THE PLANT CELL

The Plant Cell, 2024, **36**, 2709–2728

https://doi.org/10.1093/plcell/koae120
Advance access publication 24 April 2024
Research Article

Cytochrome b_5 diversity in green lineages preceded the evolution of syringyl lignin biosynthesis

Xianhai Zhao,[1] Yunjun Zhao,[1,†] Qing-yin Zeng,[2] and Chang-Jun Liu[1,*]

[1]Biology Department, Brookhaven National Laboratory, Upton, NY 11973, USA
[2]State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry and Northeast Forestry University, Beijing 100091, China
*Author for correspondence: cliu@bnl.gov
†Present address: CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, 300 Fenglin Rd., Xuhui District, Shanghai 200032, China.
The author responsible for distribution of data and materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plcell/pages/General-Instructions) is: Chang-Jun Liu (cliu@bnl.gov).

Abstract

Lignin production marked a milestone in vascular plant evolution, and the emergence of syringyl (S) lignin is lineage specific. S-lignin biosynthesis in angiosperms, mediated by ferulate 5-hydroxylase (F5H, CYP84A1), has been considered a recent evolutionary event. F5H uniquely requires the cytochrome b_5 protein CB5D as an obligatory redox partner for catalysis. However, it remains unclear how CB5D functionality originated and whether it coevolved with F5H. We reveal here the ancient evolution of CB5D-type function supporting F5H-catalyzed S-lignin biosynthesis. CB5D emerged in charophyte algae, the closest relatives of land plants, and is conserved and proliferated in embryophytes, especially in angiosperms, suggesting functional diversification of the CB5 family before terrestrialization. A sequence motif containing acidic amino residues in Helix 5 of the CB5 heme-binding domain contributes to the retention of CB5D function in land plants but not in algae. Notably, CB5s in the S-lignin–producing lycophyte *Selaginella* lack these residues, resulting in no CB5D-type function. An independently evolved S-lignin biosynthetic F5H (CYP788A1) in *Selaginella* relies on NADPH-dependent cytochrome P450 reductase as sole redox partner, distinct from angiosperms. These results suggest that angiosperm F5Hs coopted the ancient CB5D, forming a modern cytochrome P450 monooxygenase system for aromatic ring *meta*-hydroxylation, enabling the reemergence of S-lignin biosynthesis in angiosperms.

Introduction

When the pioneering ancestors of land plants transitioned from water to terrestrial habitats approximately 470 to 600 million yr ago, they immediately confronted a series of life-threatening challenges, such as exposure to damaging UV-B radiation, desiccation, temperature fluctuations, coevolving herbivores and pathogens, and lack of structural support (Raven 1984). To survive these harsh terrestrial conditions, early land plants evolved various protective mechanisms and specialized metabolic capacities. These changes enabled them to produce essential defense substances that facilitated their adaptation to the terrestrial ecosystem. The development of lignified vascular tissue marked a significant milestone in the evolution of land plants, by providing early vascular plants with physical rigidity and strengthened water-conducting cells, allowing for long-distance water transport. This innovation allowed land plants to significantly increase in size, ultimately contributing to their dominance of the Earth flora (Friedman and Cook 2000; Weng et al. 2010).

Lignin is a heteropolymer primarily derived from oxidative coupling of 3 phenylpropanoid units, i.e. *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol (collectively called monolignols), which give rise to *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) lignin subunits when incorporated into the lignin polymer (Boerjan et al. 2003). The amount and composition of lignin monomers vary substantially among major taxa of vascular plants. In general, ferns and gymnosperms deposit lignin primarily composed of G units, with a small proportion of H units. By contrast, angiosperm lignins additionally contain S units, resulting in G and S copolymers, along with some H monomers (Boerjan et al. 2003; Vanholme et al. 2010). The innovation of S-lignin is regarded as a significant and characteristic evolutionary event of angiosperms (Weng et al. 2010). Interestingly, S-lignin has also been detected in *Selaginella*, a genus that represents an extant lineage of the most basal vascular plants, the lycophytes. Its emergence in *Selaginella* has been demonstrated as an independent evolutionary event relative to its angiosperm counterpart (Weng et al. 2008a). Therefore, S-lignin might have evolved multiple times through convergent evolution in different plant lineages (Weng et al. 2008b, 2010).

In angiosperms, 3 cytochrome P450 monooxygenases (P450s) catalyze the hydroxylation of the phenyl ring of phenylpropanoid units, resulting in the structurally characteristic monolignols. Cinnamate 4-hydroxylase (C4H; CYP73A5) performs the *para*-hydroxylation of cinnamic acid, the initial aromatic intermediate of the phenylpropanoid pathway, producing *p*-coumaric acid. This compound is essential for the formation of all 3 types of lignin subunits. *p*-Coumaroyl ester 3′-hydroxylase (C3′H; CYP98A3) and ferulate/coniferyl alcohol/coniferaldehyde 5-hydroxylase (F5H, also reported as FERULIC ACID 5-HYDROXYLASE 1 [FAH1]; CYP84A1) are responsible for the subsequent *meta*-hydroxylations of the benzene ring. C3′H catalyzes the first *meta*-hydroxylation reaction, necessary for the biosynthesis of both G- and S-lignin units (Franke et al. 2002), while F5H catalyzes the second *meta*-hydroxylation reaction, leading to the formation of S-lignin subunits (Supplementary Fig. S1; Humphreys et al. 1999). Uniquely, the characterized F5H in *Selaginella moellendorffii* (SmF5H)

Received October 18, 2023. Accepted March 27, 2024

is evolutionarily distant from its angiosperm counterpart and belongs to a phylogenetically new P450 family, CYP788 (Weng et al. 2008b). Unlike angiosperm F5H, SmF5H is essentially a bifunctional *meta*-hydroxylase capable of catalyzing both 3- and 5-hydroxylations of phenylpropanoids (aldehydes and alcohols), indicating its independent evolution (Weng et al. 2010). Nevertheless, similar to its angiosperm counterpart, SmF5H can restore S-lignin biosynthesis when expressed in the Arabidopsis (*Arabidopsis thaliana*) F5H-deficient mutant *fah1* (Weng et al. 2008b).

Catalysis by P450 enzymes requires redox partner(s) to supply reducing power, i.e. electrons from pyridine dinucleotide cofactors NADPH and/or NADH (Hannemann et al. 2007). In eukaryotic cells, there are 2 endoplasmic reticulum (ER) electron transfer systems: the NADPH-cytochrome P450 oxidoreductase (CPR) chain and the NADH-cytochrome b_5 reductase (CBR)-cytochrome b_5 (CB5) chain. These systems support the activities of P450 enzymes and other oxidases (Porter 2002; Schenkman and Jansson 2003; Kandel and Lampe 2014; Liu 2022). In some cases, CB5 also functionally associates with CPR, from which it accepts electrons to regulate P450 enzymatic activity (Bhatt et al. 2017; Zhao et al. 2023). CB5 represents an ancient heme-containing protein that is commonly found in mammals, plants, yeasts, and even purple phototrophic bacteria (Schenkman and Jansson 2003). CB5 is represented by a single gene in the budding yeast (*Saccharomyces cerevisiae*) genome (Schenkman and Jansson 2003), while vertebrates have 2 CB5 genes encoding distinct isoforms, one anchored to the ER membrane and the other located at the outer mitochondrial membrane (Cowley et al. 2005; Parthasarathy et al. 2011). By contrast, flowering plant genomes often contain multiple *CB5* genes (Smith et al. 1992; Napier et al. 1995; Fukuchi-Mizutani et al. 1999; Hwang et al. 2004; Kumar et al. 2006, 2012; Maggio et al. 2007; Gou et al. 2019). For instance, the Arabidopsis genome has 5 annotated *CB5* genes, encoding isoforms AtCB5A (encoded by At1g26340), AtCB5B (At2g32720), AtCB5C (At2g46650), AtCB5D (At5g48810), and AtCB5E (At5g53560). These isoforms share only 40% to 68% amino acid sequence identity, suggesting their functional diversification.

