**Syringyl lignin biosynthesis is directly regulated by a secondary cell wall master switch**

Qiao Zhao\(^a\), Huanzhong Wang\(^a\), Yanbin Yin\(^b,c\), Ying Xu\(^b,c\), Fang Chen\(^b,c\), and Richard A. Dixon\(^a,c,1\)

\(^a\)Plant Biology Division, Samuel Roberts Noble Foundation, Ardmore, OK 73401; \(^b\)Computational Systems Biology Lab, Department of Biochemistry and Molecular Biology, and Institute of Bioinformatics, University of Georgia, Athens, GA 30602; and \(^c\)Bioenergy Sciences Center (BESC), Oak Ridge, TN 37831

Contribution by Richard A. Dixon, June 29, 2010 (sent for review June 6, 2010)

Lignin is a major component of plant secondary cell walls and is derived from p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) monolignols. Among higher plants, S lignin is generally considered to be restricted to angiosperms, which contain the S lignin-specific cytochrome P450-dependent monooxygenase, ferulic acid/coniferaldehyde/coniferyl alcohol 5-hydroxylase (F5H). The transcription factor MYB58 directly regulates expression of monolignol pathway genes except for FSH. Here we show that FSH expression is directly regulated by the secondary cell wall master switch NST1/SND1, which is known to regulate expression of MYB58. Deletion of NST1 expression in *Medicago truncatula* leads to a loss of S lignin associated with a more than 25-fold reduction of FSH expression but only around a 2-fold reduction in expression of other lignin pathway genes. A detailed phylogenetic analysis showed that gymnosperms lack both FSH and orthologs of NST1/SND1. We propose that both FSH and NST1 appeared at a similar time after the divergence of angiosperms and gymnosperms, with FSH possibly originating as a component of a defense mechanism that was recruited to cell wall biosynthesis through the evolution of NST1-binding elements in its promoter.

---

**Results**

**FSH Is Regulated Differently from Other Lignin Biosynthesis Genes in the *M. truncatula* *nst1* Mutant.** *M. truncatula* NST1 was previously identified from a *Tnt1* retrotransposon insertion mutant population screened by analysis of stem cross sections for lignin autofluorescence. The *nst1* mutants (five independent alleles) show a dramatic loss of lignin autofluorescence in the interfascicular region, and thioacidolysis analysis indicated reductions in the levels of H, G, and S monomers (14). Remarkably, S units, which account for ≈30% of the total lignin content in mature *Medicago* stems, were totally undetectable by thioacidolysis. To confirm this finding by an alternative approach, Mäule staining of stem cross sections was performed to selectively locate S lignin in the *nst1-1* line (Fig. 1). The red coloration indicates the presence of S lignin in the vascular elements and interfascicular fibers in wild-type plants, but this was almost undetectable in the *nst1* mutant, such that the fibers were not visible in the stained cross sections.

Based on microarray analysis, expression of most verified lignin biosynthetic genes, including *cinnamate 4-hydroxylase* (CA4), *caffeic acid 3-O-methyltransferase* (COMT), *caffeoyl CoA 3-O-methyltransferase* (*CCoAOMT*), and *4-coumarate: CoA ligase* (*4CL*), was reduced by between 1- and 2-fold in stems of the *Medicago* *nst1* mutant (Fig. 2). However, *FSH* gene expression was reduced by more than 25-fold, suggesting that *FSH* is regulated differently from other lignin biosynthesis genes.

**Arabidopsis** SND1, MYB46, and MYB58 Differentially Activate *M. truncatula* Lignin Biosynthesis Genes. To investigate how regulation of *FSH* might differ from that of other lignin biosynthesis genes, we used an in vitro *Medicago* leaf protoplast transcription

Author contributions: Q.Z., Y.X., and R.A.D. designed research; Q.Z., H.W., Y.Y., and F.C. performed research; Q.Z., H.W., Y.Y., and R.A.D. analyzed data; and R.A.D. wrote the paper.

The authors declare no conflict of interest.

\(^1\)To whom correspondence should be addressed. E-mail: radixon@noble.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1009170107/-/DCSupplemental.
assay to test the effects of coexpression of cell wall-related transcription factors on trans-activation of a reporter gene, firefly luciferase, driven by the promoters of various *M. truncatula* lignin biosynthesis genes. Because MYB58 can directly activate lignin biosynthesis genes, and MYB46 can regulate the entire secondary cell wall formation program (5, 13), these two transcription factors were chosen for the assay. However, because the *Medicago* orthologs of MYB58 and MYB46 have yet to be identified, *Arabidopsis* MYB58 and MYB46 were used. To be consistent with the use of *Arabidopsis* MYB58 and MYB46, *Arabidopsis* SND1 was used in place of *Medicago* NST1, because the two genes are orthologous (14).

