Journal of Experimental Botany, Vol. 59, No. 14, pp. 3831–3844, 2008
doi:10.1093/jxb/ern223 Advance Access publication 9 October, 2008
This paper is available online free of all access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

Journal of Experimental Botany
www.jxb.oxfordjournals.org

RESEARCH PAPER

Comparative phylogenetic analysis of small GTP-binding genes of model legume plants and assessment of their roles in root nodules

Bayram Yuksel* and Abdul R. Memon

Plant Molecular Biology Laboratory, Genetic Engineering and Biotechnology Institute, Marmara Research Center, TUBITAK, PO Box 21, 41400, Gebze, Kocaeli, Turkey

Received 22 May 2008; Revised 17 July 2008; Accepted 6 August 2008

Abstract

Small GTP-binding genes play an essential regulatory role in a multitude of cellular processes such as vesicle-mediated intracellular trafficking, signal transduction, cytoskeletal organization, and cell division in plants and animals. *Medicago truncatula* and *Lotus japonicus* are important model plants for studying legume-specific biological processes such as nodulation. The publicly available online resources for these plants from websites such as http://www.ncbi.nih.gov, http://www.medicago.org, http://www.tigr.org, and related sites were searched to collect nucleotide sequences that encode GTP-binding protein homologues. A total of 460 small GTPase sequences from several legume species including *Medicago* and *Lotus*, *Arabidopsis*, human, and yeast were phyletically analysed to shed light on the evolution and functional characteristics of legume-specific homologues. One of the main emphases of this study was the elucidation of the possible involvement of some members of small GTPase homologues in the establishment and maintenance of symbiotic associations in root nodules of legumes. A high frequency of vesicle-mediated trafficking in nodules led to the idea of a probable subfunctionalization of some members of this family in legumes. As a result of the analyses, a group of 10 small GTPases that are likely to be mainly expressed in nodules was determined. The sequences determined as a result of this study could be used in more detailed molecular genetic analyses such as creation of RNA inteference silencing mutants for further clarification of the role of GTPases in nodulation. This study will also assist in furthering our understanding of the evolutionary history of small GTPases in legume species.

Key words: *Arabidopsis thaliana*, *ARF*, *Lotus japonicus*, *Medicago truncatula*, nodulation, phylogenetic, *RAB*, *RAC*, *ROP*, small GTPases.

Introduction

Small GTP-binding genes are implicated in a myriad of different functions including the early and late secretory pathway (Bucci *et al*., 1992; Novick and Brennwald, 1993), abiotic and biotic stress signal transduction pathways (Ono *et al*., 2001; Morino *et al*., 2004; Wong *et al*., 2004), and mitotic spindle assembly (Kalab *et al*., 1999; Dasso, 2001). This family of genes is usually divided into four main subfamilies in plants: (i) *ARF*/*SAR*; (ii) *RAB*; (iii) *ROP* (Rho-like proteins in plants); and (4) *RAN* (Vernoud *et al*., 2003). Because of their role in basic housekeeping functions, they are evolutionarily well conserved in both Embryophyta and metazoa/fungi of the Eukaryotae (Bourne *et al*., 1990, 1991; Jekely, 2003; Jiang and Ramachandran, 2006). Despite a high level of functional conservation across the entire Eukaryotae, an occasional sub- or neofunctionalization of some of these genes in plants is suggested (Rutherford and Moore, 2002; Vernoud *et al*., 2003). It is a generally accepted hypothesis that the duplicated genes after speciation could be the main source for evolution to work on, resulting in the generation of genes with a new function, i.e. neofunctionalization, or with a slightly different function from its ancestral form, i.e. subfunctionalization (Chapman *et al*., 2006). To test this idea for small GTPases of the

* To whom corespondence should be addressed. E-mail: bayram.yuksel@mam.gov.tr

Fabaceae lineage of the Embryophyta, this family of genes was functionally and phyletically analysed by retrieving all available sequences from online databases of the main Fabaceae species; *Medicago truncatula*, *Medicago sativa*, *Cicer arietinum*, *Pisum sativum*, *Vigna radiata*, and *Vicia faba*, together with *Arabidopsis thaliana*, *Homo sapiens*, and *Saccharomyces cerevisiae*.

The assessments of the functional identities of small GTP-binding genes from different plant species suggest that these proteins for the most part are functionally very well conserved across the whole plant kingdom. However, some of these proteins might have undergone functional diversification in different lineages of the plant kingdom to regulate some lineage-specific functions such as nodulation in Fabaceae. Therefore, it is reasonable to suggest that comparative functional and phyletical analyses between legume homologues of small GTP-binding genes and other species would be very informative in terms of retrieving the genes that could have subfunctionalized in nodules since the split of Fabaceae from other main lineages of Embryophyta. The physical locations, phyletic affiliations, and functional identities of many gene families are inter-related. In other words, the genes that are phyletically closely associated also tend to be related in terms of their genomic positions. In order to gain insights into evolutionary and functional characteristics of these proteins in more detail, phylogenetic analyses of these genes have been carried out for several eukaryotic systems. The first detailed phylogenetic analysis of small GTP-binding genes was conducted by Garcia-Ranea *et al*. (1998) in yeast. Phylogenetic assessments of the evolution of *RHO* genes in terrestrial plants were conducted by Winge *et al*. (2000), with the general conclusion that this family of genes could have evolved to compensate the loss of the *RAC* subfamily of animals. Furthermore, *Arabidopsis* small GTP-binding genes have been phyletically compared with similar proteins from other eukaryotic organisms including human and yeast, allowing some postulations about the functional and expression identities of *Arabidopsis* genes (Vernoud *et al*., 2003). In a recent study, Jiang *et al*. (2006) suggested an early diversification of these genes in Eukaryotae on the basis of the K_a/K_s ratio. Thus, even though these genes are mostly functionally very strictly conserved due to their basic housekeeping roles, there are some phyletic groups that are likely to be functionally diversified.

A possible tissue-specific sub- or neofunctionalization of small-GTP binding genes in the formation and maintenance of nodular structures in fabaceous plant roots has been suggested. For example, *RAB1p* and *RAB7p* are implicated to be essential in the formation of the peribacteroid membrane in *Glycine max* and *Vigna aconitifolia* (Cheon *et al*., 1993), and Son *et al*. (2003) showed that *RAB7* could play an essential role in vesicular fusion during rhizobial endocytosis and symbiosome formation in soybean. Moreover, a total of 33 small GTP-binding genes isolated form a cDNA library of *Lotus japonicus* (Borg *et al*., 1997), the majority of which are ubiquitously expressed in all plant tissues with the exception of *RAB1*, *RAB2*, *RAB5*, *RAB7*, and a *RAC* that are predominantly expressed in nodules. Likewise, Schiene *et al*. (2004) demonstrated that *RAB11* from *M. sativa* is predominantly expressed in the root nodules. All of the aforementioned examples could suggest a nodule-specific subfunctionalization of these genes in legumes (Son *et al*., 2003). This conclusion provided the idea of looking further into the Fabaceae small GTPases with the hope of finding more of these genes that could be significant for the formation and maintenance of nodular structures through regulating functions such as shuttling of products between symbiont and host. To realize this purpose, an attempt was made to retrieve small GTPases from all freely available expressed sequence tag (EST) databases of two model legume species, *Medicago* and *Lotus*, which are possibly solely or predominantly expressed in nodular structures. It is reasonable to presume that the subfunctionalization of paralogous genes in a lineage-specific manner is likely to follow a divergence in spatial expression.

Medicago truncatula and *Lotus japonicus* are two main model plants for legume species (Frugoli and Harris, 2001; Young *et al*., 2003; Cannon *et al*., 2006; Cronk *et al*., 2006), and full genome sequences of these model organisms are likely to further our knowledge of the molecular basis of formation and functioning of symbiosis organelles, i.e. nodules. Several studies have been carried out with the purpose of describing the genes that are significant for nodules by using different experimental approaches such as transcriptomal profiling (Fedorova *et al*., 2002; Wienkoop and Saalbach, 2003; Manthey *et al*., 2004). Furthermore, comparative analyses of symbiosis-associated genes with homologues from non-leguminous species contributed to our comprehension of how symbiosis evolved; for example, one such attempt was carried out to investigate the non-legume orthologues of genes associated with arbuscular mychorrhizal symbiosis and nodulation (Zhu *et al*., 2006). By the same token, the comparative analysis of small GTP-binding genes from legumes and non-legumes would allow better understanding of the possible roles of these genes in symbiotic associations.

The main objective of this study is to shed light on the evolutionary history of small GTP-binding genes in legume species, namely *M. truncatula*, *L. japonicus*, *Glycine max*, *M. sativa*, *Cicer arietinum*, *P. sativum*, *V. faba*, and *V. radiata*, and non-legumes and animals, namely *A. thaliana*, *H. sapiens*, and *S. cerevisiae*, by comparative analyses of phyletic and expression characteristics. In more detail, this study aimed at finding those Fabaceae homologues of this gene family that could have

gained Fabaceae-specific functions after the split. The main components of this study are: (i) retrieval of all possible small GTP-binding gene homologues from *Medicago* and *Lotus* and other fabaceous species by searching through all sequence databases; (ii) phyletical analysis of Fabaceae, where the main gene source was *M. truncatula* and *L. japonicus*, compared with *A. thaliana*, a Brassicaceae, *H. sapiens*, and *S. cerevisiae*, with the main objective of better elucidation of Fabaceae-specific evolution of this family after divergence; (iii) detection of the spatiotemporal expression pattern of *Arabidopsis* small GTP-binding gene homologues from microarray experiments (Zimmermann *et al*., 2005) and correlative analysis of functional characteristics of these genes with the phyletic positions, with the main emphasis on shedding light on the functional identities of Fabaceae genes; (iv) identification of possible small GTP-binding genes that are solely or mainly expressed in nodules; (v) performing an overall evaluation of the physical and phyletic positions of the putative nodule-unique expressed genes relative to homologues from other species; and (vi) making an overall assessment of the functional role of these genes and their phyletical characteristics.