Indeed, we previously discovered that the Arabidopsis CB5 family member AtCB5D specializes as an indispensable electron shuttle protein supporting the AtF5H1-catalyzed benzene ring 5-hydroxylation required for the biosynthesis of 5-hydroxylated phenolics and S-lignin monomers (Gou et al. 2019). None of the other Arabidopsis CB5 isoforms can substitute for CB5D in augmenting AtF5H1 catalysis, highlighting the highly specialized CB5 family in angiosperms (Gou et al. 2019; Zhao et al. 2023). Moreover, the recognized CB5D member accepts electrons from both NADPH–CPR and NADH–CBR electron transfer systems and functionally associates exclusively with AtF5H1, but not C4H (or C3′H), in monolignol biosynthesis (Zhao et al. 2023). Unlike C4H and C3′H that are believed to have diverged from ancient primary metabolic cousins at a very early stage of land plant evolution (Werck-Reichhart and Feyereisen 2000; Renault et al. 2017; Alber et al. 2019), the emergence of F5H in angiosperms has been regarded as a recent evolutionary event, coinciding with the innovation of S-lignin biosynthetic branch (Weng et al. 2008b, 2010). The identification of CB5D, tightly functionally associated with F5H in angiosperms, raises an intriguing question: did CB5D coevolve with F5H and acquire its functional specification during the rise of S-lignin? Furthermore, similar to angiosperm F5H, SmF5H in *Selaginella* also catalyzes S-monomer biosynthesis. It remains unclear whether this independently evolved P450 enzyme, as its angiosperm counterparts, also requires CB5 as a redox partner for its catalysis.

In this study, we uncovered, through in planta genetic exploration and heterologous whole-cell biocatalytic assays, that the function of CB5D in enhancing F5H-catalyzed hydroxylation for syringyl type of phenolic biosynthesis originated in the charophytic algae, the closest living relatives of land plants, thus predating the advent of angiosperm F5H and S-lignin biosynthesis. Furthermore, the CB5 family has substantially expanded and diversified across land plant lineages. This early evolved CB5D function has been conserved in most embryophytes, including the basal land plants, such as liverworts, hornworts, and mosses, and the vascular lineages that do not produce S-lignin, such as ferns and gymnosperms. By contrast, even though S-lignin is present in the basal vascular plant lycophyte *Selaginella*, neither of its 2 CB5 orthologs has retained CB5D function and is required by SmF5H. Unlike AtF5H1, SmF5H, similar to the monolignol biosynthetic enzymes C4H and C3′H, recruits CPR but not CB5 as its electron donor. These findings suggest that the P450 monooxygenase F5H of angiosperms coopted the anciently evolved electron transfer protein CB5D for its newly arisen benzene ring 5-hydroxylation.

Results

CB5 is expanded and diverged in land plant lineages

Among the 5 canonical CB5 members found in Arabidopsis, only CB5D functions as an indispensable electron donor to support AtF5H1-catalyzed 5-hydroxylation of the benzene ring for the biosynthesis of syringyl-type phenolics, suggesting a functional specialization of angiosperm CB5 proteins. This functional divergence inspired us to deeply explore the evolution of the plant CB5 family, particularly in respect to the emergence of the CB5D functionality in supporting F5H-catalyzed S-lignin biosynthesis.

We performed BLAST searches using the Arabidopsis *CB5D* gene as an initial query against the available genomic and/or transcriptomic sequences from 20 representative plant species for its homologs. The selected species are across the entire green lineages, including green algae (Chlorophyceae *Volvox carteri* and *Chlamydomonas reinhardtii* and Charophyceae *Chara braunii* and *Chara vulgaris*), the bryophytes hornwort (*Anthoceros agrestis*), common liverwort (*Marchantia polymorpha*), and moss (*Physcomitrium patens*), lycophyte spike moss (*S. moellendorffii*), the monilophytes maidenhair fern (*Adiantum aleuticum*), giant fern (*Angiopteris evecta*), and aquatic floating fern (*Salvinia cucullata*), the gymnosperms ginkgo (*Ginkgo biloba*), loblolly pine (*Pinus taeda*), Norway spruce (*Picea abies*), and white spruce (*Picea glauca*), and the angiosperms *Amborella* (*Amborella trichopoda*), poplar (*Populus trichocarpa*), petunia (*Petunia axillaris*), rice (*Oryza sativa*), and sorghum (*Sorghum bicolor*; Fig. 1A). The obtained putative homologs were then used as queries to conduct reciprocal BLAST searches against the Arabidopsis genome, eventually yielding a large set of plant CB5 orthologs (Supplementary Data Set 1). In contrast to the single *CB5* gene found in yeast and green algae, the numbers of this gene commonly range from 2 to 7 in land plant species (Supplementary Data Set 1), indicating a substantial gene expansion after ancestral plants switched from their aquatic habitat to a terrestrial environment. Moreover, more complex plant species contained more copies of *CB5* homologs. For example, the flowering plants Arabidopsis, poplar, petunia, sorghum, and rice typically possess 5–7 genes, which likely resulted from the high

