Plant Mol Biol (2010) 73:673–685
DOI 10.1007/s11103-010-9648-x

Trichome specific expression of the tobacco (*Nicotiana sylvestris*) cembratrien-ol synthase genes is controlled by both activating and repressing *cis*-regions

Hanane Ennajdaoui · Gilles Vachon · Cécile Giacalone · Isabelle Besse · Christophe Sallaud · Michel Herzog · Alain Tissier

Received: 26 January 2010 / Accepted: 1 May 2010 / Published online: 21 May 2010

Abstract Tobacco (*Nicotiana sylvestris*) glandular trichomes make an attractive target for isoprenoid metabolic engineering because they produce large amounts of one type of diterpenoids, α- and β-cembratrien-diols. This article describes the establishment of tools for metabolic engineering of tobacco trichomes, namely a transgenic line with strongly reduced levels of diterpenoids in the exudate and the characterization of a trichome specific promoter. The diterpene-free tobacco line was generated by silencing the major tobacco diterpene synthases, which were found to be encoded by a family of four highly similar genes (*NsCBTS-2a*, *NsCBTS-2b*, *NsCBTS-3* and *NsCBTS-4*), one of which is a pseudogene. The promoter regions of all four CBTS genes were sequenced and found to share over 95% identity between them. Transgenic plants expressing *uidA* under the control of the *NsCBTS-2a* promoter displayed a specific pattern of GUS expression restricted exclusively to the glandular cells of the tall secretory trichomes. A series of sequential and internal deletions of the *NsCBTS-2a* promoter led to the identification of two *cis*-acting regions. The first, located between positions -589 to -479 from the transcription initiation site, conferred a broad transcriptional activation, not only in the glandular cells, but also in cells of the trichome stalk, as well as in the leaf epidermis and the root. The second region, located between positions -279 to -119, had broad repressor activity except in trichome glandular cells and is mainly responsible for the specific expression pattern of the *NsCBTS-2a* gene. These results establish the basis for the identification of *trans*-regulators required for the expression of the *CBTS* genes restricted to the secretory cells of the glandular trichomes.

Electronic supplementary material The online version of this article (doi:10.1007/s11103-010-9648-x) contains supplementary material, which is available to authorized users.

H. Ennajdaoui · M. Herzog
Laboratoire d'Ecologie Alpine, Université Joseph Fourier and CNRS—Unité Mixte de Recherche 5553, 2233, rue de la piscine, BP 53, 38041 Grenoble Cedex 9, France

G. Vachon
Laboratoire de Physiologie Cellulaire Végétale, Université Joseph Fourier—Unité Mixte de Recherche 5168 CNRS/CEA/INRA/UJF—CEA, 17 rue des Martyrs—bâtiment C2, 38054 Grenoble Cedex 9, France

C. Giacalone · I. Besse · C. Sallaud · A. Tissier
Librophyt, Centre de Cadarache, Bâtiment 185, 13115 Saint Paul-lez-Durance, France

Present Address:
H. Ennajdaoui
Molecular Cell & Developmental Biology Department, Sinsheimer Laboratories, 1156 High Street, Santa Cruz, CA 95060, USA

A. Tissier (✉)
Institut de Recherche pour le Développement, Université Montpellier 2—Unité Mixte de Recherche Diversité et Adaptation des Plantes Cultivées, 911, avenue d'Agropolis, B.P. 64501, 34394 Montpellier Cedex 5, France
e-mail: alain.tissier@univ-montp2.fr

Present Address:
C. Giacalone
Laboratoire Echanges Membranaires et Signalisation, Institut de Biologie Environnementale et de Biotechnologie, Commissariat à l'Energie Atomique, 13108 Saint Paul Lez Durance Cedex, France

Present Address:
C. Sallaud
Functional and Applied Cereal Group, Biogemma, Z.I. du Brezet, 8, rue des Frères Lumière, 63028 Clermont-Ferrand Cedex 2, France

Keywords Glandular trichomes · Promoter · Terpene biosynthesis · *cis*-regulatory regions · Tobacco · Intron hairpin

Introduction

Glandular trichomes are specialized organs located on the surface of the aerial parts of many higher plant species. They are the site of production of abundant secondary metabolites, which in certain species, such as tobacco, may represent up to 15% of the leaf dry weight (Wagner et al. 2004). This huge contribution to leaf biomass relative to the actual volume of the glandular trichomes suggests that massive metabolic fluxes are likely to occur in these structures, a feature which makes them an attractive target for metabolic engineering (Wagner 1991; Lange and Croteau 1999; Mahmoud and Croteau 2002; Wagner et al. 2004; Schilmiller et al. 2008). Because trichomes are present on the surface of leaves and stems, their secretion constitutes a potential barrier to plant pests and metabolic engineering could be used to modify the leaf surface chemicals to increase plant defence against various pests. Another potential application of metabolic engineering in the glandular trichomes is the high level production of valuable but otherwise difficult to produce secondary metabolites, such as pharmaceutical ingredients like the anticancer compound taxol. However, to reach these goals, it is necessary to gain more basic knowledge about glandular trichomes, and in particular to understand how gene expression is regulated in these organs.

Plant glandular trichomes are present in several important plant families such as the Solanaceae, Asteraceae, Lamiaceae or Cannabaceae. The nature of the compounds which are produced by these trichomes is typical of a species or a group of species. For example, the peltate trichomes of many species from the Lamiaceae family produce volatile compounds such as monoterpenes, sesquiterpenes or phenylpropanoids. Among the Solanaceae, a wide variety of compounds are produced by glandular trichomes in a species specific fashion. For example, the wild tobacco species *Nicotiana sylvestris* has secretory glandular trichomes (SGTs) which mainly produce large amounts of α- and β-cembratrien-diols (CBT-diols), macrocyclic diterpenoids of the cembrane type that are also encountered in many varieties of cultivated tobacco (*Nicotiana tabacum*) such as the Burley or Virginia types (Keene and Wagner 1985). Because it is diploid, has a simple trichome metabolic profile consisting mainly of CBT-diols and does not cross spontaneously to *N. tabacum*, thus eliminating the risk of contamination of a cultivated species, *Nicotiana sylvestris* was chosen as a model system for studies of glandular trichomes and metabolic engineering of isoprenoid biosynthesis. In addition, sequences from *N. tabacum* which originate from *N. sylvestris* are usually highly conserved with typically over 99% identity at the nucleotide level. Thus, the available genomic sequences and ESTs of *N. tabacum* should provide valuable sequence information for studies on Nicotianae species and on *N. sylvestris* in particular.

Plant diterpenoid biosynthesis proceeds in the plastids from geranylgeranyl diphosphate (GGPP), itself synthesized from isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) produced by the plastidic deoxyxylulose phosphate (DXP) pathway. In *Nicotiana tabacum*, gene candidates for the biosynthesis of the CBT-diols were recently identified (Wang et al. 2001; Wang and Wagner 2003) (Fig. 1). The first step was proposed to be catalyzed by cembratrien-ol synthase(s) (CBTS), leading to a mixture of α- and β-cembratrien-ols. The second step would then be a site specific oxidation of the CBT-ols by a cytochrome P450 oxygenase. cDNA sequence for a candidate *CBTS* was identified and silencing this gene led to reduced amounts of CBT-diols and CBT-ols (Wang and Wagner 2003), thus suggesting that this gene indeed codes for a CBT-ol synthase. However, no demonstration of CBTS activity was provided leaving the question open as to what enzymatic activity is exactly encoded by this gene. Similarly, a P450-encoding cDNA (*CYP71D16*) expressed in tobacco glandular trichomes was identified and its down-regulation *in planta* by gene silencing resulted in the

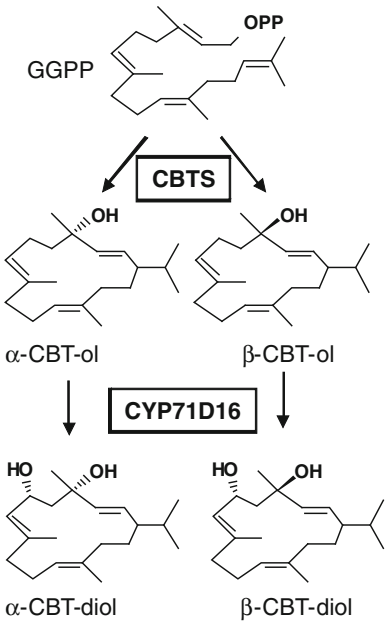


Fig. 1 Biosynthesis pathway of the CBT-diols in *Nicotiana sylvestris* trichomes. Geranylgeranyl diphosphate (GGPP) is utilized by the cembratrien-ol synthases (CBTS) which are encoded by a multigene family (this article). The next step is a site and regio- specific hydroxylation which is catalyzed by a P450 oxygenase (CYP71D16)

accumulation of cembratrien-ols (Wang et al. 2001). Interestingly, plants overproducing CBT-ols exhibited increased protection against spider-mites, most likely due to the toxicity of CBT-ols towards these insects (Wang et al. 2001, 2004). Furthermore, 1.75 kb of the *CYP71D16* promoter were isolated and shown to be sufficient to direct expression of the GUS reporter gene specifically to the trichome glandular cells in tobacco (Wang et al. 2002).