Materials and methods

Retrieval of sequences

Medicago truncatula sequences were retrieved from several publicly available online data resources including; http://www.tigr.org (TIGR, The Institute for Genomic Research); http://www.medicago.org (The International Consortium for Sequencing Medicago Genome); http://medicago.tolouse.inra.fr (European Medicago Consortium); http://www.genome.ou.edu/medicago.html (University of Oklahoma); http://www.noble.org/medicago.html (Noble Organization, Oklahoma, USA); http://www.ncbi.nih.nlm.gov (GenBank); http://notmutdb.vbi.vt.edu (database mainly containing information about the nodulation mutants, Virginia Technical University); and http://www.plantgdb.org (PGD, Plant Genome Database). Likewise, *Lotus japonicus* sequences were retrieved from http://www.kazusa.or.jp/lotus.html (*Lotus japonicus* Genome Sequencing Project); http://www.shigen.nigac.jp/lotusjaponicus (National BioResource Project, Legume Database, *Lotus japonicus*); PGD, GenBank, and TIGR.

For the rest of the fabaceous species, *G. max, M. sativa, P. sativum, V. faba, V. radiata*, and *C. arietinum*, GenBank resources were used. *Arabidopsis thaliana* sequences were obtained from http://www.arabidopsis.org (TAIR, Arabidopsis Information Resource). *Homo sapiens* sequences were acquired from two sources; http://gdbwww.gdg.org (GDB, The GDB Human Genome Database) and http://www.ensembl.org (ENSEMBL). Finally, GenBank and http://www.yeastgenome.org (SGD, Saccharomyces Genome Database) were databanks for yeast sequences.

Screening of the databases

The databases contained annotated sequences and protein sequences for *Medicago* from IMGAG (International Medicago Genome Annotation Group), EST constructs for *Medicago* and *Lotus* TCs from TIGR, and PUT from PGD, nucleotide and protein sequences from GenBank and other databases (for more detailed information about sequence resources, GIs nomenclature conventions, and other sequence characteristics, see Supplementary Table S1 available at *JXB* online). All the sequences were downloaded and local databases created with Microsoft Access. First, all databases were screened for small GTP-binding gene homologues by employing the Blast algorithm (Altschul *et al*., 1990) (Blastp for protein databases and Blastx for nucleotide databases). For the initial screening, *Arabidopsis*, human, and yeast small GTP-binding protein sequences were used and every database rescanned for the second round with the sequences obtained from the initial screening. Any sequence that produced an E-value $\leqslant 1\times10^{-7}$ were considered for further analysis. The sequence databases were also searched by using keywords for GTP-binding gene homologues.

All the nucleotide sequences that were obtained at the end of database searches were conceptually translated in six reading frames, and open reading frames (ORFs) that are longer than 100 amino acids in length were kept for further analyses (Supplementary Table S1 at *JXB* online).

Nomenclature

The details of the naming convention followed in this paper can be found in the supplementary data (Supplementary Table S1 at *JXB* online). Briefly, for sequences acquired from GenBank, the first letter of the genus name followed by the two initial letters of the species name, and followed by the gene identification number, e.g. *Msa74095371*, where 'M' stands for *Medicago*, 'sa' for *sativa*, and 74094371 for the GI of the sequence. In the nomenclature of EST assembly sequences and genome annotation sequences, the source of the sequence was denoted by adding the abbreviation of data source after the letters symbolizing the name of the species, i.e. PUT for PGD, TC for TGR EST assembly sequences, and IMGA for *Medicago* genome annotation sequences.

Multiple alignment and phylogenetic tree construction

A total of 460 small GTP-binding gene amino acid sequences (Table 1) were multiply aligned, as were the sequences of all GTP-binding genes by utilizing the ClustalX program (Thompson *et al*., 1997) with the purpose of discerning the four main subfamilies: *ARF*, *RAB*, *ROP*, and *RAN*. After that, each subfamily of genes was multiply aligned within its own group, and the sequences that did not have the expected motifs or were truncated were discarded and not used for the construction of phylogenetic trees. The Neighbor–Joining trees were constructed with MEGA3 (Kumar *et al*., 2003) after the multiple alignment of each subfamily of genes.

Table 1. *The total number of small GTPases used in this study*

More details about the sequences can be found in Supplementary Table S1 at *JXB* online.

Organisms	Families				
	ARF	*RAB*	*ROP*	*RAN*	Total
M. truncatula	31	74	16	6	127
L. japonicus	4	56	7	0	67
A. thaliana	21	57	11	4	93
Other Fabaceae[a]	1	34	18	2	55
H. sapiens	20	41	22	1	94
S. cerevisiae	5	11	6	2	24
Total	82	273	80	15	460

[a] Other Fabaceae: *Medicago sativa*, *Pisum sativum*, *Glycine max*, *Vicia faba*, *Vigna radiata*, and *Cicer arietnium*.

Detection of nodule uniqueness of some Fabaceae-specific small GTPases

The details of the procedure employed for the identification of possible nodule-unique sequences (i.e. the small GTP-binding genes that are expressed solely or mainly in nodular tissues) are explained in Fig. 5. In brief, the ESTs from all libraries of *Medicago* and *Lotus* available online (Supplementary Table S4 at *JXB* online) were downloaded together with those from two new databases containing ESTs, on the basis of the source tissue of the libraries: nodular tissues and the other tissues (Fig. 5). The ESTs that are likely to be unique to nodular tissues were determined by cross-comparison of the above-mentioned databases. To gauge the efficiency of this analysis, all the ESTs were Blasted against Embryophyta databases; it was observed that the majority of sequences shared high sequence homology with the commonly known nodule-specific proteins such as nitrate reductase, nodulins, and *RAB11F*. The ESTs obtained from this comparative analysis were Blasted against the small GTP-binding gene sequences, and small GTP-binding genes with nearly 100% identity to the nodule-unique ESTs were considered as possible candidates that are exclusively expressed in nodular tissues. For further confirmation of the results, they were checked against all the small GTP-binding genes reported to be mainly expressed in nodular tissues using online available 2D expression profiling data, at both the protein and transcript level.

Spatial and temporal expression profiling of GTP-binding genes in Arabidopsis thaliana

Two main online sources were used to determine the spatial and temporal expression profile of small GTP-binding gene homologues in *Arabidopsis*; Genevestigator Arabidopsis microarray expression profile data, http://www.genevestigator.ethz.ch (Zimmermann *et al*., 2005) and Arabidopsis MPSS (Massively Parallel Signature Sequencing) http://mpss.udel.edu/at/. Each *Arabidopsis* small GTPase gene was searched for in Genevestigator microarray data to elucidate the temporal and spatial expression characteristics, and the results obtained from Genevestigator were compared with those from MPSS data (Supplementary Table S2 at *JXB* online).

Results and Discussion

Small GTPase gene homologues

A total of 460 small GTP-binding protein sequences were obtained from 11 different organisms; nine plants, human, and a yeast species (Table 1 and Supplementary Table S1 at *JXB* online) after scanning of major sequence databases (see Materials and methods for details). The majority of these sequences belonged to two model leguminous species, 127 from *M. truncatula* and 67 from *L.aponicus*, and 55 sequences were collected from other legume species. The remaining sequences were from *A. thaliana*, 93; *H. sapiens*, 94; and *S. cerevisiae*, 24. The largest numbers of sequences, 273, were grouped in the *RAB* subfamily, followed by the *ARF* subfamily with 82, the *ROP* subfamily with 80, and the *RAN* subfamily with 15. As a result of this study, the total number of small GTPases of *Medicago* is estimated to be ~120, which is much greater than the 93 *Arabidopsis* small GTPases. Even though most of the small GTPases in this study were predicted from assembled EST sequences, meaning that some of these sequences could be artefacts, this estimation is likely to be accurate because the screening of a recently published *Populus trichocarpa* genome sequence (Tuskan *et al*., 2006) for the small GTP-binding gene orthologues yielded a total of 117 sequences; 31 *ARF*/*SAR*, 68 *RAB*, 14 *ROP*, and four *RAN*. Furthermore, the rice genome is reported to have 111 of these genes (Jiang and Ramachandran, 2006). The most recent common ancestor of eudicots and monocots was estimated to have ~34 small GTP-binding genes (Jiang and Ramachandran, 2006), which is consistent with the number of small GTPases of a single-celled photosynthetic green alga, *Chlamydomonas reinhardtii*, i.e. 28 (Merchant *et al*., 2007; Rensing *et al*., 2008). Hence, it could be argued that the rate of expansion of these genes has been different at the main lineages of plants. Since these genes are implicated to play key regulatory roles in many vital biological processes such the regulation of responses of plants to the multitude of environmental stresses including salinity and diseases (Bolte *et al*., 2000; Ono *et al*., 2001; Mazel *et al*., 2004; Morino *et al*., 2004; Jung *et al*., 2006), there could be a direct correlation between the number of small GTPases in a specific lineage and its adaptability to different environmental conditions. Thus, it is reasonable to assume that some homologues of these genes could have been subfunctionalized to perform lineage-specific regulatory roles such as nodulation in leguminous plants. The functional and structural evolution of these genes could be better illuminated with the sequencing of more genomes representing main lineages of plants.