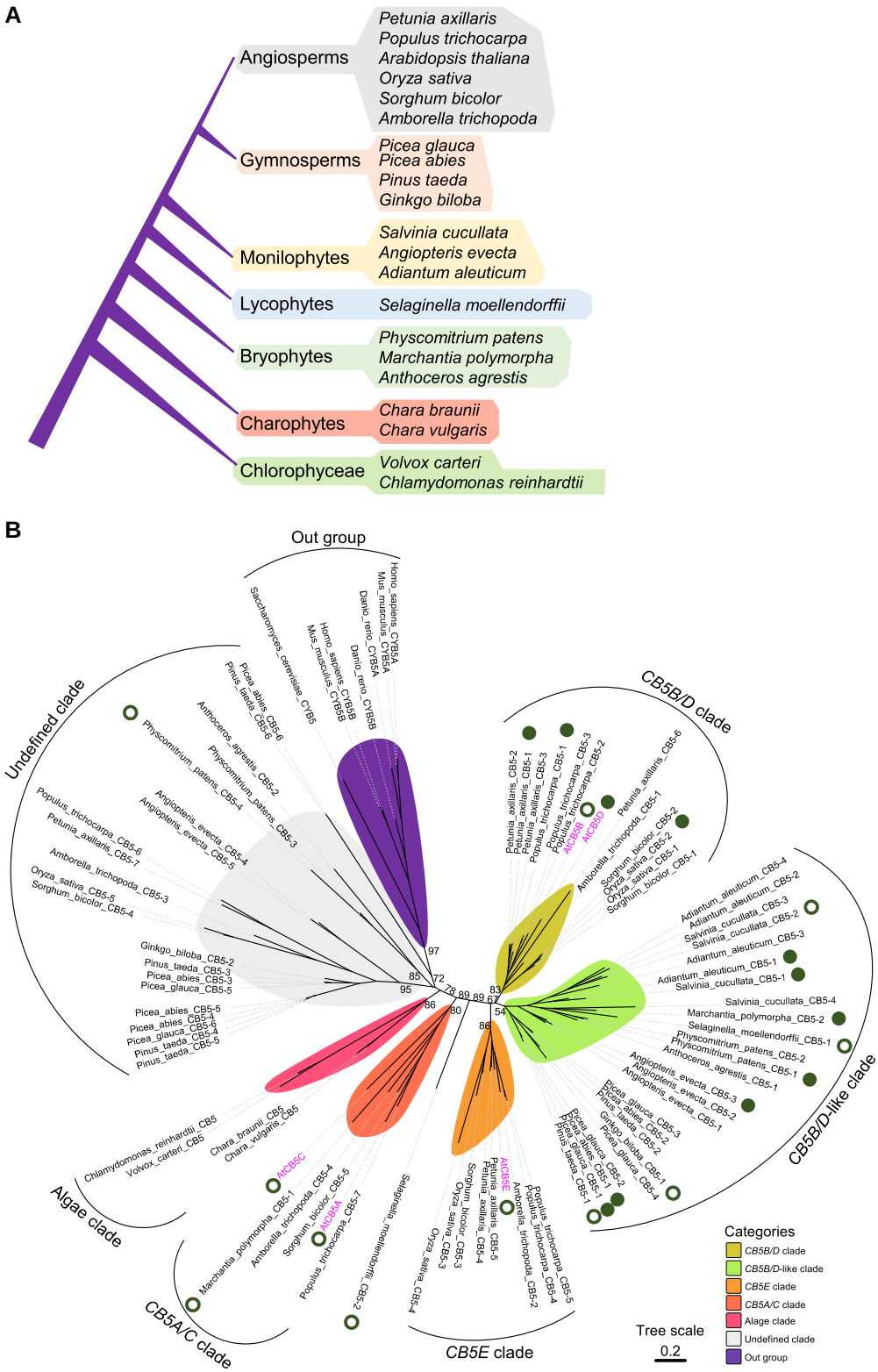

Figure 1. Phylogeny of CB5 homologs from representative plant taxa. **A)** A simplified phylogenetic tree view showing the relationship of the plant species from which the *CB5* homologs were identified. **B)** Reconstructed phylogeny of the *CB5* gene family in the green lineages, comprising 80 *CB5* homologs from the plant species indicated in **A)** and 7 homologs from *S. cerevisiae*, *H. sapiens*, *M. musculus*, and *D. rerio*, serving as outgroup. The coding sequences of candidate CB5 proteins were obtained by using AtCB5D as query via BLASTP against various genomic databases: UCSC Genome Browser (*H. sapiens*, *M. musculus*, and *D. rerio*), the Saccharomyces Genome Database (*S. cerevisiae*), Phytozome (*P. trichocarpa*, *O. sativa*, *S. bicolor*, *A. trichopoda*, *S. moellendorffii*, *M. polymorpha*, *P. patens*, *V. carteri*, and *C. reinhardtii*), PLAZA (*P. glauca*, *P. abies*, *P. taeda*, and *G. biloba*), ORCAE (*C. braunii*), Hornworts (*A. agrestis*), ONEKP (*C. vulgaris*, *A. evecta*, and *A. aleuticum*), FernBase (*S. cucullata*), and Solanaceae Genomics Network (*P. axillaris*). The full-length coding sequences of *CB5*s were aligned using the MUSCLE algorithm integrated within the MEGA X program. A maximum likelihood tree was then reconstructed using the IQ- tree2 algorithm (SYM+I+R5 model, 1,000 SH-aLRT tests, and 1,000 ultrafast bootstraps). Only major bifurcate bootstrap support values are indicated at the main branches of the tree. Arabidopsis *CB5* genes are highlighted in magenta. The filled and open circles indicate genes that successfully rescued or failed to rescue the *cb5d-1* defects in complementation assays, respectively. The published complementation assay results of *AtCB5A*, *AtCB5B*, *AtCB5C*, and *AtCB5E* (Zhao et al. 2023) were also included here.

frequency of whole-genome duplication (WGD) events that have occurred in flowering plants (Qiao et al. 2022). Such substantial gene expansion aligns with the increased metabolic complexity in land plants, contributing to their adaptation to the harsh terrestrial environment.

To determine the evolutionary relationship of the plant CB5 family members, we conducted a phylogeny reconstruction with a total of 87 *CB5* homologs found from 21 green lineages and 4 outgroup species via the maximum likelihood method using their coding sequences (Fig. 1B, Supplementary Files 1 and 2) and protein sequences (Supplementary Fig. S2 and Files 3 and 4). The topology of the reconstructed phylogenetic tree based on the nucleotide sequence showed that all green lineage *CB5*s are distinct from their yeast and mammalian counterparts and can be clustered into several distinct clades with high bootstrap value support (Fig. 1B). Among them, the *CB5*s from green algae are grouped as an independent clade distinct from all other homologs from land plants, suggesting that substantial evolution and development of CB5s occurred following plant terrestrialization. Based on the classification of Arabidopsis *CB5*s, we divided most homologs from land plant species into the highly supported clades *CB5A/C*, *CB5B/D*, *CB5B/D*-like, and *CB5E*. Additionally, several *CB5* homologs from across the entire range of embryophytes constitute a clade with less clear functional prediction, tentatively referred to as the "undefined clade." The members of this clade are more closely related to the outgroup consisting of yeast and mammalian homologs, yet they are distant from most CB5 members in land plants, suggesting their ancestral nature in the evolution of this gene family (Fig. 1B). Notably, a lycophyte homolog, *S. moellendorffii CB5-2* (*SmCB5-2*), is separated from all the defined clades, demonstrating its high divergency at least at the nucleotide sequence level (Fig. 1B). The *CB5A/C* clade represents an early diverged group, composed of *CB5*s from the basal bryophyte liverwort, the basal flowering species *Amborella*, and the angiosperms poplar, Arabidopsis, and sorghum. The absence of *CB5A/C* homologs in lycophytes, monilophytes, and gymnosperms suggests that they might have been lost during evolution but reacquired or retained in flowering plants in respect to specific environmental adaptation needs. The *CB5B/D* clade includes *AtCB5B*, *AtCB5D*, and other genes exclusively present in flowering lineages, forming a unique flowering plant *CB5* clade, as which is also the case for the *CB5E* clade (Fig. 1B). As a counterpart to the *CB5B/D* clade, we noticed the existence of a group of *CB5* homologs from diverse land species, ranging from basal bryophytes to lycophytes, monilophytes, and gymnosperms, which is termed the *CB5B/D*-like clade. Both *CB5B/D* and *CB5B/D*-like clades are positioned as a monophyletic group, suggesting their common ancestral origin. This monophyletic group of genes is conserved throughout the entire embryophyte lineage, indicating a high level of conservation and the likelihood that they serve crucial function(s) in the evolution of land plants. Notably, homologs from early emerging bryophytes such as liverwort, hornwort, and moss are clustered within the *CB5B/D*-like clade, suggesting the early existence of *CB5D*-like genes prior to the divergence and speciation of the primitive land plant lineages.

The duplicated or quadruplicated bryophyte *CB5*s can be classified into evolutionarily distant clades. For instance, 2 liverwort *CB5*s, *M. polymorpha CB5-1* (*MpCB5-1*) and *M. polymorpha CB5-2* (*MpCB5-2*), are placed in the evolutionarily distant *CB5A/C* and *CB5B/D-like* clades, respectively. In a similar vein, *CB5*s from hornwort and moss are separated within the *CB5B/D*-like clade, where they are closely related to liverwort *MpCB5-2*, and the undefined clade, where they are grouped loosely with homologs from ferns, gymnosperms, and angiosperms (Fig. 1B). These observations suggest that the *CB5* gene family has experienced early gene duplication and potential functional diversification in the basal land plants.

The phylogenetic tree reconstructed with CB5 protein sequences exhibited an overall similar topology to that of the nucleotide sequence–based tree, but also revealed some differences in branches. Noticeably, in contrast to the tree based on the nucleotide sequence where *SmCB5-2* is represented as a single branch, SmCB5-2 clustered within the CB5B/D-like clade in the protein sequence–based tree. Additionally, instead of grouping with CB5A, CB5C homologs are clustered with CB5E, forming a CB5C/E clade, which leaves CB5A homologs in a separate branch (Supplementary Fig. S2). Nevertheless, the CB5 protein sequence tree, characterized with lower branch support values, is less reproducible than the nucleotide sequence–based tree.

Complementation of CB5D loss by *CB5* orthologs from land plant species

The emergence of *CB5B/D*-like sequences in the primitive land plant lineages suggests that the functions analogous to AtCB5D might have already evolved in these ancient land species. To unequivocally demonstrate the emergence of bona fide CB5D in land plants, we selected 14 *CB5* homologs from *CB5B/D* and *CB5B/D*-like clades of the examined land plant species based on their relatively high hit score with *AtCB5D* in BLASTP (Fig. 2, Supplementary Data Set 2) to conduct a complementation test of the Arabidopsis *cb5d-1* mutant deficient in the biosynthesis of S-lignin and sinapoyl esters (Gou et al. 2019). In addition, we included 3 *CB5* homologs outside the *CB5B/D*-like clade, i.e. the liverwort *MpCB5-1*, moss (*P. patens*) *CB5-4* (*PpCB5-4*), and spike moss *SmCB5-2*, and that their encoded proteins share relatively low amino acid sequence similarity with AtCB5D, for comparison (Fig. 2, Supplementary Data Set 2).

In this assay, we placed all 17 synthesized *CB5* genes under the control of the Arabidopsis *C4H* promoter and transformed each resulting construct into the *cb5d-1* mutant. We confirmed the presence of the *CB5* transgene and its expression by genotyping PCR in primary transgenic plants and/or by reverse transcription quantitative PCR (RT-qPCR) analysis in T2 seedlings (Supplementary Fig. S3). We then determined the contents of sinapoyl esters in leaves and of S-lignin (expressed as S/G ratio) in stems in the obtained transgenic lines along with the wild type (WT, Col-0) and *cb5d-1* as controls. These quantifications served as proxies to assess the extent of functional complementation by each *CB5* ortholog in phenolic biosynthesis. Relative to the *cb5d-1* and WT plants, the analyzed transgenic lines defined 2 groups. Group I included the *cb5d-1* mutant and the transgenic lines harboring the *CB5* homologs *MpCB5-1*, *PpCB5-4*, *SmCB5-1*, *SmCB5-2*, *S. cucullata CB5-2*, *G. biloba CB5-1*, and *P. taeda CB5-1*. The S-lignin content of their stems and the content of sinapoyl esters in their leaves remained at levels similar to those of *cb5d-1*, indicative of the lack of complementation. Group II comprised WT, *cb5d-1 AtCB5D* complementation lines, and the transgenic lines harboring the remaining *CB5* homologs of land plant species. The S/G ratio and sinapoyl ester contents in these transgenic lines returned to the levels seen in the WT and the *cb5d-1 AtCB5D* line, indicating successful rescue of *cb5d-1* defects with these selected *CB5* homologs from different land plant species (Fig. 2). Most interestingly, despite the absence of lignin in nonvascular plants (Espineira et al. 2011), *CB5* homologs from the bryophytes liverwort *MpCB5-2* and moss *PpCB5-1* effectively rescued the defects of *cb5d-1*, providing evidence that CB5D