For successful metabolic engineering of tobacco trichomes, we reasoned that two main tools should be made available. The first is tobacco lines devoid of endogenous diterpenoids. This will help identify novel terpenoids produced and avoid interference between the endogenous and the heterologous compounds. In addition, one could hope that this would increase the pool of precursors, i.e. GGPP, and result in higher levels of production. Second, several trichome specific promoters are necessary and a detailed understanding of the transcriptional regulation in the glandular cells are required to take full advantage of these cell factories. We thus set out to identify the gene(s) encoding the CBTS activity and characterize the corresponding promoters. In this work, a four-member gene family from *N. sylvestris* was found to encode the trichome specific CBTS enzymes. The promoters of those genes were identified and shown to be highly similar to each other. One of those promoters was further characterized and *cis*-acting regions defining the spatial specificity of expression could be identified.

Materials and methods

Oligonucleotide primers and probes

A complete list of oligonucleotide primers and probes used in this study is available in the Supplemental data file.

Plant material

Nicotiana sylvestris seeds were germinated directly in standard commercial soil. Plants were grown in a growth room at 25°C ± 1°C and 60 µmol photons/m^2 s^{-1} illumination under a 16 h light/8 h dark regime. Lighting was provided by two types of fluorescent tubes (OSRAM L 36 W/77 and OSRAM Biolux 36 W/965) in a 1:1 ratio. Humidity in the growth room was set at 60%.

Nucleic acid preparation

Total plant DNA was extracted as in Sallaud et al. (2009). Total RNA isolation was achieved from 50 mg of leaves using the RNeasy® kit (Qiagen) according to the manufacturer's instruction. To remove traces of genomic DNA contamination, RNA samples were treated with a DNAse (Turbo DNAse-free™, Ambion) according to the manufacturer's instruction. For RT–PCR, cDNAs were synthesized from 1 µg of total RNA with the Transcriptor Reverse Transcriptase kit (Roche) using oligodT or random primers. For the total RNA extraction from trichomes, trichomes were collected from a dozen young leaves (approximately 10 cm long) by freezing the leaves lying on a sheet of aluminium itself floating at the surface of liquid nitrogen. Once frozen, the leaves were brushed with a paint brush and the trichomes collected in a mortar. After grinding, the sample was treated as for the leaves and RNA extracted with the RNeasy® kit (Qiagen).

Phylogenetic analysis

Protein sequences were aligned using the ClustalW2 program at the European Bioinformatics Institute (http://www.ebi.ac.uk/tools/clustalw2) with the default matrix. Phylogeny was inferred using the maximum likelihood method from the PHYLIP package (http://evolution.genetics.washington.edu/phylip.html), where a consensus tree was derived from the most parsimonious trees out of 1,000 data sets. The complete set of sequences used for the construction of the phylogenetic tree is available as a table in the Supplemental Data file.

ihpCBTS constructs

The Cauliflower Mosaic virus 35S promoter and the nopaline synthase terminator were amplified with primers 35Sfor and 35Srev, and tNosFor and tNosRev, respectively, from the pCambia1301 T-DNA vector. The complete description of pCambia plasmids is available online at http://www.cambia.org/daisy/cambia/585#dsy585_Description. The 35S promoter fragment was digested by SacI and BamHI and cloned into the pBluescript II SK+ plasmid digested with these enzymes. Similarly the Nos terminator fragment was cloned between the XhoI and XbaI sites of the pBluescript II SK+ plasmid. A genomic fragment of the *NsCBTS-2a* gene encompassing exon 2 and intron 2 was amplified with primers CBTS_ex2_int2For and CBTS_ex2_int2Rev, digested with BamHI and ScaI and cloned into pBluescript II SK+. Similarly, exon 2 of the CBTS-2a gene was amplified with primers CBTS_ex2_For and CBTS-ex2_Rev, digested with ScaI and XhoI and cloned into pBluescript II SK+. All these fragments were excised by the enzymes which were used to clone them into pBluescript II SK+, gel-purified, mixed and ligated together with pBluescript II SK+ digested with SacI and XbaI to make the 35S:ihpCBTS construct into the pBluescript II SK+. The 35S:ihp*CBTS* construct was then recovered by digestion of this plasmid with SacI and XbaI and ligated into the

pCambia5300 T-DNA vector digested with the same enzymes to afford the 35S:ihp*CBTS* T-DNA vector. To make the pCBTS1.7:ihp*CBTS* T-DNA vector, the 35S fragment from the 35S:ihpCBTS-pBluescript II SK+ was excised by SacI and BamHI and replaced with the 1.7 kb fragment of the *CBTS-2a* promoter amplified with primers pCBTS1.7For and pCBTS1.7Rev and digested with SacI and BamHI. The resulting pCBTS1.7:ihpCBTS in pBluescript II SK+ was then digested with SacI and BamHI to release the pCBTS1.7:ihpCBTS fragment, which was then cloned into the pCambia5300 T-DNA vector digested with the same enzymes.

Promoter deletions-GUS constructs

Two *NsCBTS2a* promoter fragments of 1.7 and 1.09 kb in length respectively were isolated by PCR and cloned into the pKGWFS7 T-DNA vector containing the uidA coding sequence.

Deletions

The *NsCBTS2a* promoter (pCBTS1.09) was amplified by PCR to add attB1 & attB2 Gateway recombination sites together with a XhoI site placed in between the attB1 and the promoter and a PacI site in between the promoter and attB2 site. The amplicon was cloned by BP reaction into the pDONR221 donor vector (Invitrogen). The resulting entry clone was used to insert the promoter into the destination vector pKGWFS7 (Plant System Biology VIB-Ghent University, Belgium) upstream of GFP:GUS coding sequences resulting in the pKGWFS7:pCBTS1.09 expression clone.

All deletions of the promoter were PCR-amplified and directly cloned into pKGWFS7 using XhoI and PacI sites. Internal deletions were obtained by first amplifying the regions upstream and downstream of the deletion and by fusing the two resulting amplicons by PCR-based overlap amplification. All constructs were checked by sequencing before transformation.

-46 construct

The -46 portion of the 35S promoter was amplified by PCR to add a XhoI and AvrII sites upstream of the minimum promoter and an omega traductional enhancer sequence together with a PacI site downstream of the minimum promoter. The resulting amplicon was cloned in XhoI and PacI sites of pKGWFS7:pCBTS thus replacing the *CBTS* promoter by the -46 35S promoter resulting into the pKGWFS7Xho/AvrII-46 vector. The 4c fragment was PCR-amplified to add XhoI and AvrII sites at the 5′ and 3′ sides, respectively and cloned individually into the XhoI and AvrII sites of the vector resulting into the pCBTS400c:p35Smin:*uidA* and pCBTS4c:p35Smin:*uidA*, respectively. All constructs were checked by sequencing before transformation.

Exudate extraction and analysis by GC–MS

Analyses were performed on ten 1 cm diameter leaf disks per plant. The disks were weighted and 10 µL of a 35 mM β-caryophyllene solution (i.e. 71.4 µg of β-caryophyllene) was added to the samples as well as 1 mL of pentane. After vortexing, the samples were extracted for 3 min. 1 mL of this extract was then injected directly into a 6890N gas chromatograph coupled to a 5973N mass spectrometer (Agilent Technologies). Separation was assured with a 30 m × 0.25-mm diameter with 0.25-µm film of HP-1 ms (Agilent Technologies). Samples were injected using a cool on-column injector at 40°C. The oven was programmed to start at 40°C and then increased by 10°C/min to 100°C, followed by 3°C/min to 280°C, with a 5-min hold. Electronic impact was recorded at 70 eV. Under these conditions, the stereoisomers of CBT-ol (α and β) and CBT-diol (α and β) are not separated.