We first tested the effect of SND1 expression on monolignol pathway gene activation. Firefly luciferase reporter gene constructs were made to include around 1 kb of DNA upstream of the start codon for each of the lignin pathway genes tested. Coexpression of cauliflower mosaic virus (CaMV) 3S promoter driven SND1 activated the expression of the firefly luciferase reporter gene driven by the F5H promoter by around 25-fold (Fig. 3A). However, SND1 could not activate the promoters of the other lignin pathway genes tested (C4H, COMT, CCoAOMT, and 4CL), all of which are involved in both G and S lignin biosynthesis (1) (Fig. 3A). The same assay was then performed with the 3S promoter driving expression of MYB46 in place of SND1. The FSH, C4H, and COMT promoters were activated by around 4-fivefold, and CCoAOMT and 4CL were activated by more than 10-fold (Fig. 3B). The fact that all promoters were able to be activated rules out the possibility that the promoter regions of C4H, COMT, CCoAOMT, and 4CL selected for the assays were not functional. Using the same procedure, the ability of MYB58 to activate the same set of promoters was also tested. The C4H, COMT, CCoAOMT, and 4CL promoters were strongly activated, by 20- to 70-fold, whereas the F5H promoter was only activated around 4-fold (Fig. 3C).

**SND1 Directly Activates Expression of F5H by Binding to Its Promoter.** To further investigate the hypothesis that F5H is, unlike other monolignol pathway genes, directly regulated by SND1, a steroidal receptor-based inducible system was used in the *Medicago* protoplast assay. The regulatory region of the human estrogen receptor (HER) (15) was first tested. Cotransfection of *Medicago* leaf protoplasts was performed with 3S promoter-driven SND1 fused with HER and the FSH promoter-driven firefly luciferase reporter construct. However, the FSH promoter was highly activated even without addition of the inducer estradiol. This could possibly be due to the high level of phytoestrogens in legume cells, which are able to mimic estradiol (16). To obviate this problem, we used a glucocorticoid receptor (GR)-based inducible system (17). The 3S promoter-driven SND1 was fused with the GR domain, and this construct was cotransfected into *Medicago* leaf protoplasts with the FSH promoter-driven firefly luciferase reporter gene (Fig. 4A). Luciferase activity was highly induced by the addition of dexamethasone (Dex), whereas Dex alone could not activate expression of the reporter gene (Fig. 4B).

To test whether SND1 is directly activating the F5H promoter-driven reporter gene in these assays, the protein synthesis inhibitor cycloheximide was used to block de novo protein synthesis. A total of 2 μM cycloheximide was sufficient to completely abolish F5H promoter-driven firefly luciferase activity induced by Dex (Fig. 4B), confirming that this concentration effectively inhibits protein synthesis. Under these conditions, genes that are directly activated by SND1 should still be induced by addition of Dex, because SND1 is under control of the 3S promoter and has been produced before cycloheximide treatment. However, indirect activation of genes by SND1 via other proteins would be inhibited because of inhibition of new protein synthesis. Dex-induced appearance of F5H transcripts via SND1 under cycloheximide treatment clearly
Lignin is often considered to be restricted to angiosperms, we have shown that F5H is regulated directly by NST1/SND1, and S. Zhao et al. PNAS Early Edition

Phylogenetic Analysis of NST, F5H, MYB46, and MYB58. We next considered the evolutionary origins of F5H, NST, and MYB genes, by conducting phylogenetic analysis of plant species including algae, ferns, moss, two angiosperms, and 12 gymnosperm species (Table S1). The large phylogeny of F5H-related proteins (Fig. 6A) included the other monolignol pathway P450s coumaroyl shikimate 3′-hydroxylase (C3′H) and C4H, and other cytochrome P450 proteins. Three Medicago proteins are closest to the two Arabidopsis F5H proteins in the phylogeny (Fig. S1). Interestingly, the five proteins were clustered with one protein translated from an assembled EST UniGene from the fern species Adiantum capillus veneris. The six proteins together were further clustered with 13 Medicago proteins and then with many gymnosperm proteins. Therefore, the single fern protein may represent the ortholog of F5H, whereas true gymnosperm F5H orthologs may be missing. The many gymnosperm P450 proteins await functional characterization. Interestingly, no moss or spike moss proteins were found in the higher plant F5H clade. The previously described spike moss (Selaginella moellenendorfii) smF5H is located in a distant clade, consistent with the independent evolution of F5H activity in spike moss (18).