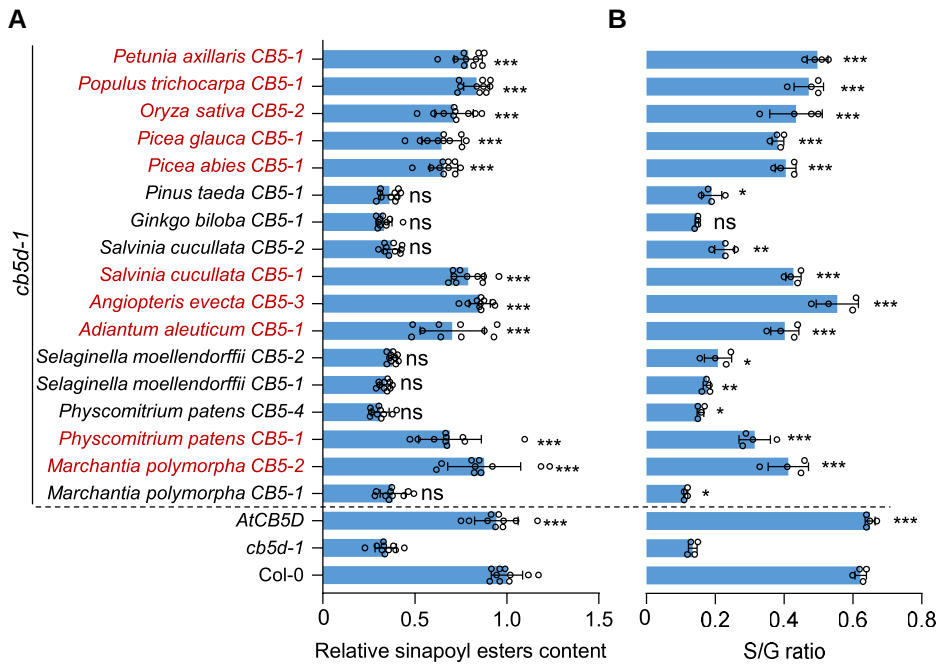

Figure 2. Complementation of *cb5d-1* with *CB5D* orthologs. **A)** Relative content in sinapoyl esters in the leaves and **B)** S/G ratio of stem lignins of the primary transgenic plants of the selected 17 *CB5D-like* orthologs in the Arabidopsis *cb5d-1* mutant background. One rosette leaf per individual 4-wk-old primary transgenic plant was sampled, and the leaves from 2 primary transgenic plants were mixed representing 1 biology replicate for sinapoyl esters analysis. Two to four stems from 11-wk-old primary transgenic plants were pooled representing 1 biology replicate for lignin thioacidolytic analysis. Data are means ± SD from 10 biological replicates for sinapoyl ester quantification and 4 biological replicates for lignin thioacidolytic analysis. The *CB5D* orthologs that rescued *cb5d-1* are highlighted in red. Asterisks indicate significant differences between the *cb5d-1* control and the indicated transgenic plants with $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$ (2-tailed Student's *t* tests). ns, not significant.

emerged in the ancient embryophytes prior to the advent of vascular plants (Fig. 2). By contrast, the paralogs *MpCB5-1* of liverwort and *PpCB5-4* of moss failed to rescue the *cb5d-1* defects (Fig. 2), indicative of the functional divergence and specialization of CB5 members in the basal land plants.

Furthermore, CB5D function appears to have been retained in descendent euphyllophytes, including ferns, gymnosperms, and angiosperms. Indeed, the homologs from ferns (i.e. *S. cucullata CB5-1*, *A. evecta CB5-3*, and *A. aleuticum CB5-1*), pines (*P. glauca CB5-1* and *P. abies CB5-1*), and the angiosperms rice (*O. sativa CB5-2*), poplar (*P. trichocarpa CB5-1*), and petunia (*P. axillaris CB5-1*) all rescued the *cb5d-1* defects in sinapoyl ester and S-lignin biosynthesis (Fig. 2). These data indicate that the early emerged CB5D function was inherited and retained in euphyllophytes during evolution.

Liverwort and hornwort CB5s effectively support Arabidopsis F5H catalysis

The rescue of *cb5d-1* with liverwort *MpCB5-2* transgene encouraged us to further dissect the biochemical properties of liverwort CB5 homologs. When we infiltrated *Nicotiana benthamiana* leaf epidermal cells with a construct encoding individual CB5 homologs fused to the yellow fluorescent protein (YFP), 2 liverwort CB5s showed distinct subcellular localization patterns. Consistent with their phylogenetic classification in the *CB5A/C* clade, we detected fluorescence from the YFP-MpCB5-1 fusion at the chloroplast outer envelope (Fig. 3A), similar to the reported localization for AtCB5A (Maggio et al. 2007), which we independently confirmed in Fig. 3B. By contrast, MpCB5-2, which rescues the *cb5d-1* defects, displayed a typical ER network localization pattern resembling that of AtCB5D (Fig. 3, C and D), suggesting its potential function as an ER electron transfer chain component.

We then produced and purified recombinant MpCB5 proteins lacking the transmembrane domains from *Escherichia coli*. Both recombinant MpCB5-1 and MpCB5-2 could be reduced by the reducing agent sodium dithionite and displayed typical cytochrome b_5 spectroscopic characteristics, with a maximal Soret absorption at 412 nm in their oxidized form and absorptions at 423, 528, and 557 nm when completely reduced (Fig. 3, E and F; Zhao et al. 2023). When we used Arabidopsis CBR1 in the presence of NADH or CPR2 (i.e. ATR2, encoded by At4g30210) in the presence of NADPH to reduce recombinant MpCB5s, we observed a more effective reduction for MpCB5-2, the CB5D-like protein, than for MpCB5-1 (Fig. 3, G and H), indicating its superior intrinsic electron accepting ability.

To further assess the potential functional association of liverwort CB5s with F5H, we developed a whole-cell biocatalytic assay within *E. coli* cells through the coproduction of truncated CB5 and F5H lacking their transmembrane domains. Accordingly, we individually fused MpCB5s, AtCB5D, or the mutant variant of CB5D, 2muCB5D, to AtF5H1 via a lambda linker (GSTSSGSG; Schuckel et al. 2012), resulting in the formation of a single polypeptide to ensure their equivalent production levels. In 2muCB5D, 2 histidine residues crucial for heme binding are replaced with alanine, thus preventing electron transfer (Gou et al. 2019); 2muCB5D was included in this assay as a control. After induction of protein production with isopropyl-β-D-thiogalactopyranoside (IPTG), we monitored the accumulation of recombinant proteins via SDS–PAGE analysis, which revealed strong Coomassie blue–stained bands (Supplementary Fig. S4). We then incubated the induced *E. coli* cells with coniferyl alcohol, the substrate of F5H, and analyzed the transformed products by HPLC and UHPLC-MS (Supplementary Fig. S5).

We first tested the AtF5H1-2muCB5D and AtF5H1-AtCB5D fusions to assess the reliability of the whole-cell catalytic system.