N. sylvestris transformation and histochemical analysis of GUS expression

Transgenic *Nicotiana sylvestris* were obtained as described earlier (Rontein et al. 2008) from leaf disks of tissue cultured plants co-cultivated in a LBA4404 (for the ihpCBTS constructs) or GV3101 (for the promoter deletion constructs) *Agrobacterium tumefaciens* suspension. The number of independent transgenic lines recovered for the different constructs and analyzed is provided in the supplemental data file.

Expression of the GUS gene was detected by histochemical staining of plantlets or young leaves incubated 12 h with GUS substrate and destained as described (Gallagher 1992). Destained samples were mounted in 20% glycerol and observed under microscope (Eclipse E600, *Nikon*) or binocular glasses (SZX12, *Olympus*) and photographed with a CCD camera (DP70 *Olympus*). Histochemical GUS staining was performed on T1 or T2 plants.

Results

Identification of a gene family encoding CBT-ol synthase related enzymes

Several primers (see "Materials and methods" section) matching the CBT-ol cyclase gene (*NtCYC-1*) from *N. tabacum* identified previously (Wang and Wagner 2003) were used to amplify cDNA fragments from *N. sylvestris*

trichome specific total RNA (see "Materials and methods"). Direct sequencing of the RT–PCR fragments revealed the existence of multiple sequences. Individual DNA molecules were cloned and sequenced. Specific primers to those sequences were designed and four distinct cDNAs could be identified. We named the corresponding genes *NsCBTS-2a*, *NsCBTS-2b*, *NsCBTS-3*, and *NsCBTS-4*. The complete cDNA sequences of these genes were determined and further analyzed. They are highly similar to each other with pairwise percentages of nucleotide identity ranging from 86%—between *NtCYC-1* and the *NsCBTS* sequences—to 99% between *NsCBTS-2a* and *NsCBTS-2b* (Table 1). The proteins encoded by *NsCBTS-2a*, *NsCBTS-2b* and *NsCBTS-3* are 598 amino acid long against 597 for NtCYC-1. The *NsCBTS-4* cDNA is shorter with several small deletions interspersed throughout the gene and a single substitution at position 259 in the cDNA introducing an in frame premature stop codon. The resulting peptide encoded by *NsCBTS-4* is only 87 amino acids long and is unlikely to have any terpene synthase activity. Thus, *NsCBTS-4* may be considered a pseudogene.

Interestingly, BLAST searches indicate that the CBTS proteins are most closely related to the tomato sesquiterpene synthases SSTLH1, SSTLH2, SSTLE1 and SSTLE2 which are specifically expressed in type VI secretory trichomes (Van der Hoeven et al. 2000). A phylogenetic tree illustrates these relationships (Fig. 2) and indicates that the tomato and tobacco trichome specific terpene synthases constitute a distinct branch of the Tps-a group of terpene synthases, as per the nomenclature established previously (Bohlmann et al. 1998; Martin et al. 2004). In contrast, other Tps-a terpene synthases of the Solanaceae, such as the 5-epi-aristolochene synthases from tobacco (NtEAS) or pepper (CaEAS) are positioned on a different branch. This suggests that there is a distinct lineage within the terpene synthases of the Solanaceae which gave rise to the trichome specific enzymes.

The *NsCBTS* promoter is specific of the trichome glandular cells

In the absence of genomic sequence information, we set out to identify the sequence of the promoter region of the *NsCBTS* genes by a technique called adapter anchor PCR (Siebert et al. 1995). By using anchor-primers specific to each of the *NsCBTS* gene and positioned on the first exon, the corresponding promoters could be isolated. 1.09 kb of all four promoters could thus be determined, as well as 1.7 kb of the *NsCBTS2a* promoter. The promoter sequences of the *NsCBTS* genes are highly similar to each other with only a few polymorphisms between them (Fig. 3). To probe the expression profile of these genes *in planta*, promoter:*uidA* fusions were constructed with the 1.7 and 1.09 kb fragments of the *NsCBTS-2a* gene. These fusions were then cloned into a T-DNA vector and the resulting plasmids used to transform *N. sylvestris* by Agrobacterium mediated leaf-disc transformation. Several transgenic lines were recovered for each construct and single copy transformants were selected for further analysis. GUS staining of leaves of these lines consistently showed a highly distinct profile with only the trichome secretory cells of the tall glandular trichomes showing the characteristic GUS blue staining (Fig. 4). This indicated that the 1.09 kb fragment was sufficient to confer the tissue specificity and could thus be used for further study to identify *cis* elements involved in transcriptional regulation.

Table 1 Percentage of identical amino acids shared between CBTS proteins

	Nt CYC1	Ns CBTS-2a	Ns CBTS-2b	Ns CBTS-3
Nt CYC1	100	86	86	86
Ns CBTS-2a		100	99	94
Ns CBTS-2b			100	94
Ns CBTS-3				100

Intron hairpin gene silencing targeting the *NsCBTS* multigene family

As mentioned above (see "Introduction"), one of our objectives was to generate a diterpene-free tobacco line which would provide a better background for metabolic engineering of terpenoids in the trichomes. The presence of several closely related *NsCBTS* genes and the high level of sequence identity between the promoters of these genes suggested that they, like *NsCBTS-2a*, may all be expressed in the trichomes and collectively contribute to the biosynthesis of the mixture of alpha and beta-cembratrienols. Therefore, a gene silencing approach was adopted to down-regulate their expression and generate diterpene-free *N. sylvestris* lines. Post-transcriptional gene silencing is highly specific but nonetheless tolerates low levels of nucleotide substitutions. Analysis of a multiple alignment of the *NsCBTS* coding sequences showed that the second exon was the fragment with the lowest level of dissimilarity between the three *NsCBTS* genes. This fragment was thus used to make an intron hairpin construct. Intron hairpin gene silencing constructs have been showed previously to be successful in achieving stable, highly efficient and tissue specific gene silencing (Smith et al. 2000; Stoutjesdijk et al. 2002). The second exon, second intron and, to ensure efficient splicing, a few bases of the third exon of *NsCBTS-2a* were cloned upstream of the second exon in antisense orientation. This cassette was cloned downstream of either the Cauliflower Mosaic Virus 35S promoter (35S:ihp*CBTS*) or the 1.7 kb fragment of the