Identification of small GTP-binding gene homologues of Lotus *and* Medicago *that are likely to be uniquely or predominantly expressed in nodular structures*

As mentioned above, one of the main purposes of this study was the identification of small GTP-binding genes from *Lotus* and *Medicago* plant, which are solely or mainly expressed in root nodules (see Materials and methods for the details). In order to identify possible nodule-unique small GTPase candidates, 38 457 EST sequences were downloaded from 19 different cDNA libraries (the source tissue of six of them is nodules) of *L. japonicus* (Table 2; Supplementary Table S4 at *JXB* online; Fig. 5); and 109 862 EST sequences originating from 75 different cDNA libraries (the source tissues of 11 libraries were several developmental stages of nodular structures) of *M. truncatula* (Table 3; Supplementary Table S4 at *JXB* online). From the comparative analyses of *Lotus* EST sequences, it is concluded that three small GTPases, two *RAB*s and a *ROP*, could be uniquely expressed in nodular structures (Table 2; Supplementary Table S3.1 and S3.2 at *JXB* online). Likewise, there were seven similar sequences from *M. truncatula*, five of which were *RAB*-like sequences and two of which were from

Table 2. *The number of possible nodule-unique transcripts from* Lotus japonicus

The methodology employed for determining the transcripts likely to be expressed solely in nodular tissues is explained in Fig 5.

Tissue origin	No. of libraries	No. of of transcripts	No. of small GTP-binding gene homologues[a]
Nodule	6	9122	106[b]
Nodule-unique	–	1228	3[c]
All others	13	29 335	–
Total	19	38 457	108

[a] The sorted EST sequences Blasted all the Fabaceae small GTP-binding gene homologues identified in this study (Table 1), and any sequence that had a Blast score $>10^{-7}$ was considered to be a small GTP-binding gene homologue.
[b] The total number of EST sequence with a blast score $>10^{-7}$; these EST sequences were also present in libraries constructed from tissues other than nodule.
[c] These were not present in libraries other than nodules.

Table 3. *The list of small GTP-binding genes sorted relative to their nodular expression profiles*

The details of the method employed for the separating out of the list of small GTP-binding genes that could be only expressed in nodular tissues is denoted in Fig. 5.

Tissue	Stages	No. of days post-infection[a]	No. of libraries	No. of transcripts	No. of GTPases
Nodule[b]	Early	0–72 h	3	6,611	3
	Mid	3–15 d	5	3610	3
	Late	>15 d	2	4110	–
	Mixed[c]	–	1	2014	1
Not nodule	–	–	64	93 517	47[d]
Total	–	–	75	109 862	55

[a] The number of days after the inoculation of *Rhizobium* in the roots.
[b] Some of these libraries may also contain tissues from the root; however, it needs to be clarified that the uniqueness of these transcripts to the nodules was decided on the basis of their presence only in libraries containing nodular tissues
[c] The true characteristics of the libraries in terms of nodulation stage were not specified.
[d] These small GTP-binding transcripts were also present in tissues other than nodules.

ARF subfamily groups (Table 3; Supplementary Table S3.1 at *JXB* online; Fig. 5). For *Medicago* sequences, an attempt was made to pinpoint the possible developmental stages of nodules in which the small GTPases are increased (Table 3). As a result of these analyses, a total of 10 small GTPase homologues were identified that are likely to be either only or predominantly expressed in nodular tissues (Tables 2, 3; Supplementary Tables S5, S3.1, S3.2 at *JXB* online). However, it should be noted that these sequences are mostly predicted from EST assemblies; thus, these results need to be interpreted cautiously until experimentally proven. The main purpose of conducting these analyses was to determine possible phyletic groups of small GTPases with mainly nodular expression, thereby playing a significant role in the formation and maintenance of nodular structures. It is also important to emphasize that it can only be asserted that some phyletical groups of Fabaceae small GTPases evolved to be specialized to perform lineage-specific functions (i.e. nodulation) through a likely subfunctionalization. It has been shown by several studies that neofunctionalization and subfunctionalization are the two main mechanisms shaping the evolution of paralogous genes such as in yeast (He and Zhang, 2005) and in *DOF* gene families of poplar, *Arabidopsis*, and rice (Yang *et al*., 2006). Besides, the correct directing of shuttling of products between symbiotic partners, rhizobial bacteria and leguminous plants, is critical for nodular functioning; hence, it is reasonable to suggest that the small GTPases are the main regulators of this trafficking. Furthermore, several previous studies have shown that the expression level of some small GTPases is altered in nodules of several leguminous plants (Cheon *et al*., 1993; Borg *et al*., 1997; Fedorova *et al*., 2002; Son *et al*., 2003; Manthey *et al*., 2004; Schiene *et al*., 2004), hinting at a possible involvement of these genes. In conclusion, it could be suggested that some of the GTPases identified in this study are likely to play a role in legume–rhizobium symbiosis associations. However, to clarify better the possible roles of small GTPases in the establishment of root–rhizobium interaction and in the development of nodules, further experimental studies such as expression profiling need to be conducted.

Spatial expressional analysis of GTP-binding genes in Arabidopsis thaliana

In order to assess the possible correlative associations between phyletic position and spatial expression diversification of small GTPases, microarray-based expression profiling databases from Genevestigator (Zimmermann *et al*., 2005) and the MPSS database from the website http://mpss.udel.edu/at/ were searched for all 93 *A. thaliana* small GTPases (see Materials and methods for further details, and Supplementary Table S2 at *JXB* online). By considering the assumption of inter-relatedness of phyletic position and functional identity, an attempt was made to make some extrapolations regarding the spatial expression characteristics of Fabaceae small GTPase homologues from the spatial expressional pattern of *Arabidopsis* genes. At the end of this analysis, 47 of 93 these genes from *Arabidopsis* were found to be ubiquitously expressed (Table 4; Supplementary Table S2 at *JXB* online). Before making any further postulations, it needs to be noted that the expression of some of these genes was shown to be increased in more than one tissue, meaning some genes were equally dominant in more than one tissue, for instance both root, and stamen and pollen. This result was expected, since the majority of these genes regulate basic cellular functions such as intracellular

Table 4. *The spatial expression analysis of small GTP-binding genes in* Arabidopsis

The spatial expression pattern of each *Arabidopsis* small GTP-binding gene was characterized with the gene chip data of Genevestigator (Zimmermann *et al*., 2005); the expression profiling results are summarized on the basis of dominance of tissue specificity of the gene. More detailed results can be found in Supplementary Table S2 at *JXB* online.

Spatial expression profile	Subfamilies			
	RAB	*ARF*	*ROP*	*RAN*
Ubiquitous	30	12	5	4
Root tissue	14 (8)[a]	3	4	–
Regenerative tissue[b]	6	5	3	–
Stamen and pollen	17 (14)	6 (2)	3 (2)	–
All other[c]	3	–	–	–

[a] The numbers in parentheses denote the number of small GTP-binding genes expressed mainly in the specified tissue.
[b] Regenerative tissues are hypocotyl, callus, shoot apex, nodes, etc.
[c] The expression level was very low in all tissues or the transcriptomal profiles did not fit into any of the described categories.

protein trafficking (Lee *et al*., 2002; Kotzer *et al*., 2004; Memon, 2004; Munro, 2005). Root, stamen and pollen were the major organelles where the expression levels of some small GTPase homologues were significantly higher relative to those of other tissues. Hence, it could be suggested that these genes are likely to be the main molecular switches coordinating the busy vesicle trafficking and protein targeting required in rapidly regenerating plant tissues (Cheung *et al*., 2002; Nielsen *et al*., 2002; Arthur *et al*., 2003; Chen *et al*., 2003; Preuss *et al*., 2004; Schiene *et al*., 2004; de Graaf *et al*., 2005). For instance, 17 of the small GTP-binding genes of the *RAB* subfamily were expressed at a high level in pollen and stamen tissues, 14 of which were mainly present in pollen and stamen (Table 4). Likewise, 14 of the *RAB* GTPase homologues were expressed at a high level in root tissues, and eight of these were principally present only in roots (Table 4). *RAB* subfamily genes were much more diversified in their spatial expression characteristics. This result could be explained partly by the fact that *RAB* subfamily genes are the most expanded and diverged among different lineages of Embryophyta, suggesting a possible lineage-specific subfunctionalization of some paralogues (Rutherford *et al*., 2002). Further supporting this conclusion, seven out of 10 of the small GTPases identified in this study to be mainly expressed in nodular structures of leguminous plants were from this family. However, it should be noted that it is very likely that the expansion rate of different subfamilies of these genes shows variation among different lineages of Embryophyta; for instance, *A. thaliana* has 21, *V. vinifera* has 20, *P. trichocarpa* has 31, and *Oryza sativa* has 43 *ARF/SAR* subfamily genes. This could mean that different subfamilies could have followed different evolutionary paths in terms of both copy number expansion and functional divergence. Likewise, about half of a total of 21 *ARF/SAR* homologues of *Arabidopsis* were noticeably different in terms of their spatial expression divergence, implying a possible subfunctionalization (Table 4). Even though almost all of the *ROP* genes have some level of expression in all tissues, some of them were predominantly or solely expressed in root, stamen, and pollen, or other regenerative tissues such as shoot apex and node stem, corroborating previous studies conducted to shed light on the role of *ROP* genes in *Arabidopsis* (Molendijk *et al*., 2001; Fu *et al*., 2002; Gu *et al*., 2003). Finally, there was no noteworthy difference in the spatial expression pattern of *RAN* GTPases. To sum up, it could be said that small GTPases are ubiquitously expressed, but some of them could have undergone functional modifications to perform tissue-specific functions, thereby contributing to organelle differentiation.