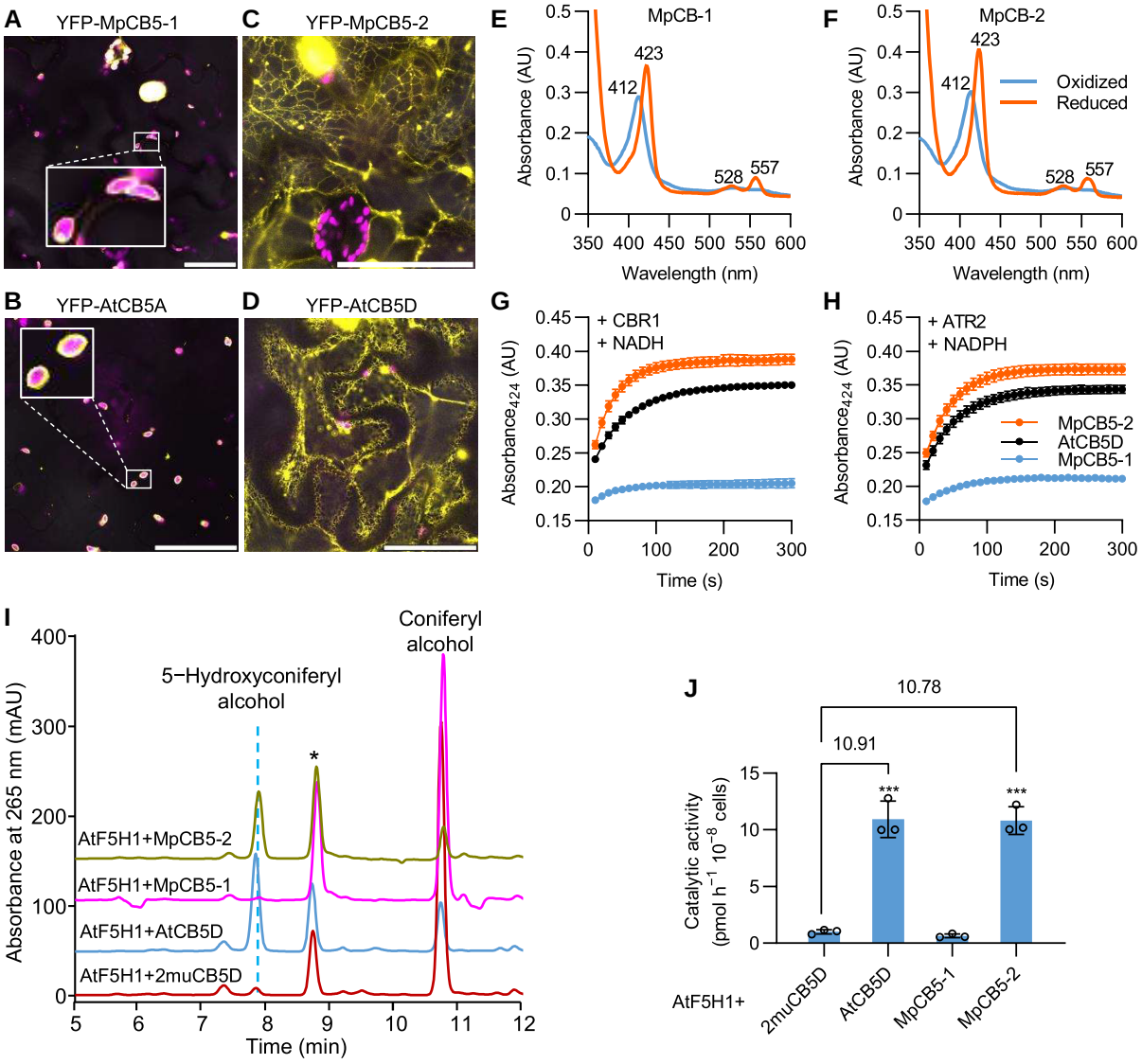

Figure 3. Characterization of MpCB5s from liverwort. **A to D)** Representative fluorescence pattern of YFP-MpCB5-1 **A)**, YFP-AtCB5A **B)**, YFP-MpCB5-2 **C)**, and YFP-AtCB5D **D)** in *N. benthamiana* leaf epidermal cells. The insets show enlarged images of the white boxed region indicating fluorescence in the chloroplast outer envelope. Scale bars, 50 μm. **E and F)** UV-Vis absorbance spectra of recombinant MpCB5-1 **E)** and MpCB5-2 **F)** in the oxidized or sodium dithionate–reduced state. **G and H)** Changes in absorbance at 424 nm of MpCB5s (7 μM) over the indicated reaction duration with 7.2 nM CBR1 in the presence of 40 μM NADH **G)** or 720 nM ATR2 in the presence of 40 μM NADPH **H)**. The absorption was recorded for 30 cycles in 10-s intervals. Data are means ± SD from 3 independent experiments. **I)** HPLC detected under UV illumination (HPLC-UV) profiles of the F5H-catalyzed biotransformation of coniferyl alcohol to 5-hydroxyconiferyl alcohol in *E. coli* cells. The asterisk indicates a nonspecific peak. 2muCB5D, a mutant version of AtCB5D lacking electron transfer property, was included as a negative control. **J)** Catalytic activity of different F5H-CB5 fusion proteins. Coniferyl alcohol was fed as substrate, and the 5-hydroxyconiferyl alcohol product was monitored by HPLC. The fold changes are shown above the horizontal lines, compared to control cells producing the 2muCB5D variant protein with no activity. Data are means ± SD from 3 biological replicates. Asterisks indicate significant differences with $^{***}P < 0.001$ (2-tailed Student's *t* tests).

When compared to 2muCB5D, coproduction of AtF5H1 with AtCB5D resulted in an up to 11-fold increase of F5H catalytic activity in converting coniferyl alcohol to the 5-hydroxyconiferyl alcohol product (Fig. 3, I and J). This result highlights the functionality of the truncated AtF5H1 and AtCB5D proteins within *E. coli* cells and confirms that AtCB5D is indeed an effective electron donor for AtF5H1-catalyzed benzene ring 5-hydroxylation. In addition, this result also suggests that the endogenous native *E. coli* redox systems can effectively reduce CB5, consistent with previous observations (Ichinose et al. 2004; Hatakeyama et al. 2016). We then tested the AtF5H1-MpCB5-1 and AtF5H1-MpCB5-2 fusions. Consistent with the in planta complementation assay (Fig. 2), MpCB5-2 substantially enhanced AtF5H1 activity, reaching a level comparable to that of the AtF5H1-AtCB5D fusion (Fig. 3, I and J). However, MpCB5-1, when fused to AtF5H1, essentially showed no effect on F5H catalytic activity (Fig. 3, I and J).

We obtained a similar result with hornwort CB5 homologs. When we fused *A. agrestis* CB5-1 (AaCB5-1), which clustered within the CB5B/D-like clade, to AtF5H1 and produced the fusion in *E. coli*, AtF5H1 displayed strong catalytic activity. This finding sharply contrasted with *A. agrestis* CB5-2 (AaCB5-2), a homolog classified in the undefined CB5 clade. Its combination with AtF5H1 showed no effect on AtF5H1 activity, similar to the effect observed with 2muCB5D (Supplementary Fig. S6). These data strongly support the presence of an authentic CB5D in bryophyte species.

CB5D-type function has emerged in green algae

With the detection of CB5D function supporting F5H catalytic activity in the primitive land plants, we were curious whether such a function ever emerged in algal CB5 proteins. Green algae typically have 1 copy of the *CB5* gene. We thus individually cloned the synthesized Chlorophyceae *C. reinhardtii CB5* (*CrCB5*) and *V. carteri CB5* (*VcCB5*) and the Charophyte *C. braunii CB5* (*CbCB5*) in-frame with *AtF5H1* and coproduced the respective F5H-CB5 fusion proteins in *E. coli* for a whole-cell catalytic assay. Although neither of the Chlorophyte CB5 proteins exhibited an effect on AtF5H1 catalytic activity, the Charophyte CbCB5 surprisingly showed a substantial enhancement of AtF5H1 catalytic activity for 5-hydroxylation of coniferyl alcohol (Fig. 4). These results indicate that CB5 from Charophyceae algae, the closest living relatives of early land plants, has CB5D function.

F5H did not emerge in bryophyte, fern, and gymnosperm species

The discovery of CB5D function in bryophytes, ferns, and gymnosperms inspired us to reexamine whether S-lignin and *F5H*-like genes ever emerged in these land lineages, even though previous studies have clearly established that S-lignin biosynthesis and the presence of F5H enzyme are restricted to angiosperms and the primitive vascular plant Selaginellales (Weng et al. 2008a, 2010). Applying the conventional thioacidolysis method that cleaves the β-aryl ether bonds of lignin polymers, we examined lignin composition in the fern *A. aleuticum* and the gymnosperm Norway spruce (*P. abies*). Both species possess the *CB5D-like* genes that rescued *cb5d-1* defects in transgenic plants (Fig. 2). Additionally, we examined lignin composition of loblolly pine (*P. taeda*), a species with a *CB5D-like* gene that failed to rescue the *cb5d-1* defects (Fig. 2). Consistent with previous studies (Weng et al. 2008b), we detected no S-lignin monomers in any of the examined plants (Supplementary Fig. S7A). This result suggests that the retention of CB5D function in a fern (*A. aleuticum*) and Norway spruce (*P. abies*) is not related to S-lignin biosynthesis. Furthermore, we conducted a search for potential F5H homologs using AtF5H1 as a query against protein databases of liverwort (*M. polymorpha*), moss (*P. patens*), and white spruce (*P. glauca*) whose genomes contain a *CB5D-like* gene capable of rescuing the *cb5d-1* mutant. We also included Ginkgo (*G. biloba*), a species reported to produce S-lignin subunits (Uzal et al. 2009), although its *CB5D-like* gene failed to rescue *cb5d-1* deficiency (Fig. 2). All obtained hits shared low sequence similarity to AtF5H1, ranging from 34% to 44% at the amino acid level. We chose 4 of these candidate genes, i.e. Mapoly0025s0014.1, Pp3c1_1880, PGL00005126, and GBI00006168, with relatively higher similarity to AtF5H1, synthesized their coding sequences, and placed them under the control of the Arabidopsis *C4H* promoter, for transformation into the mutant *fah1-2* defective in S-lignin biosynthesis due to a mutation in *AtF5H1* (Chapple et al. 1992; Supplementary Fig. S7, B and C). However, we detected neither 5-hydroxylated phenolics (sinapoyl esters) in the leaves nor S-lignin monomers in the stems of the primary transformants. These data indicate that the selected candidates do not functionally complement *F5H* deficiency (Supplementary Fig. S7, D and E). Therefore, it is most likely that no functional *F5H* homologs emerged in bryophytes or gymnosperm, regardless of whether these plants possess a functional *CB5D*.