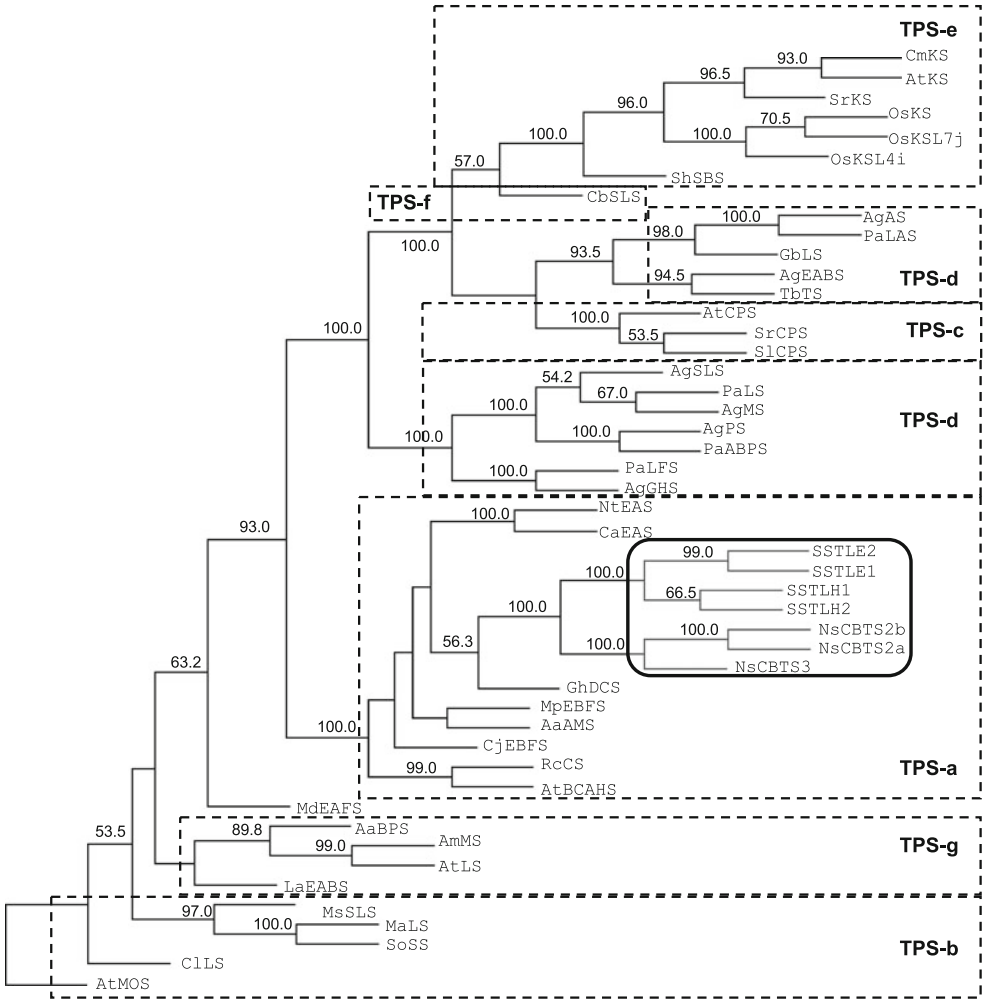


Fig. 2 Unrooted phylogenetic tree of plant TPS including the tobacco CBTS. The nomenclature of TPS groups is as described in Martin et al. (2004). The box in the TPS-a group indicates trichome specific sequences of TPS from the Solanaceae, namely CBTS for tobacco and SSTL from tomato. The complete list of sequences is provided in the Supplemental Data. Figures at nodes indicate percentage values (above 50%) from the bootstrap sampling (1,000 replicates)

NsCBTS-2a promoter (pCBTS1.7:ihp*CBTS*) and upstream of the *Nos* terminator. These two constructs were cloned in T-DNA vectors which were used to generate transgenic *N. sylvestris* lines via *Agrobacterium tumefaciens* mediated transformation. Nine and ten transgenic lines were thus obtained for the 35S:ihp*CBTS* and pCBTS1.7:ihp*CBTS* constructs respectively and were further analyzed. The exudate of the first generation transformants was analyzed by GC–MS. The results are summarized in Fig. 5. Most transgenics carrying the pCBTS1.7:ihp*CBTS* construct showed highly reduced amounts of CBT-diols (less than 10% of the wild type content) while only one 35S:ihp*CBTS* transgenic showed similar reduction in the levels of CBT-diols. Typical GC–MS chromatograms of the wild type and a pCBTS1.7:ihp*CBTS* line are shown in Fig. 5b, c. The transgenic lines expressing the pCBTS1.7:ihp*CBTS* construct displayed no apparent difference in their growth and development compared to wild type plants in our growing conditions. However, upon transforming these lines by a second T-DNA, we noted that the frequency and efficiency of transformation in *CBTS* silenced lines were higher than with wild type plants (data not shown, to be published separately) possibly because of the bacteriostatic activity of the CBT-diols.

Silencing the *CBTS* genes does not affect the trichome specific expression of the pCBTS:GUS fusions

To determine whether the *CBTS* genes were feed-back regulated by CBT-diols, 1.7 kb and 1.09 pCBTS fusions were cloned into another T-DNA vector, pKGWFS7 and transformed into the ihpCBTS line. In addition to the *uidA* reporter gene, this vector carries a kanamycin resistance marker which is different from the hygromycin resistance gene on the ihp*CBTS* T-DNA constructs. These new T-DNA vectors were used to transform the CBTS silenced

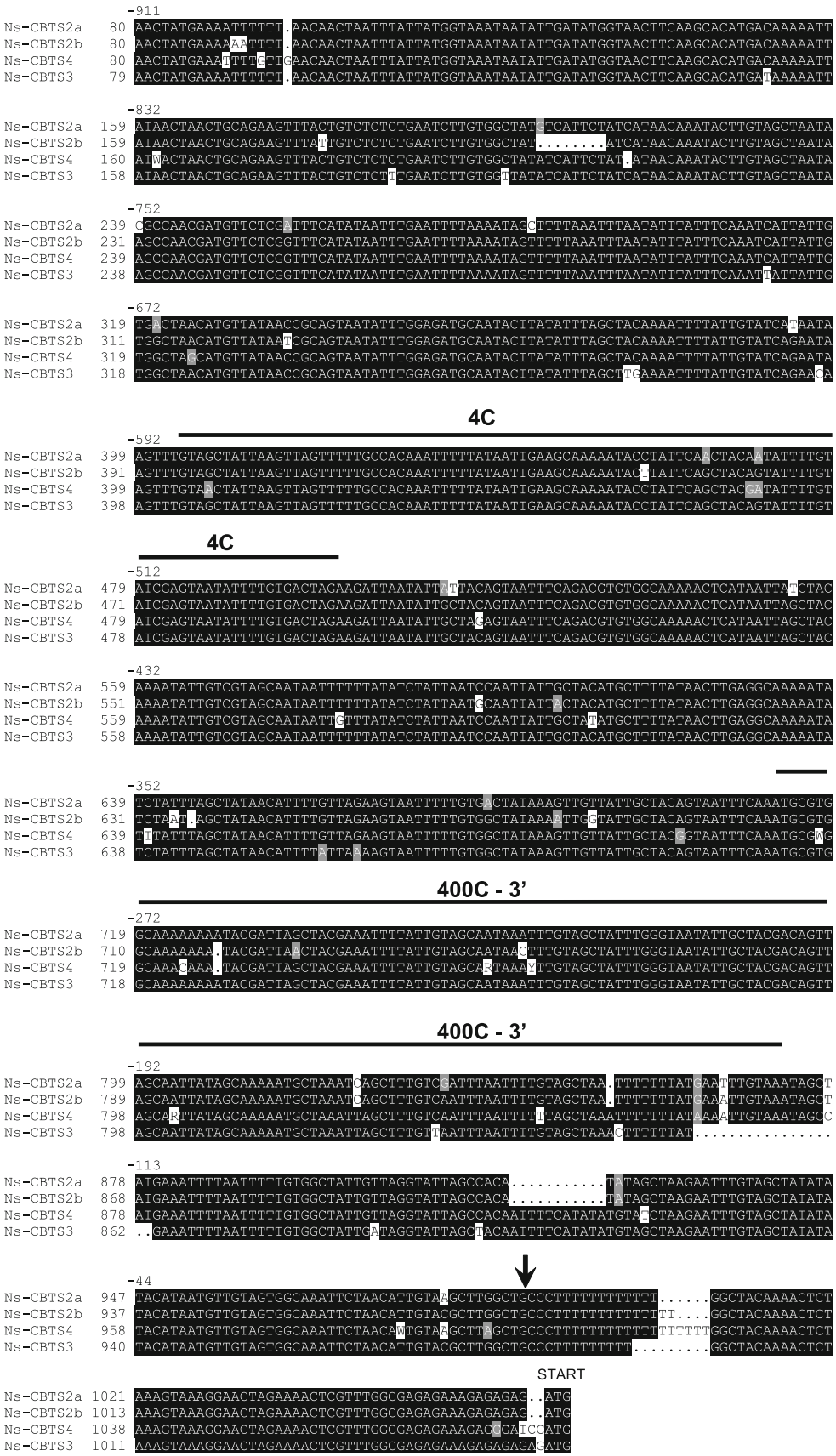


Fig. 3 Alignment of the *NsCBTS* promoter sequences. START indicates the translation initiation codon and the *downward pointing arrow* the transcription initiation site. The 4C region and the 3′ end of the 400C fragment identified as *cis*-regulating elements are indicated by *horizontal bars*