Phylogenetic analyses of GTP-binding genes in Arabidopsis*, Fabaceae, human, and yeast, and functional phylogenetic comparative associations*

Amino acid sequences of four subfamilies of small GTPases, 460 in total, from mainly two fabaceous species, *Medicago* and *Lotus*, human, and yeast were multiply aligned and the alignments were edited using Jalview (http://www.jalview.org). After the alignment, the sequences that contained major errors such as large deletions and inversions, or those that were too diversified or too truncated were not considered for further analysis. Neighbor–Joining trees were constructed by using the default parameters of Mega 3 (Kumar *et al*., 2003). The confidentiality of branch nodes was tested by bootstrapping 10 000 times, and values <50 were not denoted (see Materials and methods for the details of tree construction). In order to make trees more legible, the protein sequences that are >99% identical were not depicted on the final trees. The same notations were utilized as by Vernoud *et al*. (2003) for *Arabidopsis* homologues. In order to make better correlative assessments about the phyletic position and functional identity of genes, spatial expression characteristics of each *Arabidopsis* gene are depicted on the tree picture (Figs 1–4) (for the details of notations and abbreviation used in tree pictures, see Materials and methods and figure legends).

Comparative phyletic analysis of Fabaceae ARF *genes*: ADP-ribosylation factor 1 (ARF1) has been shown to have a very critical role in COPI-mediated retrograde trafficking in yeast, mammalian, and plant systems (Aoe *et al*., 1998; Gaynor *et al*., 1998; Lee *et al*., 2002; Takeuchi *et al*., 2002). *Arabidopsis* have five nearly identical duplicates of *ARF1*, *AthARFA1a*, *AthARFA1b*, *AthARFA1c*, *AthARFA1d*, and *AthARFA1e* (Vernoud *et al*., 2003) (Fig. 1). These duplicate copies of the gene

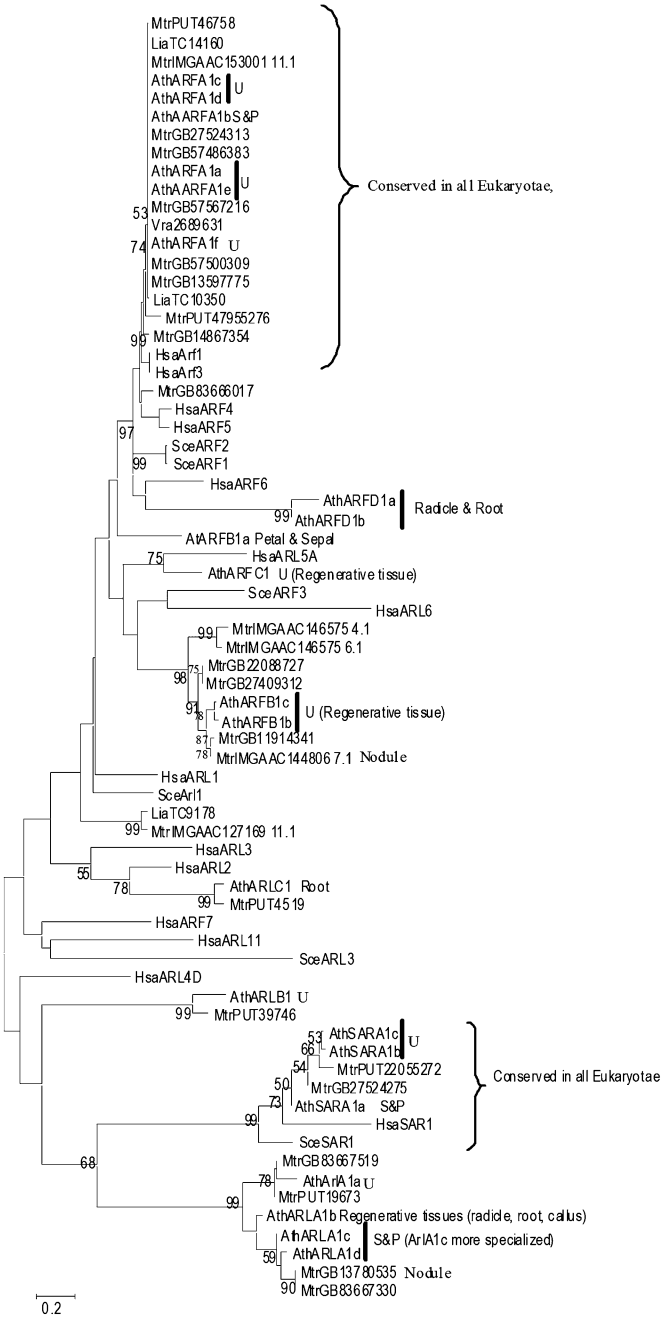

Fig. 1. *ARF* and *SAR* homologues from several fabaceous species, *Arabidopsis*, human, and yeast were phyletically analysed by construction of a Neighbor–Joining tree utilizing Mega3.1 (Kumar *et al*., 2003) after multiple alignment of the protein sequences with ClustalX (Thompson *et al*., 1997). The confidence levels of nodes were tested by bootstrapping 1000 replications, and the bootstrap values that are >50% are denoted on the branches. The sequences are named using the convention that the first letter of the name of the sequence is the first letter of the genus and the following two letters are the initial two letters of the species: Ath, *Arabidopsis thaliana*; Mtr, *Medicago truncatula*; Lja, *Lotus japonicus*; Msa, *Medicago sativa*; Gma, *Glycine max*; Psa, *Pisum sativum*; Car, *Cicer arietnium*; Vra, *Vigna radiata*; Vfa, *Vicia faba*; Hsa, *Homo sapiens*; Sce, *Saccharomyces cerevisiae*. The spatial expression characteristics of each *Arabidopsis* small GTP-binding gene (Supplementary Table S2 at *JXB* online) are denoted at the side of each sequence. For the tissues in which the genes are predominantly expressed: U, ubiquitous, i.e. expressed in all tissues including roots, pollen and stem, nodes, leaf; S&P, stamen and pollen; Regenerative tissues, tissues, with the exception of stamen and pollen, where the active growth occurs, such as callus, root tips; the tissues in parentheses denotes that the gene is predominantly expressed in the specified tissue. The genes that are likely to be expressed only in nodular tissues, *MtrGB13780535* and *MtrIMGAAC144806_7.1*, are shown.

are likely to be functionally redundant albeit that there may not be a full functional overlap (Takeuchi *et al*., 2002). This argument is supported by the spatial expression pattern of these genes, i.e. all four copies are ubiquitously expressed, but one of them, *AthARFA1b*, was more highly expressed in stamen and pollen tissues, hinting at a possible functional divergence (Fig. 1). *Medicago* has more duplicate copies in this clade, implying a probable more functional diversification of these genes in fabaceous species relative to *Arabidopsis*. In conclusion, since these genes play such an essential role in COPI-mediated vesicular trafficking from Golgi to endoplasmic reticulum (i.e. retrograde transport) (Nakano and Muramatsu, 1989; BarPeled and Raikhel, 1997; Takeuchi *et al*., 2000; Aridor *et al*., 2001; Lee *et al*., 2005), having nearly identical multiple copies of these genes is likely to be a prudent strategy to ensure the survival of the cell; thus, all copies of these genes are likely to be almost functionally overlapping, with slight divergence in expression. Similar conclusions could be reached about the *SAR* homologues, which are known to be involved in the intracellular COPII-mediated protein (Memon, 2004) trafficking from the endoplasmic reticulum to Golgi apparatus. However, the copy number of the duplication rate of the *SAR* phyletic group was much lower than that of *ARF*; only three copies were found in the *Arabidopsis* genome, *AthSARA1a*, *AthSARA1b*, and *AthSARA1c* (Fig. 1). Similarly, only two homologues were retrieved from *Medicago* sequences. Hence, the expansion of this group of genes was much more limited in both *Medicago* and *Arabidopsis*. Another phyletic group that could have undergone a burst of copy number expansion in plants is a Plantae-specific phyletic group, *AthARFB1c*, implying a possible lineage-specific functional diversification. In fact, *MtrIMGAAC144806_7.1*, denoted as being mainly or uniquely expressed in nodular tissues, belongs to this phyletic group. Moreover, a human-specific cladistic group of *HsaARL4A*, *HsaARL4C*, and *HsaARL4D* is implicated to be involved in mammalian-specific functions such as somitogenesis, neurogenesis, and early spermatogenesis, thus further supporting the idea of lineage-specific functional divergence of these genes (Lin *et al*., 2000). Similarly, *HsaARL3* is suggested to play a role in vesicular transportation or cell signalling in photoreceptors of the human eye together with *RP2*. Furthermore, spatial expression characteristics of *ARF*-like genes, *AthARLA1c* and *AthARLA1d* in stamen and pollen tissues, and *AthARL1b* in regenerative tissues, is

Fig. 2. *RAB* homologues from several fabaceous species, *Arabidopsis*, human, and yeast were multiply aligned by using ClustalX (Thompson *et al*., 1997), and the multiply aligned protein sequences were phyletically characterized by the construction of a Neighbor–Joining tree using Mega V3 (Kumar *et al*., 2003). The confidence levels of nodes were tested by bootstrapping 1000 times, and scores >50% are denoted. The sequences are named using the convention that the first letter of the name of the sequence is the first letter of the genus and the following two letters are the initial two letters of the species: Ath, *Arabidopsis thaliana*; Mtr, *Medicago truncatula*; Lja, *Lotus japonicus*; Msa, *Medicago sativa*; Gma, *Glycine max*; Psa, *Pisum sativum*; Car, *Cicer arietnium*; Vra, *Vigna radiata*; Vfa, *Vicia faba*; Hsa, *Homo sapiens*; Sce, *Saccharomyces cerevisiae*. Different colours are used to depict the sequences from different families: with *Arabidopsis* sequences in magenta, legume sequences in cyan, and human and yeast sequences in yellow. The spatial expression characteristics of each *Arabidopsis* small GTP-binding gene (Supplementary Table S2 at *JXB* online) are denoted at the side of each sequence. For the tissues in which the genes are predominantly expressed: U, ubiquitous, i.e. expressed in all tissues including roots, pollen and stem, nodes, leaf; S&P, stamen and pollen; Regenerative tissues, tissues, with the exception of stamen and pollen, where the active growth occurs, such as callus, root tips; the tissues in parentheses denotes that the gene is predominantly expressed in the specified tissue. The genes that are likely to be solely expressed in nodular tissues, *MtrIMGAAC1221648.2*, *MtrPUT45174*, *LjaTC12838*, *MtrPUT18368*, *MtrGB13781173*, *LjaTC18154*, and *MtrGB27408074*, are shown with arrows.