In the large phylogeny of NAC transcription factors (Fig. 6B), those related to secondary cell wall synthesis (12, 19, 20) form one clade, except for SND2 and SND3. Medicago NST1 is orthologous to Arabidopsis NST1, -2, and -3 (SND1) in this clade (Fig. S2). No gymnosperm proteins are found to be orthologous to Arabidopsis and Medicago NST proteins. The closest gymnosperm proteins to MtNST1 (plantGDB ID: PUT-163a-Picea_glauca-31332_2 and PUT-162b-Picea_allspecies-73182_2) form a separate clade with AT4G10350.1 (ANAC070), AT1G33280.1 (ANAC015), and Medtr4g041220.1, all of which are not yet functionally characterized. All of the remaining Arabidopsis NAC proteins form a separate subclade, consisting of AtVND1-7, a few

![Fig. 4](image-url) Direct activation of F5H by SND1. M. truncatula leaf protoplasts were first transfected with 3SS:SND1 fused with GR. The protoplasts were then treated with dexamethasone (Dex) or cycloheximide (CHX) plus Dex. Luciferase activity or F5H expression (luciferase transcript level) in control samples (no Dex) is set to 1.0. (A) The 3SS:SND1-GR construct. (B) Luciferase activities in protoplasts cotransfected with SND1-GR and F5H promoter-luciferase. Dex alone cannot induce F5H promoter activity, and 2 μM CHX can completely abolish luciferase activity, indicating that de novo protein synthesis is inhibited. (C) Induction of F5H gene expression. M. truncatula Actin was used as internal control. Error bars represent SE of three biological replicates.

![Fig. 5](image-url) SND1 binds directly to the F5H promoter. (A) EMSA showing direct binding between SND1 and the F5H promoter. The NAC domain of SND1 fused with MBP was expressed in E. coli, and the purified recombinant protein incubated with biotin-labeled F5H promoter. (B) Competition analysis of SND1 binding to the F5H promoter shows that P2, P3, and P4 fragments of the F5H promoter contain the binding site. Lanes are as follows: 1, no competitor; 2, wild type 249-bp probe; 3-8, fragments P1-P6, respectively. MBP was used as a control protein. Unlabeled competitor DNA fragments are in 200-fold molar excess relative to the labeled probes.

Zhao et al.
Medicago proteins, a few gymnosperm proteins, and two spike moss proteins basal to the other proteins.

MYB transcription factors usually contain at least two Pfam Myb-DNA-binding domains (21). In our large phylogeny (Fig. 6C), 13 MYB proteins related to secondary cell wall biosynthesis (12, 19, 20) are located in two major clades each consisting of four subclades. One of the major clades contains the N-terminal Myb-DNA-binding domain, whereas the other contains the C-terminal Myb-DNA-binding domain. ATMYB58 and -63 form a subclade, ATMYB20, -42, -43, and -85 form a subclade, ATMYB46, MYB53, MYB26, and MYB103 form a subclade, and ATMYB52, -54, and -69 form a subclade. In the subclade containing ATMYB58 (Figs. S3-S5), two Medicago EST UniGenes are orthologous to ATMYB58, ATMYB65, and the other two Arabidopsis proteins (AT1G56160.1 and AT3G12820.1); in the subclade containing MYB46 (Figs. 1S and S7), Medtr2g117145.1 is the Medicago ortholog.

We did not find algal genes homologous to NST transcription factors, but did find algal homologs of F5H enzymes and MYB transcription factors (Fig. 6). Recently it was reported that one red alga, Calliardthron cheliosporoides, makes lignins (22). However, none of the algal homologs is close to the known lignin synthesis-related F5H, C3H, and C4H genes (Fig. 6C). This indicates that either the green algae included in this study do not have lignins or that algal lignin synthesis genes evolved independently compared with their higher plant equivalents, as seen for smF5H (18). In the MYB phylogeny (Fig. 6C), a total of 32 green algal and 37 red algal homologs are found, many in the two major clades containing the N and C termini of the 13 secondary cell wall-related MYB proteins, which further locate in four subclades of each of the two major clades. Clearly three of the four subclades, excepting the one containing MYB52/54/69 homologs, are closer to each other and one single green algal protein is found basal to them, suggesting they have diverged after the green algae appeared. On the other hand, all of the remaining algal proteins of the two major clades are basal to the subclade containing MYB52/54/69 homologs, suggesting that these MYB subclades diverged from the other three before the appearance of the green algae.