Selaginella does not possess a functional CB5D

Selaginella synthesizes S-lignin in its cortex (Weng et al. 2010). An independently evolved cytochrome P450 CYP788A1, namely SmF5H, was characterized as being responsible for its S-lignin

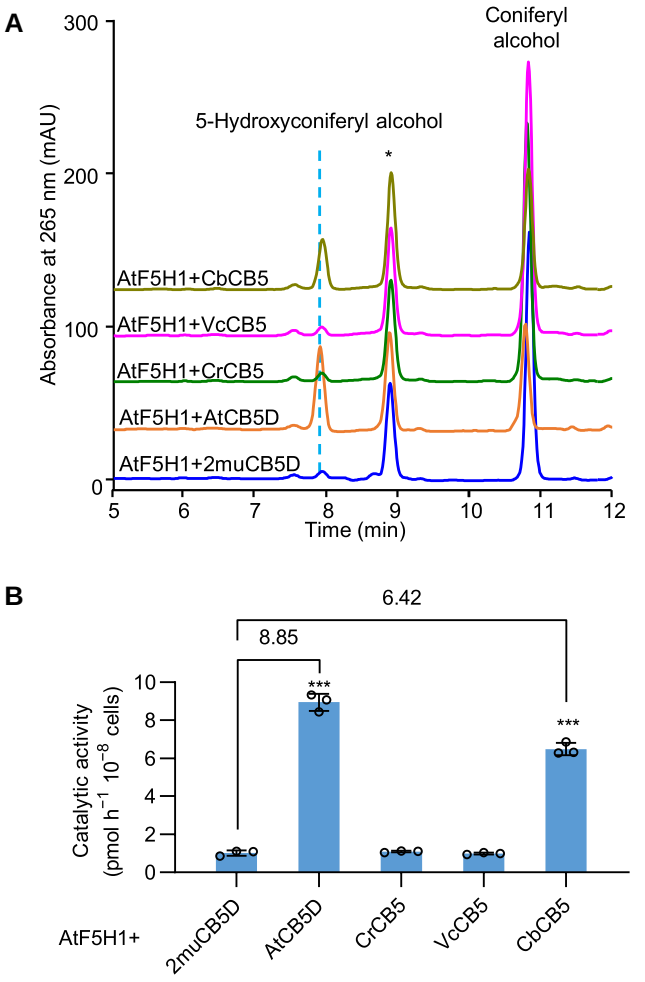

Figure 4. Characterization of algal CB5s via whole-cell biocatalytic assays. **A)** HPLC-UV profiles of the F5H-catalyzed biotransformation of coniferyl alcohol to 5-hydroxyconiferyl alcohol in *E. coli*. The asterisk indicates a nonspecific peak. AtF5H1 was fused to *C. reinhardtii* CB5 (CrCB5), *V. carteri* CB5 (VcCB5), or *C. braunii* CB5 (CbCB5), respectively, and the resulting fusion protein was produced in *E. coli* cells. Arabidopsis CB5D (AtCB5D) and its mutant variant 2muCB5D were included as the controls. **B)** Calculated catalytic activities of different AtF5H1-CB5 fusion proteins. Coniferyl alcohol was fed as substrate, and the 5-hydroxyconiferyl alcohol product was monitored by HPLC. The fold changes are shown above the horizontal lines, compared to control cells harboring the CB5 variant 2muCB5D. Data are means ± SD from 3 biological replicates. Asterisks indicate significant differences with ***$P < 0.001$ (2-tailed Student's *t* tests).

biosynthesis (Weng et al. 2008b, 2010). However, it is unclear whether this early evolved primitive vascular plant F5H also requires CB5 as the electron donor for its catalysis, similar to its counterpart AtF5H1. Upon scrutinizing the *Selaginella* genome, we identified 2 *CB5* homologs, Sm271829 (*S. moellendorffii* CB5-1, *SmCB5-1*) and Sm148632 (*SmCB5-2*). Phylogenetically, *SmCB5-1* clusters with hornwort *AaCB5-1*, liverwort *MpCB5-2*, and moss *PpCB5-1* and *CB5-2*, positioned within the *CB5B/D*-like clade, while *SmCB5-2* does not cluster within any clade (Fig. 1B), suggesting both sequence and potential functional divergence. When constructs encoding a fusion for either SmCB5 with YFP were infiltrated in *N. benthamiana* leaf epidermal cells, we observed a typical ER membrane localization pattern (Fig. 5, A and B), reminiscent of AtCB5D localization (Fig. 3D; Gou et al. 2019). We asked whether SmCB5s like AtCB5D might function as an electron donor for F5H-catalyzed reaction in planta by introducing each *SmCB5*

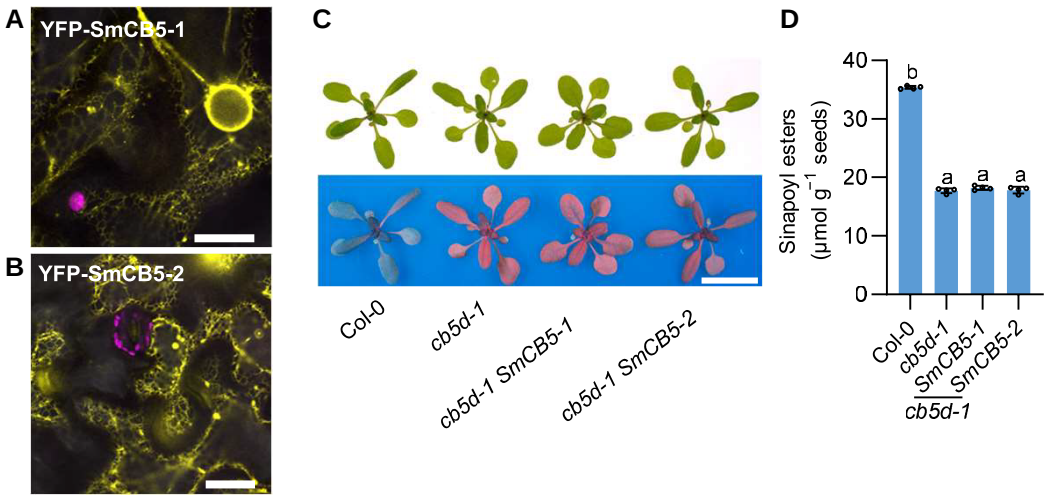

Figure 5. Subcellular localization of SmCB5s and complementation assay of the *cb5d-1* mutant. **A and B)** Representative fluorescence patterns of YFP-SmCB5-1 **A)** and YFP-SmCB5-2 **B)** in *N. benthamiana* leaf cells. Scale bars, 20 μm. **C)** Representative photographs of 18-d-old Col-0 (WT), *cb5d-1*, and 2 T2 transgenic lines harboring *SmCB5s* in the *cb5d-1* background under visible light (upper) and UV light (lower). Scale bar, 2 cm. **D)** Content of sinapoyl esters in mature seeds of Arabidopsis T3 transgenic lines. Seeds from a single plant were pooled as 1 biology replicate. About 8 mg of them was weighted for analysis. Data are means ± SD from 4 biological replicates. Different lowercase letters indicate significant differences ($P < 0.05$, 1-way ANOVA test followed by Tukey's test).

gene, along with a set of *CB5* homologs from other species, under the control of the *AtC4H* promoter into the *cb5d-1* mutant (Fig. 2). We examined their abilities to rescue the *cbd5-1* defects for biosynthesis of sinapoyl esters and S-lignin, as described in Fig. 2.

Despite confirmed expression of each transgene in the generated transgenic lines (Supplementary Fig. S8), the rosette leaves of all primary transgenic plants showed a red appearance when illuminated under UV, similar to the *cb5d-1* mutant but unlike WT plants (Fig. 5C). The red color is from chlorophyll autofluorescence and is a typical indicator of no or a low accumulation of sinapate phenolics that generate blue autofluorescence (Ruegger and Chapple 2001). When measuring the contents of leaf and seed sinapoyl esters and stem S-lignin of transgenic plants, we found that all 5-hydroxylation–derived phenolics remain at levels similar to those detected in the *cb5d-1* mutant, confirming no restoration of biosynthesis for syringyl type of phenolics in the transgenic plants harboring either *SmCB5-1* or *SmCB5-2* (Figs. 2 and 5D). These results suggest that neither SmCB5s possess CB5D function that can support AtF5H1 catalysis.