consistent with their phyletic properties (Fig. 1). Likewise, it was indicated that one of the *Medicago* homologues, *MtrGB13780538* (Fig. 1), could be one of two transcripts that are mainly expressed in nodular structures, further supporting the conclusion about the functional diversification of the genes belonging to this phyletic group. On the

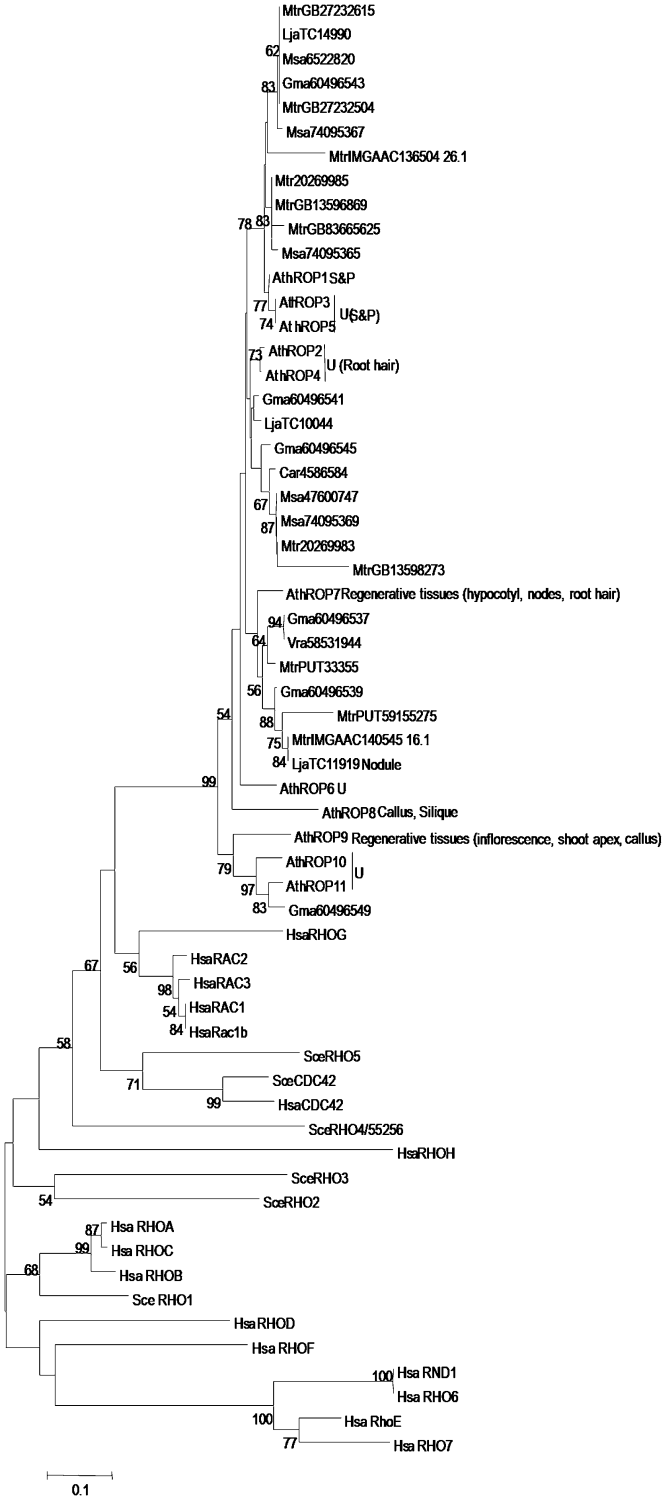

Fig. 3. ROP (Rho of plants) homologues from different fabaceous, *Arabidopsis*, human, and yeast were multiply aligned by using ClustalX (Thompson *et al*., 1997) and the multiply aligned protein sequences were phyletically characterized by construction of a Neighbor–Joining tree using Mega V3 (Kumar *et al*., 2003). The confidence levels of nodes were tested by bootstrapping 1000 times. and scores >50%, are denoted. The sequences are named using the convention that the first letter of the name of the sequence is the first letter of the genus and the following two letters are the initial two letters of the species: Ath, *Arabidopsis thaliana*; Mtr, *Medicago truncatula*; Lja, *Lotus japonicus*; Msa, *Medicago sativa*; Gma, *Glycine max*; Psa, *Pisum sativum*; Car, *Cicer arietnium*; Vra, *Vigna radiata*; Vfa, *Vicia faba*; Hsa, *Homo sapiens*; Sce, *Saccharomyces cerevisiae*. The spatial expression characteristics of each *Arabidopsis* small GTP-binding gene (Supplementary Table S2 at *JXB* online) are denoted at the side of each sequence. For the tissues in which the genes are predominantly expressed: U, ubiquitous, i.e. expressed in all tissues including roots, pollen and stem, nodes, leaf; S&P, stamen and pollen; Regenerative tissues, tissues, with the exception of stamen and pollen, where the active growth occurs, such as callus, root tips; the tissues in parentheses denotes that the gene is predominantly expressed in the specified tissue. A single gene that is likely to be expressed only in nodular tissues, *LjaTC11919*, is shown.

other hand, some of these genes, such as *HsaARL1*, could take part in one of the universal eukaryotic functions such as localization of the TGN (*trans*-Golgi network) (Burd *et al*., 2004; Latijnhouwers *et al*., 2005; Munro, 2005; Shin *et al*., 2005; Stefano *et al*., 2006). Two genes identified as possible candidates of nodule-specific sub-functionalization from *Medicago*, *MtrIMGAAC144806_7.1* and *MtrGB13780538*, belong phyletically only to plant-specific clades and were more diversified in sequence. Similarly, *Arabidopsis* orthologues closely associated with these two sequences, *AthARFB1c*, *AthARFB1d*, *AthARLA1c*, and *AthARLA1d*, were also more diversified in terms of expression. The phyletic characteristics of probable nodule-unique sequences further strengthens the hypothesis about specialization of some members of the *ARF* subfamily in nodular structures because organelle-specific functionalization would logically follow an diversification in expression and function.

Comparative phyletic analysis of Fabaceae RAB *genes*: The *RAB* subfamily of these genes are the most divergent and expansive group. This could suggest that it is very likely that this group of proteins are also functionally very diverged; in fact, corroborating this idea, these proteins are implicated in a multitude of biological functions such as in the regulation of early and late endocytic pathways (Chavrier *et al*., 1991; Bucci *et al*., 1992), and *AthRABF1* (also known as *Ara-6* or *AtRAB5c*) in early endocytic sterol transport in an actin-dependent manner (Grebe *et al*., 2003). RAB proteins have also been functionally linked to the molecular basis of nodulation in legumes. For instance, RAB1, RAB2, RAB5, and RAB7 proteins showed increased expression in nodular structures (Borg *et al*., 1997); moreover, RAB1 and RAB7 proteins are known to be part of peribacterioid membrane biogenesis (Cheon *et al*., 1993; Son *et al*., 2003). Thus, the divergence of functional and spatiotemporal expression is a common phenomenon for some members of RAB proteins, yet some of them are ubiquitously expressed and universally conserved across the whole Eukoryatae. In the present analyses, an attempt was made to understand the possible correlative relationships between the phyletic, expressional, and functional identities of these genes, and

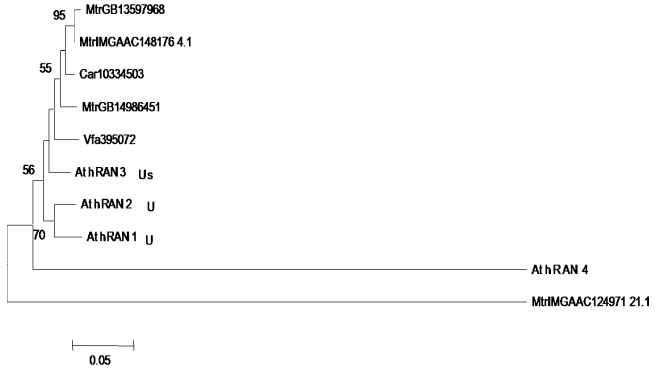

Fig. 4. *RAN* homologues from several fabaceous species and *Arabidopsis* were multiply aligned by using ClustalX (Thompson *et al*., 1997) and the multiply aligned protein sequences were phyletically characterized by the construction of a Neighbor–Joining tree using Mega V3 (Kumar *et al*., 2003). The confidence levels of nodes were tested by bootstrapping 1000 times, and scores >50% are denoted. The sequences are named using the convention that the first letter of the name of the sequence is the first letter of the genus and the following two letters are the initial two letters of the species: Ath, *Arabidopsis thaliana*; Mtr, *Medicago truncatula*; Lja, *Lotus japonicus*; Msa, *Medicago sativa*; Gma, *Glycine max*; Psa, *Pisum sativum*; Car, *Cicer arietnium*; Vra, *Vigna radiata*; Vfa, *Vicia faba*; Hsa, *Homo sapiens*; Sce, *Saccharomyces cerevisiae*. The spatial expression characteristics of each *Arabidopsis* small GTP-binding gene (Supplementary Table S2 at *JXB* online) are denoted at the side of each sequence. For the tissues in which the genes are predominantly expressed: U, ubiquitous, i.e. expressed in all tissues including roots, pollen and stem, nodes, leaf; S&P, stamen and pollen; Regenerative tissues, tissues, with the exception of stamen and pollen, where the active growth occurs, such as callus, root tips; the tissues in parentheses denotes that the gene is predominantly expressed in the specified tissue.

to get a clearer idea of the possible lineage-specific functionalization in terms of their likely roles in nodular structures.