Fig. 6. Phylogenies of F5H, NST, and MYB homologous proteins from 21 plant species. The multiple sequence alignments were of (A) 1,331 proteins homologous to Arabidopsis F5H, (B) 500 proteins homologous to Arabidopsis NST proteins, and (C) 2,209 proteins homologous to Arabidopsis MYB proteins, respectively. Selected supporting values >70% are shown. Plant proteins of interest are identified and highlighted by names.

Discussion

Syringyl units are the second most abundant monolignol units in the lignin of angiosperms, and, apart from a report of a sinapyl alcohol-specific cinnamyl alcohol dehydrogenase from aspen (23), F5H appears to be the only truly S-lignin-specific monolignol biosynthetic enzyme. S lignin accumulation can be decreased or increased simply by down-regulation or up-regulation of F5H (24, 25). Given the special role of F5H in the lignin biosynthetic pathway, its regulation is particularly critical for the determination of lignin composition. Early reports of transcript accumulation in plants grown under different light/dark periods first hinted that F5H may be differently regulated than other lignin pathway genes (26).

Recently, several transcription factors involved in lignin biosynthesis have been characterized (5, 10, 11, 13). NST1 and SND1 can activate the entire secondary cell wall formation program and regulate the expression of MYB46 and MYB58. Although MYB46 is a downstream target of NST1 and SND1, it alone can also activate the whole secondary cell wall biosynthesis program. MYB58 directly activates lignin pathway genes, except for F5H, by binding to the AC elements in their promoter regions (5), and there is no AC element in the F5H promoter (9). We now show that F5H is directly regulated by SND1, which in parallel regulates the secondary cell wall biosynthesis program, in part by activation of the lignin-specific MYB58.

A series of in vivo and in vitro data support the concept that F5H is in a different regulatory network from the other lignin pathway genes. First, F5H is down-regulated to a much greater degree than all of the other lignin biosynthesis genes in the Medicago nst1 mutant, which appears to essentially lack S lignin (14); it should be noted that the genes analyzed in Fig. 2 have been functionally proven to be involved in lignification in Medicago (27), and some are present in the M. truncatula genome as single genes (C3H, HCT, and F5H). Second, in a transient expression system, the F5H promoter is activated by coexpression of SND1, the Arabidopsis ortholog of Medicago NST1, but other lignin pathway promoters are not. Third, SND1 can bind the promoter region of F5H and activate F5H expression without de novo protein synthesis, indicating that the regulation of F5H by SND1 is through direct binding to the F5H promoter rather than through downstream transcription factors such as MYB46 and MYB58.

Most gymnosperms and ferns do not synthesize S lignin. Based on our phylogenetic analysis, angiosperms including Arabidopsis and Medicago possess F5H, NST, MYB46, and MYB58. However, gymnosperms do not appear to have orthologs of F5H and NST, but do have orthologs of MYB46 and the lignin-specific transcription factor MYB58. Ferns, moss, and algae do not appear to have NST, MYB46, or MYB58, although there is currently only one fern species
with EST data (16,944 UniGenes), compared with the EST data from 12 gymnosperm species. Ferns do have an FSH ortholog, the biochemical function of which remains to experimentally determined whereas the mosses and algae analyzed do not appear to have FSH homologs, with the exception of Selaginella moellendorffii, which has an independently evolved FSH (18).

Although all 12 gymnosperms in our analysis lack complete genome information (relying on EST analysis), it seems unlikely that the datasets for all 12 species would contain MYB46 and MYB85 but lack FSH and SND1. The apparent absence of FSH genes in lower plants, the absence of green algal homologs of NAC genes, and the limited distribution of algal MYB genes therefore strongly support the evolution of S lignin synthesis as starting after the emergence of land plants, and the expansion and maturation of the regulatory machinery has taken a very long time, starting no earlier than the appearance of angiosperms. However, as indicated previously, the presence of S lignin in spike moss (18) and lower algal plants (22) suggests that there may have been independent, alternative evolutionary routes for the generation of S lignin synthesis and its regulation in lower plants.

The current model for secondary cell wall biosynthesis suggests that both NST1/SND1 and MYB46 can activate the entire cell wall formation program including cellulose, hemicellulose, and lignin biosynthesis. Although MYB46 is a downstream target of NST1/SND1, these two different master switches at first sight seem to be redundant. However, the fact that the NST1/SND1 deletion mutant grows quite well, whereas the growth of the MYB46:MYB83 double deletion mutant is arrested immediately after germination (28) suggests that the more recently evolved NST1/SND1 control is more dispensable. Our model suggests that NST1/SND1 directly activates both FSH and MYB46 (Fig S8). The two different master switches may be necessary to guarantee that FSH is activated coordinately with the secondary cell wall program that was originally under the master control of MYB46 alone.