Since SmF5H evolved independently from its counterpart AtF5H1, it is possible that SmCB5s coevolved with SmF5H and prefer to functionally associate with it rather than AtF5H1. To investigate this possibility, we first expressed *SmF5H* under the control of the *AtC4H* promoter in the *fah1-2* mutant and confirmed the substantial restoration of the accumulation of both leaf sinapoyl esters and stem S-lignin in these transgenic lines (Supplementary Fig. S9). This result indicates that SmF5H can function as a ferulate 5-hydroxylase in Arabidopsis, as previously reported (Weng et al. 2008b).

Subsequently, we generated a homozygous *fah1-2 cb5d-2* double mutant via genetic cross (Supplementary Fig. S10) and introduced *SmF5H* alone or in combination with *SmCB5-1* or *SmCB5-2* into the mutant background (Supplementary Fig. S11). Similar to *fah1-2*, the *fah1-2 cb5d-2* double mutant exhibited nearly no accumulation of syringyl-type phenolics (Fig. 6). Introduction of *SmF5H* into the *fah1-2 cb5d-2* mutant rescued the biosynthesis and accumulation of seed sinapoyl esters neary to the WT level, although leaf sinapoyl esters and stem S-lignin were only partially rescued (Fig. 6, A to D). However, coexpression of *SmF5H* with either *SmCB5* in *fah1-2 cb5d-2* did not show an additional enhancement in sinapoyl ester content or S-lignin accumulation compared to the double mutant lines harboring *SmF5H* alone (Fig. 6). These findings strongly indicate that SmCB5 homologs are not the obligatory electron donors for SmF5H-catalyzed reactions in planta.

Notably, expressing *SmF5H* in *fah1-2 cb5d-2* resulted in a discernibly higher level of restoration of leaf and seed sinapoyl esters and stem S-lignin, compared to lines expressing *AtF5H1* in the same background (Fig. 6). These data further suggest that SmF5H functions independently from the CB5 electron donor in planta. Taken together, these results indicate the absence of a functional CB5D in *Selaginella*.

SmCB5s are functional electron carriers but not required by SmF5H

Given the inability of SmCB5s to rescue *cb5d-1* deficiency in syringyl-type phenolic biosynthesis, we aimed to further examine whether SmCB5s are capable of electron transfer. To this end, we produced recombinant SmCB5 proteins with their C-terminal transmembrane domains removed in *E. coli*. Both purified recombinant SmCB5s displayed the typical CB5 absorptive Soret band at 413 nm. Upon treatment with the reducing agent sodium dithionite, they were readily reduced, showing a characteristic absorbance at 423 nm (Fig. 7, A and B). This spectral shift resembles that of other characterized CB5s (Zhao et al. 2023), indicating their proper functionality.

When incubated with recombinant Arabidopsis CBR1 in the presence of NADH, both SmCB5s displayed a steady transition to the absorbance at 423 nm, indicative of an effective reduction (Fig. 7C). These data suggest that both SmCB5s can similarly accept electrons from reductant NADH through CBR. However, when NADPH and recombinant ATR2, a CPR from Arabidopsis, were incubated with SmCB5s, the 2 SmCB5 proteins displayed different reducing behaviors. SmCB5-2, akin to typical CB5 proteins, showed a steady transition from an oxidized state to a reduced state, eventually reaching and staying at the maximal reduction level (Fig. 7, D and F). However, SmCB5-1, the CB5D-like protein, appeared to be rapidly reduced by ATR2 at the initial reaction stage, followed by a gradual return to its oxidized state (Fig. 7, D and E).

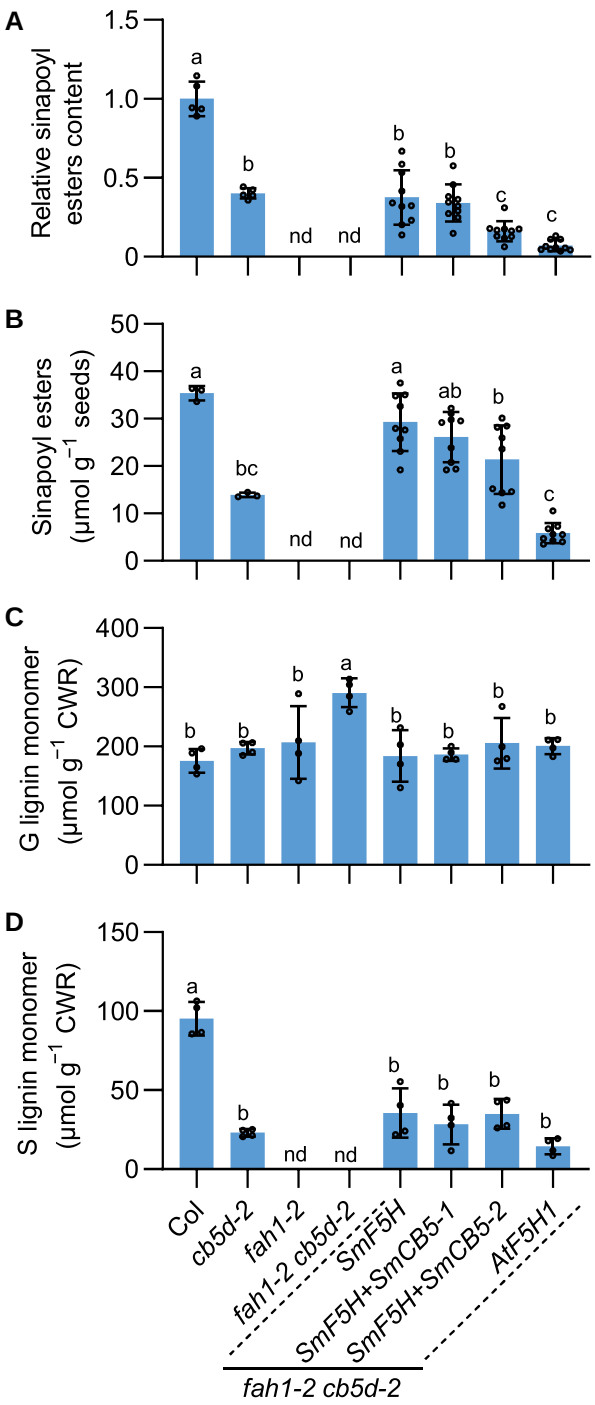

Figure 6. Complementation of the *fah1-2 cb5d-2* mutant with *SmF5H* or *SmCB5* genes. **A)** Relative contents of sinapoyl esters in rosette leaves of 4-wk-old primary transformants. One rosette leaf was sampled from individual 4-wk-old primary transgenic plants, with 2 leaves combined to represent 1 biology replicate for analysis of sinapoyl esters. **B)** Content of sinapoyl esters in T2 mature seeds. Seeds from a single transgenic plant represent 1 biology replicate. **C and D)** Contents of G-lignin monomers **C)** and S-lignin monomers **D)** determined by thioacidolytic analysis in 11-wk-old stems of primary transformants. Stems from 2 to 4 plants were pooled to represent 1 biology replicate. Data in **A to D)** are means ± SD from 3 to 10 biological replicates. Different lowercase letters indicate significant differences ($P < 0.05$, 1-way ANOVA test followed by Tukey's test). nd, not detected; CWR, cell wall residuals.

To better assess the electron transfer capacity of SmCB5s to F5H, we employed the *E. coli* whole-cell biocatalytic system. We fused SmCB5-1 or SmCB5-2 to AtF5H1 and produced the fusion protein in *E. coli*; neither of the fusions enhanced AtF5H1 catalytic activity (Fig. 7G), in stark contrast to the substantial enhancement of AtF5H1 activity via its coproduction with AtCB5D in the same assay system (Fig. 3J). In parallel, we fused SmF5H to the Arabidopsis electron transfer component AtCB5D or ATR2. Coproduction of SmF5H with AtCB5D had no effect on the performance of SmF5H. By contrast, when ATR2 was coproduced with SmF5H, the catalytic activity of SmF5H was enhanced up to 7-fold (Fig. 7G), strongly suggesting that SmF5H, like Arabidopsis C4H and C3′H (Gou et al. 2019; Zhao et al. 2023), relies solely on CPR as its redox partner.