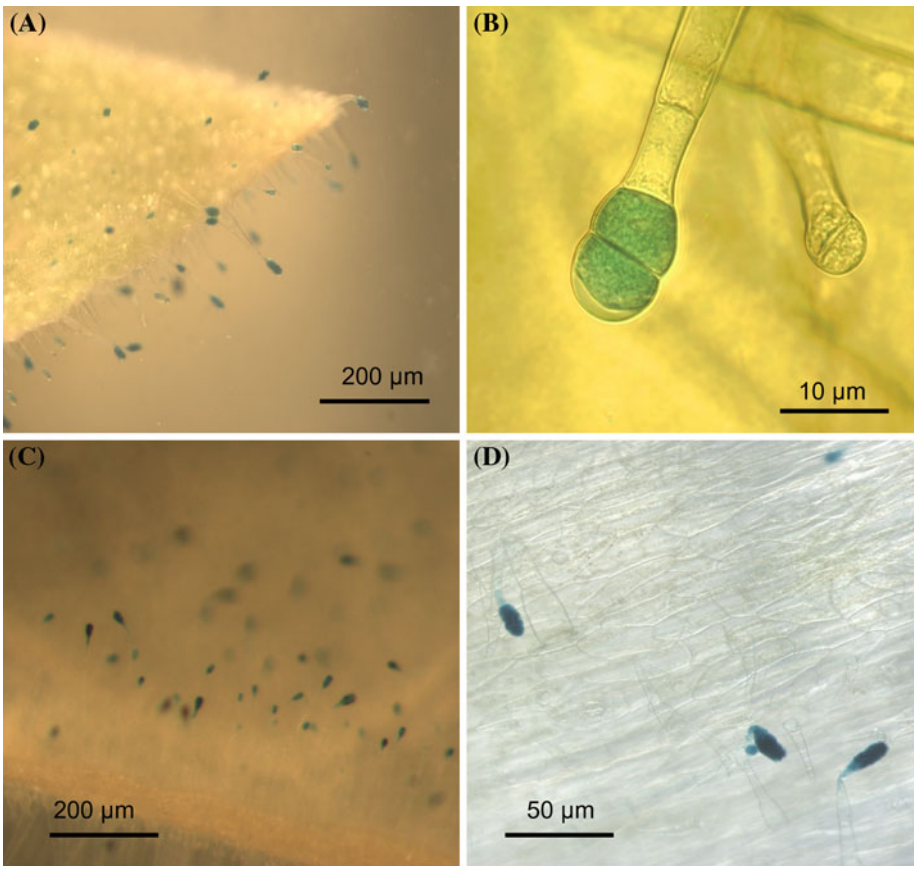


Fig. 4 GUS staining of *N. sylvestris* lines expressing the pCBTS-2a:*uidA* fusions. **a**, **b** Fusion with the 1.1 kb fragment of the *CBTS-2a* promoter. **c**, **d** Fusion with the 1.7 kb fragment of the *CBTS-2a* promoter

line. Transgenic lines expressing the 1.7 and 1.09 kb pCBTS:*uidA* fusions were recovered and stained for GUS activity. As shown in Fig. 6A, both promoters conferred a GUS expression profile specific to the secretory cells of the glandular trichomes, as was observed in the wild type background. This demonstrated that the *NsCBTS-2a* promoter activity was not affected in *CBTS* silenced plants suggesting that CBT-diols do not control CBTS transcription, and thus the ihp*CBTS* line provided a suitable background for the analysis of the *NsCBTS-2a* promoter.

Promoter deletion analysis

To identify regulatory regions that control trichome specific expression of the *CBTS-2a* gene, we generated by PCR a series of seven 5′ or internal deletions within the 1.09 kb promoter region: Δ5 s (1060 bp), Δ4 s (880 bp), Δ1 s (410 bp), Δ4c (110 bp), Δ3c (200 bp), Δ1c (190 bp), Δ400c (360 bp), as depicted in Fig. 6B. The resulting promoters were cloned into the pKGWFS7 Gateway® vector. These vectors were used to transform a pCBTS1.7:ihp*CBTS* line and several independent transformants were recovered for each construct (for a complete list of transformants see the Supplemental Data file). As expected, no GUS expression is detected when the whole promoter region is deleted as in Δ5 s (data not shown) or when only 119 bp of the promoter were kept (construct Δ4 s, Fig. 6B4). Similarly, the Δ1c deletion (190 bp) which removed the TATA box resulted in total absence of GUS expression (data not shown). Data with these constructs also indicated that surrounding sequences in the T-DNA vector do not cause expression of the reporter gene. On the other hand, the Δ1 s deletion (coordinates -910 to -589) had no effect on GUS expression intensity and pattern indicating that it does not contain essential regulatory elements (Fig. 6B5). Therefore, regulatory elements should be located between coordinates -589 and -119. This was confirmed by investigation of the expression pattern of the Δ4c, Δ3c and Δ400c deletions. First, the Δ4c (-589 to -479) deletion resulted in a complete loss of expression (Fig. 6B7). This suggested that the 4c region should contain an essential transcription activator element. In contrast, the Δ400c deletion led to significant ectopic GUS expression in the epidermis as well as in trichomes (Fig. 6B6). Expression was also visible in the roots as patches on epidermal cells (data not shown). The Δ3c deletion, which covers approximately the 5′ half of the fragment deleted in Δ400c does not lead to any apparent loss of specificity or strength of expression (Fig. 6B8). These results indicate that there are two types of *cis* elements in the CBTS