Arabidopsis genes belonging to the *RABE1* family are one of the most highly conserved phyletic groups, hinting that these genes are likely to be involved in housekeeping functions, and further supporting this concept is that almost all of these genes are ubiquitously expressed in all tissues (Fig. 2). In fact, *AthRABE1* is involved in the regulation of the secretory pathway in the Golgi or after the Golgi (Zheng *et al*., 2005). The spatial expression pattern of *AthRABA4a*, *AthRABA4b*, *AthRABA4c*, and *AthRABA4d* is specialized; most of these genes were highly or mainly expressed in root tissues, except *AthRABA4d* which shows stamen- and pollen-specific expression (Fig. 2; Supplementary Table S2 at *JXB* online). This conclusion is consistent with the studies conducted to reveal the functional identities of these genes; for instance, *AthRABA4b* is involved in the polarized secretion of cell wall components during the growth of root hair cells (Preuss *et al*., 2006). Furthermore these genes, the *AtRABA4* group, are more diverged in sequence relative to the *AthRABE1* phyletic group, hinting at a possible functional divergence. There are certain phyletic groups of *RAB* GTPases such as *HsaRAB3A*, *HsaRAB3B*, *HsaRAB3C*, and *HsaRAB3D* which are duplicated either only in metazoa/fungi or in Embryophyta. The first assumption is likely to be correct, because the phyletic topography of these genes shows the characteristics of long internal branch lengths and shorter external branches, suggesting a recent burst in copy numbers. These genes are likely to have undergone functional specialization, e.g. the *RAB3A* gene is required for calcium-dependent exocytosis during the acrosome reaction of spermatozoa (Michaut *et al*., 2000). Some groups of *RAB* GTPases are specific to certain Embryophyta lineages such *AthRABH1a*, *AthRABH1b*, *AthRABH1c*, *AtRABH1d*, and *AthRABH1e*; there were no gene homologues from any of the fabaceous species. This could mean that these genes were present in the early common ancestors of all Embryophyta and have being deleted in some lineages such as Fabaceae while being retained in some lineages such as *Arabidopsis*. These *Arabidopsis* genes are also much more diverged in their spatial expression patterns; for instance, all of *AthRABH1* group are mainly expressed in stamen and pollen tissues. Overall, even though a common conclusion about the phyletic characteristics of *RAB* genes is that the majority of them are universally conserved with regards to their functional and phyletic characteristics, there are occasional cladistic groups that are comparatively more divergent in terms of their expression, functional, and phyletic characteristics as explained by the above-mentioned examples. Thus, it would be reasonable to assert that some members of Fabaceae-specific *RAB* cladistic groups are more likely to be specialized in nodules to regulate shuttling of materials between symbiont and host. Seven of the 10 genes identified in this study are *RAB* genes, agreeing with this assumption.

There were three genes, *MtrPUT18368*, *LjaTC12838*, and *MtrGB13781173*, that were identified as being possible candidates that are uniquely expressed in nodular structures (see Materials and methods for the details of the analysis used in the identification of these genes), thus implying a likely involvement in the regulation of nodule-specific functions. The logical evolutionary sequence leading to nodular-specific expression would follow the root specificity. In other words, if there are nodule-specific small GTPases, they are likely to have evolved from root-specific small GTPases. Thus, the close phyletic association of putative nodule-specific genes would be an expected result. Further supporting this argument, three other small GTPases identified to be nodule specific, *MtrGB27408074*, *LjaTC18154*, and *MtrPUT45174*, are also closely associated with *Arabidopsis* genes that shown an increased expression in root tissues, and these sequences were also much more diverged in terms of sequence (Fig. 2; Supplementary Table S2 at *JXB* online). The remaining possible nodule-unique small GTPase gene, *MtrIMGAAC1221648.2*, is phyletically associated with *Arabidopsis AthRABC1*, which was universally

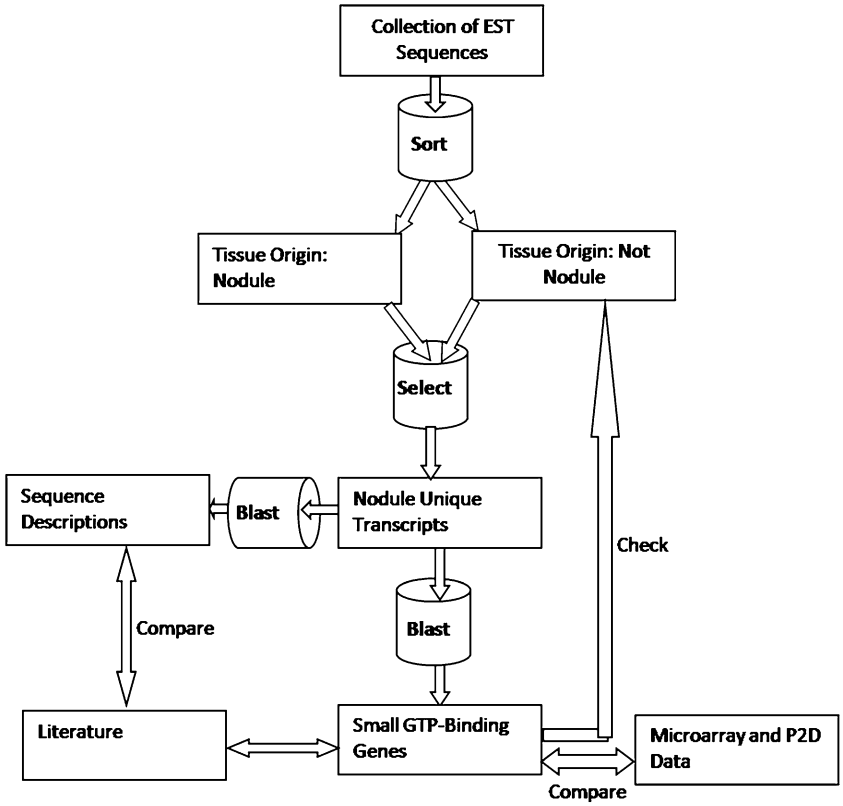

Fig. 5. A schematic drawing illustrating the steps followed to identify small GTP-binding genes that are likely to be mainly expressed in nodular structures. All the sequences from a variety of resources, TC (tentative consensus) sequences, were separated on the basis of the tissue origin of the library as libraries made from nodular structures and others, by utilizing Microsoft Access. These two local databases were compared with each other, and the sequences which are likely to be specifically expressed in nodular structures were determined. The nodule-unique transcripts were blasted against all plant protein databases (downloaded from GenBank) and the minimum 10^{-7} accepted as a threshold for significance; the sequence with the best hit was taken into consideration. The sequences were compared with the sequences reported to be uniquely expressed in nodules. The nodule transcripts were also screened with the small GTP-binding genes collected in this study (see Supplementary Table S1 at *JXB* online for details) in order to separate out the small GTP-binding gene homologues; the identified transcripts were checked again for nodule uniqueness and the sequences compared with the reports from the literature and 2D protein profiling and microarray data available online. The detailed results of this study can be seen in Supplementary Table S2 at *JXB* online.

expressed. In summary, it could be said that possible nodule-unique genes are usually cladistically associated with the genes that have undergone divergence in expression. One other observation was that the majority of the genes that are assigned as being nodule specific have closely associated Fabaceae homologues, which is expected because these genes are likely to have evolved in fabaceous plants in a lineage-specific manner. However, it needs to be clarified that the tissue source of EST sequences used in this study could have been mixed with root tissues, albeit that the origin of the tissue source is specified as nodules; thus, some of these transcripts could be specific to roots rather than nodular structures. An experimental transcriptomal profiling would clarify this ambiguity.

Comparative phyletic analysis of Fabaceae ROP *genes*: The most obvious characteristic of the *ROP* tree was that Embryophyta Rho-like proteins, ROPs, and RHO proteins of metazoa/fungi are clustered in different groups (Fig. 3). This type of phyletic characteristics agrees with previous findings (Vernoud *et al*., 2003). The other typical topological distinction between these two main lineages of Eukaryotae genes is that the plant genes were much better conserved and homogenous relative to metazoa/fungi genes, which were more diversified (Fig. 3). This fact can be interpreted as the plant genes being likely to be functionally more uniform. In fact, the spatial expression pattern of *Arabidopsis ROP* genes is consistent with this conclusion; the majority of Arabidopsis homologues, *AthROP2*, *AthROP3*, *AthROP4*, *AthROP5*, *AthROP6*, *AthROP10*, *AthROP11*, and *AthROP12*, are mostly equally present in all tissues (Fig. 3) even though some of them showed a comparatively increased expression, i.e. *AthROP2* and *AthROP4* in root tissues, and *AthROP3* and *AthROP5* in stamen and pollen tissues. Furthermore, there were some *Arabidopsis* genes that were mainly expressed in a specific tissue, such as *AthROP1* in roots, *AthROP7* and *AthROP9* in regenerative tissues (callus, root hair), and *AthROP8* in callus and siliques. *AthROP1* is very well

conserved across all Embryophyta lineages and its role in pollen tube tip growth is very well known (Chen *et al*., 2003). Some *Arabidopsis* homologues, *AthROP8*, *AthROP6*, and *AthROP7*, did not cluster with any of the Fabaceae *ROP* genes (Fig. 3); this could mean that the common ancestor of all eudicots probably had all these genes and these genes are likely to have been lost in Fabaceae. In fact, in support of this, two plant genomes, *Vitis vinifera* (French–Italian Public Consortium for Grapevine Genome Characterization, 2007) and *Populus trichocarpa* (Tuskan *et al*., 2006), have homologues of *AthROP8* and *AthROP7*, but not *AthROP6* (unpublished data). This could be interpreted as in earlier evolutionary phases of Eurosid II (*Arabidopsis* and *Populus*) both *AthROP7* and *AthROP8* were present, and were deleted in some lineages. The conclusion drawn from these analyses is that *ROP* genes are much more diverged between the main lineages of Embryophyta. In other words, the phyletic groups were homogeneous with regards to the origin of sequences (Fig. 3), i.e. Fabaceae or *Arabidopsis*. This could suggest that *ROP* genes are evolving much faster than other subfamilies of small GTPases.