Monolignols have strong antimicrobial activity, and are produced as soluble glycosides in many plant species, including Arabidopsis roots (29, 30). Among the different lignin monomers, S lignin seems to be particularly involved in plant defense, and some plants selectively accumulate S lignin in response to pathogen attack (31). Arabidopsis plants overexpressing FSH under the C4H promoter accumulate high levels of S lignin and show increased resistance to nematode infection (29). It is possible that S monolignol biosynthesis initially evolved as a defense response following the appearance of FSH, and that the new pathway was subsequently recruited for cell wall biosynthesis once its regulatory transcription factor, NST1, gained the ability to bind the MYB46 promoter. This model is consistent with the involvement of many NAC transcription factors in biotic stress tolerance (32, 33).

Materials and Methods

Mäule Staining and Microscopy. The fifth internodes from the top of M. truncatula stems were cross-sectioned to 100 μm. To visualize S lignin, Mäule staining and microscopy. The same amount of ethanol used to dissolve Dex was applied to the control protoplasts. A total of 2 μM cycloheximide was applied 30 min before addition of Dex to inhibit new protein synthesis. The treated protoplasts were then used for dual-luciferase reporter activity assay or real-time quantitative PCR analysis of luciferase transcripts. The expression of FSH-luciferase in the control protoplasts was set to 1.0 and the data are presented as averages ± SD of three biological replicates.

Real-Time PCR. cDNA samples were used for quantitative real-time PCR (qRT-PCR) with technical duplicates. The 10-μL reaction included 2 μL primers (0.5 μM of each primer), 5 μL Power SYbr (Applied Biosystems), 2 μL 1:20 diluted cDNA from the reverse transcription step, and 1 μL water. qRT-PCR data were analyzed using SDS 2.2.1 software (Applied Biosystems). PCR efficiency was estimated using the LinRegPCR software (36) and transcript levels determined by relative quantification (37) using the M. truncatula actin gene as a reference.

EMSA. To express the recombinant NAC domain of the SND1 protein, a 191-amino-acid peptide was fused in frame with maltose binding protein (MBP) and purified using amylose resin. The resulting protein was used for EMSA with the FSH promoter fragments. The promoter fragments were PCR amplified with one 5′-biotin-labeled primer and purified using a PCR purification kit (Invitrogen). The biotin-labeled DNA fragments were incubated for 20 min with 100 ng of SND1-MBP in EMSA binding buffer (Pierce). Reaction mixtures were resolved by polyacrylamide gel electrophoresis, and DNA was electroblotted onto nitrocellulose membrane and detected by chemiluminescence. Generation of nonlabeled promoter fragments for competition assays is described in SI Materials and Methods.

Homology Searching. The following hidden Markov models (HMMs) were used to search against the translated peptides of the 21 species in Table S1 using the hmmsearch command of the HMMER package (38). For MYB proteins, the Pfam MYb_DNA-binding domain HMM model (PF00249) was used; for NAC proteins, the A to E subdomain HMM model was built as described previously (33); and for FSH proteins, a HMM model was generated using the hmmbuild and hmmscan commands of the HMMER package based on multiple sequence alignment of known FSH protein sequences. If the plant genome was sequenced, the automated predicted peptides were used; otherwise the translated peptides in six frames of the assembled EST UniGenes by plantGDB were used (39).

Phylogenetic Analyses. The homologs found in the three searches under the E-value cutoff <0.01 were collected, respectively, and subjected to multiple sequence alignments using MAFFT v6.717 program (40) with the L-INS-I method. The phylogenies were further reconstructed using the FastTree v2.1.1 program (41). The local support values beside the nodes were computed by resampling the site likelihoods 1,000 times and performing the Shimodaira Hasegawa test to show the confidence levels with regard to the clustering of relevant proteins into one group.

The Arabidopsis MYB genes reported previously (20), the two Arabidopsis FSH genes listed at http://cellwall.genomics.purdue.edu/families/1-3.html, and the previously reported Arabidopsis NAC genes (12, 19, 20) were identified in the phylogenies, and the subclades containing these marker genes were taken out and displayed using the Interactive Tree of Life (ITOL) Web server (42).

Note Added in Proof. A recent review (43) observed that S lignin is present in a few gymnosperm species but that its evolutionary/biochemical origin is not yet clear. It will be interesting to determine whether it arose from a NAC-controlled classical FSH system that was then lost in later gymnosperm lineages or whether it has arisen independently as in Selaginella.