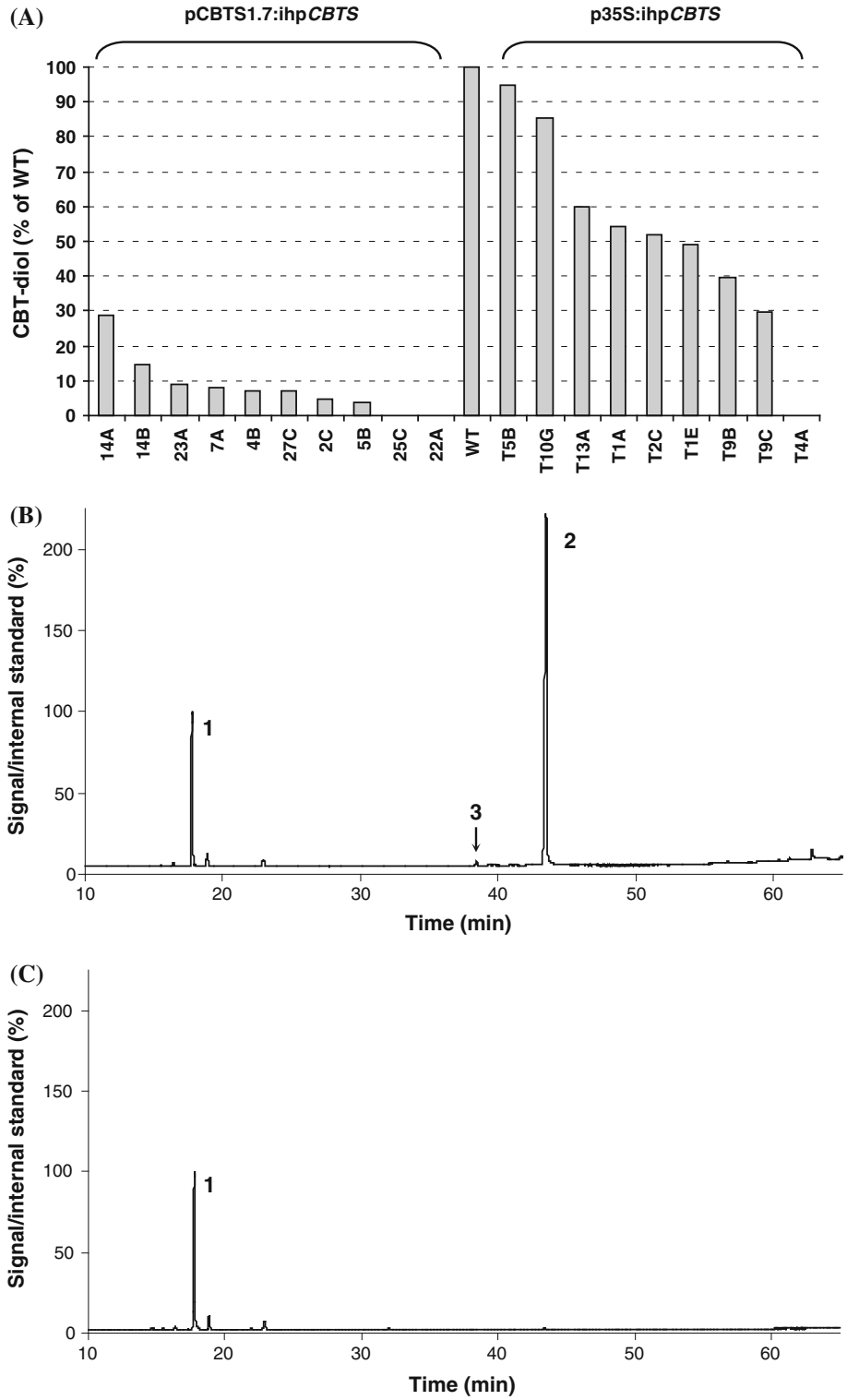
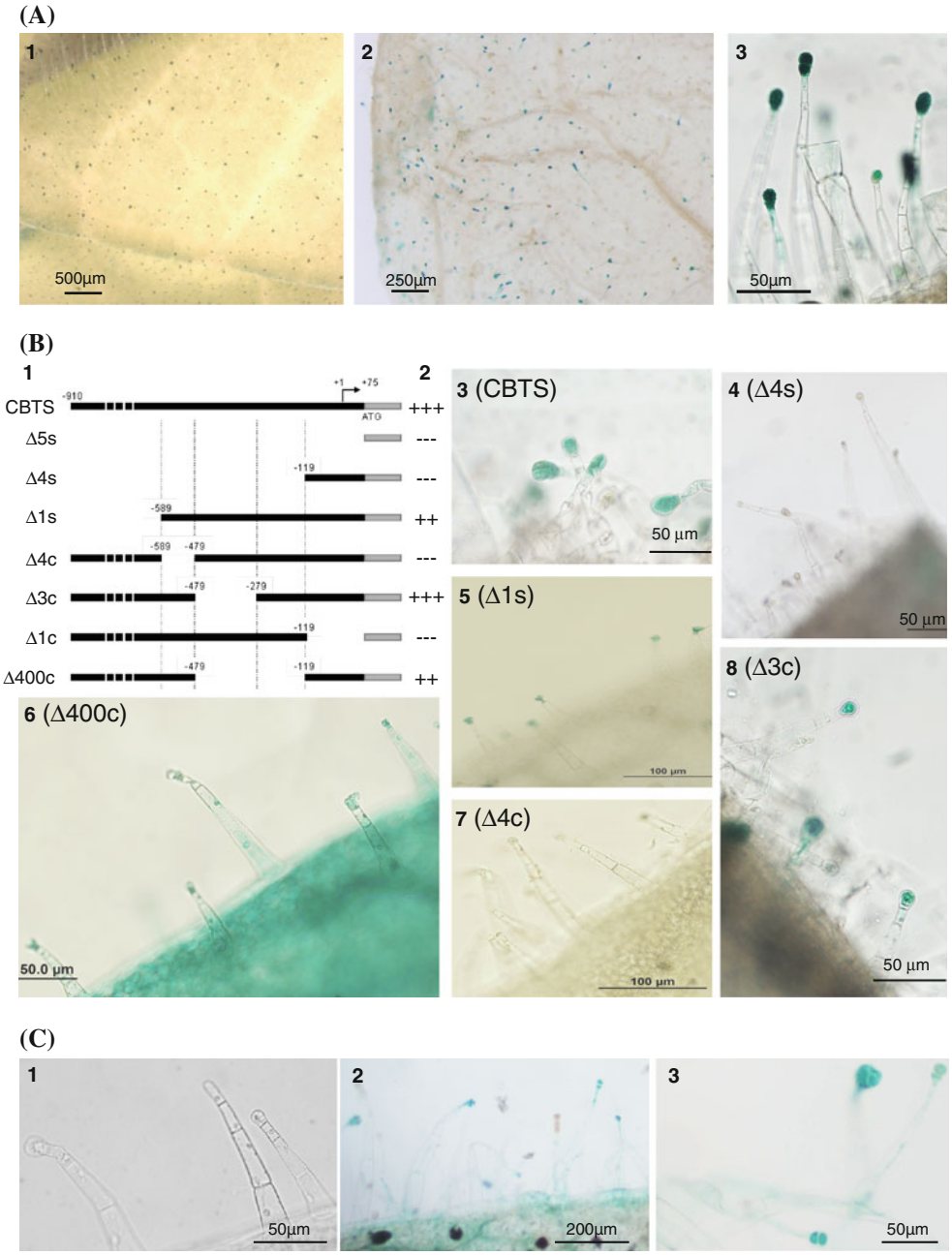


Fig. 5 Silencing the *NsCBTS* genes leads to reduced levels of cembranoids. In our GC–MS conditions, the two stereoisomers of cembratrien-diol and cembratrien-ol (α and β) cannot be separated and co-elute. The mass spectra of cembratrien-ols and cembratrien-diols are provided in the Supplemental data file. **a** Amount of CBT-diols, expressed as a percentage of WT levels, measured in the leaf exudate of transgenic lines expressing either the pCBTS1.7:ihpCBTS or the 35S:ihpCBTS constructs. **b**, **c** GC–MS chromatograms of the leaf exudate of WT (**b**) and line 22A. *Peak 1*: extraction internal standard (caryophyllene); *Peak 2*: CBT-diols; *Peak 3*: CBT-ols

promoter. The first, located in the -589 to -479 region contains an activating element conferring a relatively broad expression profile, including in the leaf epidermis and in the roots. The second, located in the -279 to -119 region, contains a repressor element which restricts the expression of the CBTS to the trichome glandular cells.

Confirmation of the enhancing activity of the 4c region

To verify its enhancer activity, the 4c fragment was cloned upstream of a minimal 35S promoter ($P35S_{min}$) of 46 bp. This minimal promoter fragment by itself does not drive transcription of a transgene but contains a TATA box



which is responsive to transcriptional enhancers (Fang et al. 1989). p35S$_{min}$ was cloned upstream of the *uidA* reporter gene and the construct integrated in the pKGWFS7 T-DNA vector for plant transformation. Similarly, a pCBTS4c:p35S$_{min}$:*uidA* was made and cloned into the pKGWFS7 T-DNA vector. Both T-DNA vectors were used to transform a *N. sylvestris* ihp*CBTS* line and transgenic lines were analyzed for GUS expression. Staining of plants harbouring the p35S$_{min}$:*uidA* construct revealed no visible expression in the trichomes or in other parts of the plant (Fig. 6C1). On the other hand, plants with the pCBTS4c:p35Smin:*uidA* construct showed staining not only in the trichome heads, but also in the epidermis (Fig. 6C2, C3). This pattern is very similar to that of the pCBTS Δ400c:*uidA* fusion (Fig. 6B6) and thus indicates that the 4c region is responsible for the broad expression profile of the pCBTS Δ400c:*uidA* construct.

Discussion

A *NsCBTS* multigene family

Glandular trichomes of the cultivated tobacco *Nicotiana tabacum* and the wild tobacco species *Nicotiana sylvestris* produce the macrocyclic diterpenoids α- and β-CBT-diols

◀ **Fig. 6** Analysis of the *NsCBTS-2a* promoter in transgenic plants. *Panel A*: Silencing the *CBTS* genes does not affect the activity of the *NsCBTS-2a* promoter. GUS staining of transgenic lines carrying pCBTS1.7:*uidA* (*1*) or pCBTS1.1:*uidA* (*2*) in the ihp*CBTS* background showing comparable expression pattern restricted to the tall trichomes. A detailed view (*3*) derived from the leaf observed in (*2*) showing the GUS signal limited to the glandular trichome head cells. *Panel B*: Deletion analysis of the *NsCBTS-2a* promoter in *N. sylvestris*. *1* Schematic representation of the *NsCBTS-2a* promoter deletion constructs for assaying GUS expression in transgenic *N. sylvestris* leaves. The sequential 5′-deletions (Δ5s, Δ4s, Δ1s) as well as the internal deletions (Δ4c, Δ3c, Δ400c, Δ1c) constructs of the *NsCBTS-2a* promoter were fused to the *uidA* gene. *2* Relative GUS signal intensities observed in transgenic lines harboring the modified promoter constructions. *3–7* Representative views of GUS expression pattern from T2 generation plants of *N. sylvestris* transgenic lines harboring deleted promoter constructs. GUS-stained leaf views from the most informative deletion constructs are shown. *3* Full 1.1 kb promoter region conferring GUS staining specifically located in the secreting trichomes head cells. *4* Δ4s deletion (−910 to −119) prevents GUS expression in any cell type. *5* Δ1s deletion (−910 to −589) has no effect on promoter activity. *6* Δ400c deletion is responsible for ectopic GUS expression in trichome head, trichome stalk and epidermis. *7* Δ4c deletion abolishes GUS staining. *8* Δ3c confers GUS staining similar to that of the full length 1.1 kb promoter. *Panel C*: The 4c region of the *NsCBTS-2a* promoter is able to activate a 35S (-46) minimal promoter. The minimal 35S promoter sequence (-46) was inserted into the pKGWFS7 vector immediately upstream of the *uidA* coding sequence. The 4c fragment (110 bp) from the *NsCBTS-2a* trichome-specific promoter was inserted upstream of the minimal promoter. Resulting constructions were transformed into *N. sylvestris* via *Agrobacterium*-mediated transformation. GUS expression was monitored in T1 plants. *1* No GUS expression is detected in plants transformed with the 35S minimal promoter alone. *2*, *3* Strong GUS staining is detected in plants transformed with the 35S minimal promoter fused to the 4c fragment from the *NsCBTS-2a* promoter, thus confirming that 4c fregion contains a cis-activating element

