du Y, Chen S, et al. (2011). Dynamic metabonomic responses of tobacco (*Nicotiana tabacum*) plants to salt stress. *J. Proteome Res*. 10: 1904-1914.