As a result of the nodule specificity analysis using all available EST sequences for two model legume species, *Medicago* and *Lotus*, a single *ROP* gene was identified, *LjaTC11919* from *Lotus* (Fig. 3), which could be mainly expressed in nodular structures. The expressionally nodule-unique gene from *Lotus* is closely associated with a gene from *Medicago*, *MtrIMGAAC140545-16.1*. If the assumption about the functional nodular specificity of this gene is correct, the existence of very similar homologues of this gene in closely related leguminous species would be expected. *ROP* genes are involved in biological processes that require extensive cytoskeletal rearrangement such as polar growth of root tips. It could be hypothesized that this gene may play a role during the formation of nodular structures, where extensive cytoskeletal reorganization of the host cell occurs. However, this speculation needs to be tested experimentally to reach more decisive conclusions about the putative roles of of *ROP* genes in nodular structures.

Comparative phyletic analysis of Fabaceae RAN *genes*: The *RAN* subfamily of small GTPases is the most strictly conserved set of genes throughout all lineages of plants and animals. These genes are known to play a key role in basic housekeeping functions such as the regulation of nuclear trafficking (Gorlich and Kutay, 1999; Gorlich *et al*., 2003) and the assembly of mitotic spindles (Kalab *et al*., 1999; Dasso, 2001). Despite a high level of conservation, there is some divergence in sequence and expression among *RAN* genes (Fig. 4). For instance, *AthRAN4* is phyletically more diverged in comparison with other *RAN* genes, and it was predominantly expressed in stem tissues (Fig. 5 and Supplementary Table S2 at *JXB* online). *Medicago RAN*, *MtrIMGAAC124971 21.1*, diverged from the other group of Fabaceae *RAN* geses such as *AthRAN4*, meaning that the functional identity of this gene could be distinct. Thus, it would be very interesting to investigate the role of this gene further. *AthRAN2* and *AthRAN1* did not group with any other Fabaceae *RAN* genes obtained in this study. In other words, even though there was no expansion in copy number of these genes in both lineages of Embryophyta, Fabaceae, and Brassicaceae, since the divergence, the rate of sequence divergence of *RAN* genes of these two families could have been different.

Supplementary data

Supplementary data for this manuscript can be found at *JXB* online.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215,** 403–410.

Aoe T, Lee AJ, van Donselaar E, Peters PJ, Hsu VW. 1998. Modulation of intracellular transport by transported proteins. Insight from regulation of COPI-mediated transport. *Proceedings of the National Academy of Sciences, USA* **95,** 1624–1629.

Aridor M, Fish KN, Bannykh S, Weissman J, Roberts TH, Lippincott-Schwartz J, Balch WE. 2001. The Sar1 GTPase coordinates biosynthetic cargo selection with endoplasmic reticulum export site assembly. *Journal of Cell Biology* **152,** 213–229.

Arthur KM, Vejlupkova Z, Meeley RB, Fowler JE. 2003. Maize ROP2 GTPase provides a competitive advantage to the male gametophyte. *Genetics* **165,** 2137–2151.

BarPeled M, Raikhel NV. 1997. Characterization of AtSEC12 and AtSAR1—proteins likely involved in endoplasmic reticulum and Golgi transport. *Plant Physiology* **114,** 315–324.

Bolte S, Schiene K, Dietz KJ. 2000. Characterization of a small GTP-binding protein of the Rab 5 family in *Mesembryanthemum crystallinum* with increased level of expression during early salt stress. *Plant Molecular Biology* **42,** 923–936.

Borg S, Brandstrup B, Jensen TJ, Poulsen C. 1997. Identification of new protein species among 33 different small GTP-binding proteins encoded by cDNAs from *Lotus japonicus*, and expression of corresponding mRNAs in developing root nodules. *The Plant Journal* **11,** 237–250.

Bourne HR, Sanders DA, McCormick F. 1990. The GTPase superfamily—a conserved switch for diverse cell functions. *Nature* **348,** 125–132.

Bourne HR, Sanders DA, McCormick F. 1991. The GTPase superfamily—conserved structure and molecular mechanism. *Nature* **349,** 117–127.

Bucci C, Parton RG, Mather IH, Stunnenberg H, Simons K, Hoflack B, Zerial M. 1992. The small GTPase Rab5 functions as a regulatory factor in the early endocytic pathway. *Cell* **70,** 715–728.

Burd CG, Strochlic TI, Setty SRG. 2004. Arf-like GTPases, not so Arf-like after all. *Trends in Cell Biology* **14,** 687–694.

Cannon SB, Sterck L, Rombauts S, *et al*. 2006. Legume genome evolution viewed through the *Medicago truncatula* and *Lotus*

japonicus genomes. *Proceedings of the National Academy of Sciences, USA* **103,** 14959–14964.

Chapman BA, Bowers JE, Feltus FA, Paterson AH. 2006. Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication. *Proceedings of the National Academy of Sciences, USA* **103,** 2730–2735.

Chavrier P, Gorvel JP, Stelzer E, Simons K, Gruenberg J, Zerial M. 1991. Hypervariable C-terminal domain of Rab proteins acts as a targeting signal. *Nature* **353,** 769–772.

Chen CYH, Cheung AY, Wu HM. 2003. Actin-depolymerizing factor mediates Rac/Rop GTPase-regulated pollen tube growth. *The Plant Cell* **15,** 237–249.

Cheon CI, Lee NG, Siddique ABM, Bal AK, Verma DPS. 1993. Roles of plant homologs of Rab1p and Rab7p in the biogenesis of the peribacteroid membrane, a subcellular compartment formed *de novo* during root-nodule symbiosis. *EMBO Journal* **12,** 4125–4135.

Cheung AY, Chen CY, Glaven RH, de Graaf BHJ, Vidali L, Hepler PK, Wu H-M. 2002. Rab2 GTPase regulates vesicle trafficking between the endoplasmic reticulum and the Golgi bodies and is important to pollen tube growth. *The Plant Cell* **14,** 945–962.

Cronk Q, Ojeda I, Pennington RT. 2006. Legume comparative genomics: progress in phylogenetics and phylogenomics. *Current Opinion in Plant Biology* **9,** 99–103.

Dasso M. 2001. Running on ran: nuclear transport and the mitotic spindle. *Cell* **104,** 321–324.

de Graaf BHJ, Cheung AY, Andreyeva T, Levasseur K, Kieliszewski M, Wu H-M. 2005. Rab11 GTPase-regulated membrane trafficking is crucial for tip-focused pollen tube growth in tobacco. *The Plant Cell* **17,** 2564–2579.

Fedorova M, van de Mortel J, Matsumoto PA, Cho J, Town CD, VandenBosch KA, Gantt JS, Vance CP. 2002. Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. *Plant Physiology* **130,** 519–537.

French–Italian Public Consortium for Grapevine Genome Characterization. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449,** 463–467.

Frugoli J, Harris J. 2001. Medicago truncatula on the move! *The Plant Cell* **13,** 458–463.

Fu Y, Li H, Yang ZB. 2002. The ROP2 GTPase controls the formation of cortical fine F-actin and the early phase of directional cell expansion during Arabidopsis organogenesis. *The Plant Cell* **14,** 777–794.

Garcia-Ranea JA, Valencia A. 1998. Distribution and functional diversification of the Ras superfamily in *Saccharomyces cerevisiae*. *FEBS Letters* **434,** 219–225.

Gaynor EC, Chen CY, Emr SD, Graham TR. 1998. ARF is required for maintenance of yeast Golgi and endosome structure and function. *Molecular Biology of the Cell* **9,** 653–670.

Gorlich D, Kutay U. 1999. Transport between the cell nucleus and the cytoplasm. *Annual Review of Cell and Developmental Biology* **15,** 607–660.

Gorlich D, Seewald MJ, Ribbeck K. 2003. Characterization of Ran-driven cargo transport and the Ran GTPase system by kinetic measurements and computer simulation. *EMBO Journal* **22,** 1088–1100.

Grebe M, Xu J, Mobius W, Ueda T, Nakano A, Geuze HJ, Rook MB, Scheres B. 2003. Arabidopsis sterol endocytosis involves actin-mediated trafficking via ARA6-positive early endosomes. *Current Biology* **13,** 1378–1387.

Gu Y, Vernoud V, Fu Y, Yang ZB. 2003. ROP GTPase regulation of pollen tube growth through the dynamics of tip-localized F-actin. *Journal of Experimental Botany* **54,** 93–101.

He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169,** 1157–1164.