as major compounds. The pathway to these compounds is highly specific of the glandular cells and proceeds in two steps from the diterpene precursor GGPP. The first step is catalyzed by terpene synthases to give a mixture of α- and β- cembratrien-ols and the second step is due to the activity of a cytochrome P450 monooxygenase, CYP71D16, mediating hydroxylation on the carbon 6 of the cembratrien-ols. A candidate gene, *NtCYC-1*, for the CBT-ol synthase activity was identified from *N. tabacum* and its function was probed by a gene silencing approach (Wang and Wagner 2003). Relying on the similarity of *N. tabacum* and *N. sylvestris*, the *CYC-1* sequence was used to design primers for amplification of *N. sylvestris* sequences. Thus, four highly similar genes could be identified, *NsCBTS2a*, *NsCBTS2b*, *NsCBTS3* and *NsCBTS4*. To generate lines devoid of CBT-diols and CBT-ols, a gene silencing approach with an intron hairpin construct based on the highly conserved exon 2 was implemented. In the most strongly silenced lines, the levels of CBT-diols and CBT-ols were very low to almost undetectable levels in the T1 generation. This indicates that, collectively the *NsCBTS* genes are responsible for the biosynthesis of CBT-ols. A more detailed analysis of the catalytic activities of each NsCBTS protein will be required to identify their product specificity, if any. *Nicotiana sylvestris* is one of the presumed parents of the allotetraploid *N. tabacum*. Because the other presumed parent, *Nicotiana tomentosiformis*, does not produce cembranoid diterpenes in its trichomes (Alain Tissier, unpublished results) it seems reasonable to assume that the cembranoid diterpene biosynthesis genes of *N. tabacum* were inherited from *N. sylvestris*. However, an ortholog of the *CYC-1* gene in N. sylvestris could not be detected. On the other hand, we have preliminary evidence that orthologs of the *NsCBTS* genes are present and expressed in *N. tabacum* trichomes (unpublished results). Therefore, the exact role of the *CYC-1* gene in *N. tabacum* remains to be established.

Glandular trichome specific terpene synthases in the Solanaceae

Phylogenetic analysis of terpene synthases of the Tps-a group has shown that the CBTS proteins are most closely related to sesquiterpene synthases of tomato which are also expressed specifically in glandular trichomes (Fig. 2). This group of enzymes forms a distinct branch from the stress-related terpene synthases, such as the 5-epi-aristolochene synthase from tobacco or the vetispiradiene synthase from tomato. Several interesting observations can be drawn from these results. First, this indicates that glandular trichome specific terpene synthases in tobacco and tomato share a common ancestor and further that tomato and tobacco glandular trichomes have a common evolutionary origin, implying in turn that the developmental origin and regulation is likely to be conserved between those two species. Another interesting aspect is that both the tobacco CBTS diterpene synthases and the tomato SSTLH sesquiterpene synthases catalyze the formation of macrocyclic terpenes, albeit with different substrates (see Supplemental Fig. 2). These cyclizations are analogous to each other and therefore suggest that the common evolutionary origin of these terpene synthases did not only confer a similar pattern of expression but also resulted in a certain level of conservation in the reaction mechanisms with a diversification in the substrates. Thus, it would be of interest to test the activity of the CBTS enzymes in the presence of *E,E*-FPP and conversely of the SSTLH enzymes in the presence of GGPP. Also, given the sequence similarity between CBTS and SSTLH enzymes, the substrate and product specificities of these enzymes could be probed by homology modelling and the identification of key residues involved in substrate and product specificities.

The *NsCBTS-2a* promoter is highly specific of the secreting cells of the tall glandular trichomes

Our results indicate that a 1.1 kb fragment of the promoter of the *NsCBTS-2a* gene directs expression in a highly specific cell type, namely the secreting cells of the tall glandular trichomes. To our knowledge, the only other reported tobacco gene whose promoter has such a precise specificity is *CYP71D16* from *Nicotiana tabacum* (Wang et al. 2002). This gene encodes the subsequent step in the biosynthesis of the CBT-diols. This, together with our data is coherent with early observations that the biosynthesis of the CBT-diols takes place specifically in trichome glandular cells (Keene and Wagner 1985). Other compounds, such as the labdane diterpene *cis*-abienol and the sucrose esters, are also synthesized in glandular trichomes (Kandra and Wagner 1988; Guo et al. 1994). The biosynthesis genes of these compounds should thus provide additional promoters for the specific expression in glandular cells.

A number of promoters from other species, mostly cotton (*Gossypium sp.*) and *Arabidopsis thaliana*, have been tested in tobacco for trichome expression. When introduced in tobacco however, many of these genes show an expression profile which includes trichome cells but is otherwise poorly specific. For example, the promoters from the cotton fiber genes *SUS3* (Ruan et al. 2009), *GhRGP1* (Wu et al. 2006), *GhGlcAT1* (Wu et al. 2007), *GhGAL1* (Wu and Liu 2006) also confer expression in vascular tissues, roots, flowers and cotyledons. Other promoters such as the cotton *LTP3* (Liu et al. 2000; Hsu et al. 2005) and *LTP6* (Hsu et al. 1999) genes show specificity to the whole glandular trichomes of tobacco (*Nicotiana tabacum*), but do not differentiate between the secreting cells and the cells of the trichome stalk. The promoter of the *OASA1* gene from Arabidopsis also confers whole trichome specific expression and in addition does not differentiate between the short and the tall tobacco (*N. tabacum*) glandular trichomes (Gutierrez-Alcala et al. 2005). The only promoters from other species which seem to confer glandular cell specific expression are from the cotton (*Gossypium arboreum*) *GaMYB2* gene (Shangguan et al. 2008) and the *Solanum americanum SaPIN2* gene (Liu et al. 2006).

Synergistic action of repressing and activating *cis*-fragments to confer trichome glandular cells specificity

By constructing a series of sequential and contiguous deletions of the *NsCBTS-2a* 1 kb promoter fragment, two types of elements playing an important role in the expression pattern could be uncovered. The first, contained in the 4c fragment (-589 to -479), completely abolished expression in any tissue when it is deleted. This is indicative of the presence of a strong transcriptional enhancer in this fragment. In contrast, a deletion of the 400c fragment (-479 to -119) resulted in a much broader expression profile with strong expression in the trichome secreting cells, but also in the trichome stem, and in the leaf epidermis (Fig. 6B6). In addition, expression in the root meristem and as patchy spots in the differentiation zone could be seen (data not shown). This is indicative of the presence of a repressor element in the 400c fragment whose role is to restrict the expression strictly to the glandular cells. This also suggests that its absence in the Δ400c deletion construct reveals the expression conferred by the activating element of the 4c fragment. In agreement with this hypothesis, the 4c fragment cloned in front of a minimal 35S promoter and the *uidA* gene gave a pattern of expression similar to the Δ400c deletion construct. The localisation of the repressor element could be further narrowed to a 160 bp fragment (-279 to -119) because the Δ3c deletion (-479 to -289) construct gave an expression profile similar to the complete 1 kb promoter.

The co-existence of *cis*-activating and *cis*-repressing elements in a transcriptional promoter has been reported before. For example, the first intron of the cotton (*Gossypium hirsutum*) sucrose *synthase 3* (*SUS3*) gene is required for repression of the expression of *SUS3:uidA* fusions in the pollen of transgenic Arabidopsis plants (Ruan et al. 2009). Also, an AT-rich region of the *Fsltp4* gene from cotton was found to be required to repress the expression outside the cotton fibres and in addition an AT-hook transcription factor able to bind this AT-rich region was identified (Delaney et al. 2007).