Jekely G. 2003. Small GTPases and the evolution of the eukaryotic cell. *Bioessays* **25,** 1129–1138.

Jiang SY, Ramachandran S. 2006. Comparative and evolutionary analysis of genes encoding small GTPases and their activating proteins in eukaryotic genomes. *Physiological Genomics* **24,** 235–251.

Jung YH, Agrawal GK, Rakwal R, *et al*. 2006. Functional characterization of OsRacB GTPase—a potentially negative regulator of basal disease resistance in rice. *Plant Physiology and Biochemistry* **44,** 68–77.

Kalab P, Pu RT, Dasso M. 1999. The Ran GTPase regulates mitotic spindle assembly. *Current Biology* **9,** 481–484.

Kotzer A, Brandizzi F, Neumann U, Paris N, Moore I, Hawes C. 2004. AtRabF2b (Ara7) acts on the vacuolar trafficking pathway in tobacco leaf epidermal cells. *Journal of Cell Sciences* **117,** 6377–6389.

Kumar S, Tamura K, Nei M. 2003. MEGA3: an integrated software for molecular evolutionary genetic analysis and sequence alignment. *Integrative and Comparative Biology* **43,** 947–947.

Latijnhouwers M, Hawes C, Carvalho C, Oparka K, Gillingham AK, Boevink P. 2005. An Arabidopsis GRIP domain protein locates to the trans-Golgi and binds the small GTPase ARL1. *The Plant Journal* **44,** 459–470.

Lee MCS, Orci L, Hamamoto S, Futai E, Ravazzola M, Schekman R. 2005. Sar1p N-terminal helix initiates membrane curvature and completes the fission of a COPII vesicle. *Cell* **122,** 605–617.

Lee MH, Min MK, Lee YJ, Jin JB, Shin DH, Kim DH, Lee KH, Hwang I. 2002. ADP-ribosylation factor 1 of Arabidopsis plays a critical role in intracellular trafficking and maintenance of endoplasmic reticulum morphology in Arabidopsis. *Plant Physiology* **129,** 1507–1520.

Lin CY, Huang PH, Liao WL, Cheng HJ, Huang CF, Kuo JC, Patton WA, Massenburg D, Moss J, Lee FJS. 2000. ARL4, an ARF-like protein that is developmentally regulated and localized to nuclei and nucleoli. *Journal of Biological Chemistry* **275,** 37815–37823.

Manthey K, Krajinski F, Hohnjec N, Firnhaber C, Puhler A, Perlick AM, Kuster H. 2004. Transcriptome profiling in root nodules and arbuscular mycorrhiza identifies a collection of novel genes induced during *Medicago truncatula* root endosymbioses. *Molecular Plant-Microbe Interactions* **17,** 1063–1077.

Mazel A, Leshem Y, Tiwari BS, Levine A. 2004. Induction of salt and osmotic stress tolerance by overexpression of an intracellular vesicle trafficking protein AtRab7 (AtRabG3e). *Plant Physiology* **134,** 118–128.

Memon AR. 2004. The role of ADP-ribosylation factor and SARI in vesicular trafficking in plants. *Biochimica et Biophysica Acta* **1664,** 9–30.

Merchant SS, Prochnik SE, Vallon O, *et al*. 2007. The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* **318,** 245–250.

Michaut M, Tomes CN, De Blas G, Yunes R, Mayorga LS. 2000. Calcium-triggered acrosomal exocytosis in human spermatozoa requires the coordinated activation of Rab3A and N-ethylmaleimide-sensitive factor. *Proceedings of the National Academy of Sciences, USA* **97,** 9996–10001.

Molendijk AJ, Bischoff F, Rajendrakumar CSV, Friml J, Braun M, Gilroy S, Palme K. 2001. *Arabidopsis thaliana* Rop GTPases are localized to tips of root hairs and control polar growth. *EMBO Journal* **20,** 2779–2788.

Morino K, Shimamoto K, Umemura K, Iwata M, Kawata M. 2004. The rice small GTPase OsRac1, a regulator of the programmed cell death in plant disease resistance, functions during development. *Plant and Cell Physiology* **45,** S136–S136.

Munro S. 2005. The Arf-like GTPase Arl1 and its role in membrane traffic. *Biochemical Society Transactions* **33,** 601–605.

Nakano A, Muramatsu M. 1989. A novel GTP-binding protein, Sar1p, is involved in transport from the endoplasmic-reticulum to the Golgi-apparatus. *Journal of Cell Biology* **109,** 2677–2691.

Nielsen EE, Santos-Serna J, Schmitz AJ. 2002. The plant Rab GTPase, AtRabA4b, localizes to the tips of growing root hairs in *Arabidopsis thaliana*. *Molecular Biology of the Cell* **13,** 504A–505A.

Novick P, Brennwald P. 1993. Friends and family—the role of the Rab GTPases in vesicular traffic. *Cell* **75,** 597–601.

Ono E, Wong HL, Kawasaki T, Hasegawa M, Kodama O, Shimamoto K. 2001. Essential role of the small GTPase Rac in disease resistance of rice. *Proceedings of the National Academy of Sciences, USA* **98,** 759–764.

Preuss ML, Schmitz AJ, Thole JM, Bonner HKS, Otegui MS, Nielsen E. 2006. A role for the RabA4b effector protein PI-4K beta 1 in polarized expansion of root hair cells in *Arabidopsis thaliana*. *Journal of Cell Biology* **172,** 991–998.

Preuss ML, Serna J, Falbel TG, Bednarek SY, Nielsen E. 2004. The Arabidopsis Rab GTPase RabA4b localizes to the tips of growing root hair cells. *The Plant Cell* **16,** 1589–1603.

Rensing SA, Lang D, Zimmer AD, *et al*. 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* **319,** 64–69.

Rutherford S, Moore I. 2002. The Arabidopsis Rab GTPase family: another enigma variation. *Current Opinion in Plant Biology* **5,** 518–528.

Schiene K, Donath S, Brecht M, Puhler A, Niehaus K. 2004. A Rab-related small GTP binding protein is predominantly expressed in root nodules of *Medicago sativa*. *Molecular Genetics and Genomics* **272,** 57–66.

Shin HW, Kobayashi H, Kitamura M, Waguri S, Suganuma T, Uchiyama Y, Nakayama K. 2005. Roles of ARFRP1 (ADP-ribosylation factor-related protein 1) in post-Golgi membrane trafficking. *Journal of Cell Science* **118,** 4039–4048.

Son O, Yang HS, Lee HJ, Lee MY, Cheon CI, *et al*. 2003. Expression of srab7 and SCaM genes required for endocytosis of Rhizobium in root nodules. *Plant Science* **165,** 1239–1244.

Stefano G, Renna L, Hanton SL, Chatre L, Haas TA, Brandizzi F. 2006. ARL1 plays a role in the binding of the GRIP domain of a peripheral matrix protein to the Golgi apparatus in plant cells. *Plant Molecular Biology* **61,** 431–449.

Takeuchi M, Ueda T, Sato K, Abe H, Nagata T, Nakano A. 2000. A dominant negative mutant of Sar1 GTPase inhibits protein transport from the endoplasmic reticulum to the Golgi apparatus in tobacco and Arabidopsis cultured cells. *The Plant Journal* **23,** 517–525.

Takeuchi M, Ueda T, Yahara N, Nakano A. 2002. Arf1 GTPase plays roles in the protein traffic between the endoplasmic reticulum and the Golgi apparatus in tobacco and Arabidopsis cultured cells. *The Plant Journal* **31,** 499–515.

Thompson J, Gibson T, Plewniak F, Jeanmougin F, Higgins D. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25,** 4876–4882.

Tuskan GA, DiFazio S, Jansson S, *et al*. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313,** 1596–1604.

Vernoud V, Horton AC, Yang Z, Nielsen E. 2003. Analysis of the small GTPase gene superfamily of Arabidopsis. *Plant Physiology* **131,** 1191–1208.

Wienkoop S, Saalbach G. 2003. Proteome analysis. Novel proteins identified at the peribacteroid membrane from *Lotus japonicus* root nodules. *Plant Physiology* **131,** 1080–1090.

Winge P, Brembu T, Kristensen R, Bones AM. 2000. Genetic structure and evolution of RAC-GTPases in *Arabidopsis thaliana*. *Genetics* **156,** 1959–1971.

Wong HL, Sakamoto T, Kawasaki T, Umemura K, Shimamoto K. 2004. Down-regulation of metallothionein, a reactive oxygen scavenger, by the small GTPase OsRac1 in rice. *Plant Physiology* **135,** 1447–1456.

Yang X, Tuskan GA, Cheng MZ-M. 2006. Divergence of the Dof gene families in poplar, Arabidopsis, and rice suggests multiple modes of gene evolution after duplication. *Plant Physiology* **142,** 820–830.

Young ND, Mudge J, Ellis TN. 2003. Legume genomes: more than peas in a pod. *Current Opinion in Plant Biology* **6,** 199–204.

Zheng H, Camacho L, Wee E, Batoko H, Legen J, Leaver CJ, Malho r Hussey PJ, Moore I. 2005. A Rab-E GTPase mutant acts downstream of the Rab-D subclass in biosynthetic membrane traffic to the plasma membrane in tobacco leaf epidermis. *The Plant Cell* **17,** 2020–2036.

Zhu H, Riely BK, Burns NJ, Ane J-M. 2006. Tracing nonlegume orthologs of legume genes required for nodulation and arbuscular mycorrhizal symbioses. *Genetics* **172,** 2491–2499.

Zimmermann P, Hennig L, Gruissem W. 2005. Gene-expression analysis and network discovery using Genevestigator. *Trends in Plant Science* **10,** 407–409.