It should be noted however, that most published promoter studies, at least those which looked at trichome specific ones, used sequential deletions, i.e. by progressively deleting larger and larger fragments from the 5′ end of the promoter. If we had not tested internal deletions (Δ400c and Δ4c) we would not have been able to uncover the repressing element because it is more proximal to the transcription initiation site than the activating element. Thus our work underscores the importance of carrying internal in addition to sequential deletions to fully analyze the *cis*-acting elements of a promoter.

In conclusion, we have shown that the CBTS activity in the tobacco glandular trichomes is encoded by a multigene family. Over 1 kb of the promoters of the *NsCBTS* genes were sequenced and were found to be highly similar to each other, suggesting they are co-regulated. The promoter of one of these genes, *NsCBTS-2a*, was further characterized and shown to direct the expression of *uidA* specifically to trichome glandular cells. Thus, this promoter may be used for metabolic engineering of tobacco glandular trichomes. In addition, we were able to identify *cis*-regulating regions required for the specific expression pattern of *NsCBTS-2a*. This establishes the foundation to identify

transcription factors both within and outside the trichome glandular cells which synergistically contribute to the highly specific expression profile of the CBTS genes. Availability of these transcription factors will open more opportunities for trichome metabolic engineering.

Acknowledgments We acknowledge the support of the Agence Nationale de la Recherche (programme Réseau Innovation Biotechnologies 2005), grant number ANR-05-PRIB02102 to AT and MH. We wish to thank the Groupement de Recherches Appliquées en Phytotechnologies of the CEA Cadarache for assistance in growing the plants.

References

Bohlmann J, Crock J, Jetter R, Croteau R (1998) Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (E)-alpha-bisabolene synthase from grand fir (Abies grandis). Proc Natl Acad Sci USA 95:6756–6761

Delaney SK, Orford SJ, Martin-Harris M, Timmis JN (2007) The fiber specificity of the cotton FSltp4 gene promoter is regulated by an AT-rich promoter region and the AT-hook transcription factor GhAT1. Plant Cell Physiol 48:1426–1437

Fang R-X, Nagy F, Sivasubramaniam S, Chua N-H (1989) Multiple cis-regulatory elements required for maximal expression of the Cauliflower Mosaic Virus 35S promoter in transgenic plants. Plant Cell 1:41–150

Gallagher SR (ed) (1992) Quantitation of GUS activity by fluorometry. Academic Press Inc., San Diego, pp 47–59

Guo ZH, Severson RF, Wagner GJ (1994) Biosynthesis of the diterpene cis-abienol in cell-free extracts of tobacco trichomes. Arch Biochem Biophys 308:103–108

Gutierrez-Alcala G, Calo L, Gros F, Caissard J-C, Gotor C, Romero LC (2005) A versatile promoter for the expression of proteins in glandular and non-glandular trichomes from a variety of plants. J Exp Bot 56:2487–2494

Hsu CY, Creech RG, Jenkins JN, Ma DP (1999) Analysis of promoter activity of cotton lipid transfer protein gene LTP6 in transgenic tobacco plants. Plant Sci 143:63–70

Hsu CY, Jenkins JN, Saha S, Ma DP (2005) Transcriptional regulation of the lipid transfer protein gene LTP3 in cotton fibers by a novel MYB protein. Plant Sci 168:167–181

Kandra L, Wagner GJ (1988) Studies of the site and mode of biosynthesis of tobacco trichome exudate components. Arch Biochem Biophys 265:425–432

Keene CK, Wagner GJ (1985) Direct demonstration of duvatrienediol biosynthesis in glandular heads of tobacco trichomes. Plant Phys 79:1026–1032

Lange BM, Croteau R (1999) Genetic engineering of essential oil production in mint. Current Opin Plant Biol 2:139–144

Liu HC, Creech RG, Jenkins JN, Ma DP (2000) Cloning and promoter analysis of the cotton lipid transfer protein gene Ltp3. Bioch Biophys Act 1487:106–111

Liu J, Xia KF, Zhu JC, Deng YG, Huang XL, Hu BL, Xu XP, Xu ZF, Liu J, Xia KF, Zhu JC, Deng YG, Huang XL, Hu BL, Xu XP, Xu ZF (2006) The nightshade proteinase inhibitor IIb gene is constitutively expressed in glandular trichomes. Plant Cell Phys 47:1274–1284

Mahmoud SS, Croteau RB (2002) Strategies for transgenic manipulation of monoterpene biosynthesis in plants. Trends Plant Sci 7:366–373

Martin DM, Fäldt J, Bohlmann J (2004) Functional characterization of nine Norway Spruce TPS genes and evolution of Gymnosperm Terpene Synthases of the TPS-d subfamily. Plant Physiol 135:1908–1927

Rontein D, Onillon S, Herbette G, Lesot A, Werck-Reichhart D, Sallaud C, Tissier A (2008) CYP725A4 from yew catalyzes complex structural rearrangement of taxa-4(5), 11(12)-diene into the cyclic ether 5(12)-oxa-3(11)-cyclotaxane. J Biol Chem 283:6067–6075

Ruan MB, Liao WB, Zhang XC, Yu XL, Peng M (2009) Analysis of the cotton sucrose synthase 3 (Sus3) promoter and first intron in transgenic Arabidopsis. Plant Sci 176:342–351

Sallaud C, Rontein D, Onillon S, Jabès F, Duffé P, Giacalone C, Thoraval S, Escoffier C, Herbette G, Leonhardt N, Causse M, Tissier A (2009) A Novel Pathway for Sesquiterpene Biosynthesis from Z, Z-Farnesyl Pyrophosphate in the Wild Tomato Solanum habrochaites. Plant Cell 21:301–317

Schilmiller AL, Last RL, Pichersky E (2008) Harnessing plant trichome biochemistry for the production of useful compounds. Plant J 54:702–711

Shangguan X-X, Xu B, Yu Z-X, Wang L-J, Chen X-Y (2008) Promoter of a cotton fibre MYB gene functional in trichomes of Arabidopsis and glandular trichomes of tobacco. J Exp Bot 59:3533–3542

Siebert PD, Chenchick A, Kellogg DE, Lukyanov KA, Lukyanov SA (1995) An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res 23:1087–1088

Smith NA, Singh SP, Wang MB, Stoutjesdijk PA, Green AG, Waterhouse PM (2000) Total silencing by intron-spliced hairpin RNAs. Nature 407:319–320

Stoutjesdijk PA, Singh SP, Liu Q, Hurlstone CJ, Waterhouse PA, Green AG (2002) hpRNA-mediated targeting of the Arabidopsis FAD2 gene gives highly efficient and stable silencing. Plant Physiol 129:1723–1731

Van Der Hoeven RS, Monforte AJ, Breeden D, Tanksley SD, Steffens JC (2000) Genetic control and evolution of sesquiterpene biosynthesis in *Lycopersicon esculentum* and *L. hirsutum*. Plant Cell 12:2283–2294

Wagner GJ (1991) Secreting glandular trichomes: more than just hairs. Plant Physiol 96:675–679

Wagner GJ, Wang E, Shepherd RW (2004) New approaches for studying and exploiting an old protuberance, the plant trichome. Ann Bot 93:3–11

Wang E, Wagner GJ (2003) Elucidation of the functions of genes central to diterpene metabolism in tobacco trichomes using posttranscriptional gene silencing. Planta 216:686–691

Wang E, Wang R, DeParasis J, Loughrin JH, Gan S, Wagner GJ (2001) Suppression of a P450 hydroxylase gene in plant trichome glands enhances natural-product-based aphid resistance. Nat Biotech 19:371–374

Wang E, Gan S, Wagner GJ (2002) Isolation and characterization of the CYP71D16 trichome-specific promoter from *Nicotiana tabacum L.* J Exp Bot 53:1891–1897

Wang EM, Hall JT, Wagner GJ (2004) Transgenic *Nicotiana Tabacum L.* with enhanced trichome exudate cembratrieneols has reduced aphid infestation in the field. Mol Breed 13:49–57

Wu AM, Liu JY (2006) Isolation of the promoter of a cotton beta-galactosidase gene (GhGal1) and its expression in transgenic tobacco plants. Sci Chin Ser C-Life Sci 49:105–114

Wu AM, Ling C, Liu JY (2006) Isolation of a cotton reversibly glycosylated polypeptide (GhRGP1) promoter and its expression activity in transgenic tobacco. J Plant Phys 163:426–435

Wu AM, Lv SY, Liu JY (2007) Functional analysis of a cotton glucuronosyltransferase promoter in transgenic tobaccos. Cell Res 17:174–183