Next, we fused SmF5H to SmCB5-1, SmCB5-2, or SmCPR and coproduced each fusion protein in *E. coli*. Again, neither SmCB5s improved the catalytic efficiency of SmF5H. However, SmCPR showed substantial enhancement of the catalytic activity of SmF5H (Fig. 7H). All these results strongly indicate that SmF5H relies solely on the CPR-mediated electron transfer system and is independent of CB5 as an electron donor.

Helix 5 of the CB5 heme-binding domain influences its functional specialization

While neither SmCB5-1 nor SmCB5-2 supports F5H catalysis, it is worth noting that *SmCB5-1* phylogenetically clusters with *MpCB5-2* and *PpCB5-1*, grouped within the *CB5B/D*-like clade (Fig. 1B). Both MpCB5-2 and PpCB5-1 are functional CB5Ds (Fig. 2). This phylogenetic relationship suggests that SmCB5-1 may theoretically also possess the evolved CB5D functionality as does its evolutionary sister MpCB5-2. This paradox prompted us to explore the underlying reasons behind the loss of CB5D function in SmCB5-1.

Homology modeling–based structural determination of AtCB5D and SmCB5-1 via AlphaFold2 (Jumper et al. 2021) revealed a similar overall structure with 5 α-helices constituting their cytochrome b_5 domains, with Helices 2 to 5 (H2 to H5) encompassing the heme-binding domain (Supplementary Fig. S12, A and B). Comparing CB5 orthologs that show CB5D function in the *cb5d-1* complementation assay (classified as Group 1) with those lacking CB5D function (Group 2; Supplementary Fig. S12C), we noticed notable distinctions in the amino acid residues of H5 within the cytosolic heme-binding domain (Fig. 8A, Supplementary Fig. S12C). Among the 10 amino acid residues of the H5 fragment, observable substitutions occurred between negatively charged acidic amino acid residues and uncharged or positively charged residues in the 2 classified groups (Fig. 8A, Supplementary Fig. S12C). In Group I, H5 contains a conserved X7-D/E-E/D-X sequence, designated as the H5-D/E motif, where X represents nonnegatively charged amino acid residues (Fig. 8A), whereas those 2 acidic residues (Asp and/or Glu) were frequently changed in Group 2 of CB5 proteins. Conversely, there was a high frequency of acidic residues occurring in the X7 region of Group 2 proteins. Specifically, the negatively charged Asp-73 and Glu-74 residues in AtCB5D were respectively replaced with His and Ser in SmCB5-1 and with the positively charged Lys (for Asp-73) in SmCB5-2. Additionally, the uncharged Ala-70 in the X7 region of the AtCB5D H5 sequence was replaced with Asp in SmCB5-1 (Fig. 8B).

To assess whether these changes affected the functionality of SmCB5s as bona fide CB5D, we generated 2 mutant variants for the CB5D-like SmCB5-1 with 3 mutations D67A, H70D, and S71E, named SmCB5-1-H5, and for the non-CB5D-like SmCB5-2 with the K81D mutation, designated SmCB5-2-H5. Before conducting an in planta complementation test, we produced and purified

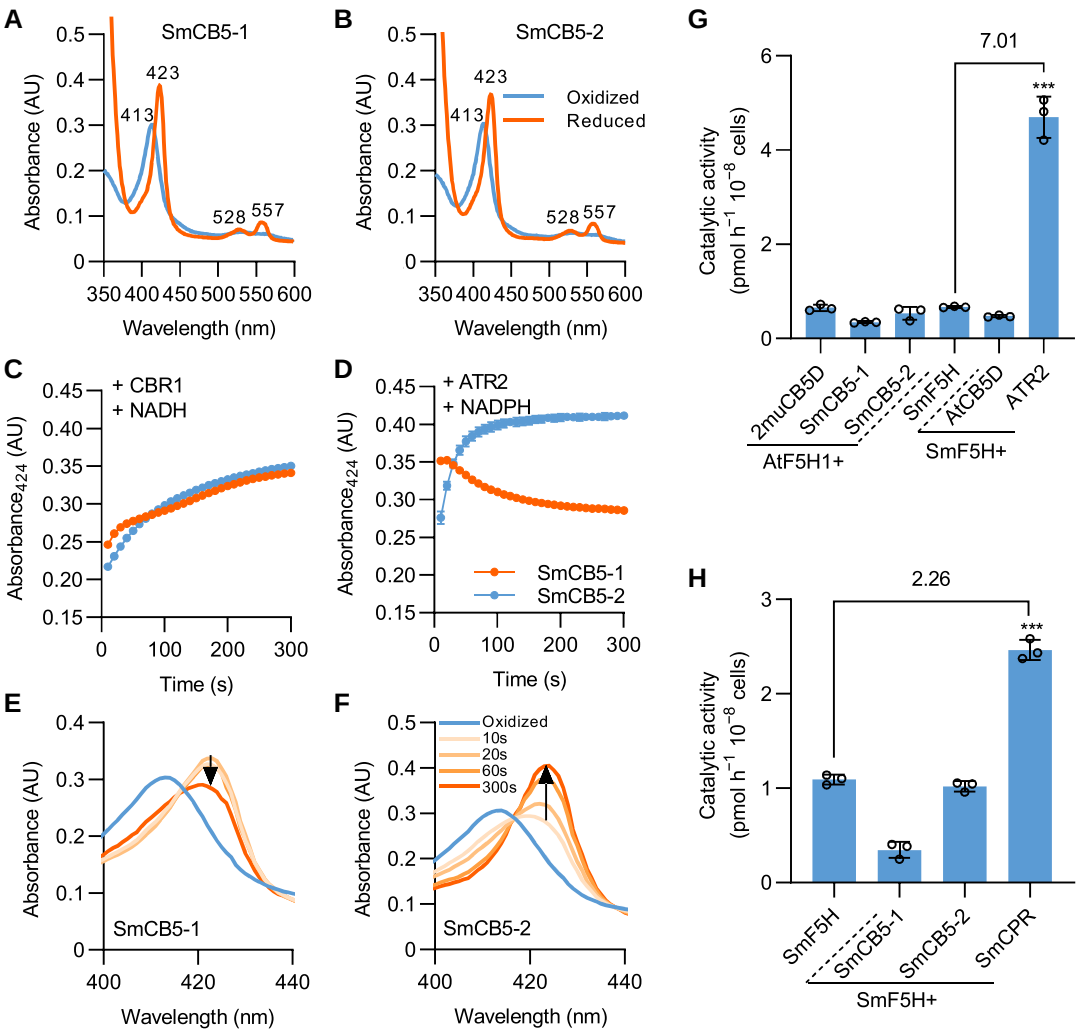

Figure 7. Characterization of SmCB5 functional properties. **A and B)** UV-Vis absorbance spectra of recombinant SmCB5-1 **A)** and SmCB5-2 **B)** in the oxidized or sodium dithionate–reduced state. **C and D)** Changes in absorbance at 424 nm of SmCB5s (7 μM) over the indicated duration of reaction with 7.2 nM CBR1 in the presence of 40 μM NAHD **C)** or 720 nM ATR2 in the presence of 40 μM NADPH **D)**. Absorption was recorded over 30 cycles in 10-s intervals. Data are means ± SD from 3 independent experiments. **E and F)** UV-Vis absorbance spectra of SmCB5-1 **E)** and SmCB5-2 **F)** in the oxidized state (prior to the addition of NADPH and ATR2) and reduced state (after the addition of NADPH and ATR2 for 10, 20, 60, or 300 s). Three replicates were performed with the same results; 1 is presented. The arrows indicate the direction of the change of maximum Soret absorption. **G and H)** Catalytic activities of different F5H-CB5 or F5H-CPR fusion proteins in *E. coli* whole-cell assays. Coniferyl alcohol was used as substrate, and the 5-hydroxyconiferyl alcohol product was monitored by HPLC. The fold changes are shown above the horizontal lines, compared to control cells harboring SmF5H alone. Data are means ± SD from 3 biological replicates. Asterisks indicate significant differences with $^{***}P < 0.001$ (2-tailed Student's *t* tests).

the truncated SmCB5-1-H5 variant lacking its transmembrane domain from *E. coli* and confirmed its redox property spectroscopically. The oxidized and sodium dithionite–reduced SmCB5-1-H5 exhibited typical CB5 characteristics (Fig. 8C). When SmCB5-1-H5 was incubated with CBR1 in the presence of NADH, it exhibited a more efficient reduction compared to its parental SmCB5-1, suggesting a higher electron transfer capacity (Fig. 8D). Moreover, upon incubation with ATR2 and NADPH, SmCB5-1-H5 began to display a typical CB5 reducing pattern, similar to that of AtCB5D (Zhao et al. 2023), which sharply contrasted with that of its parental SmCB5-1 protein (Fig. 8E).

Subsequently, we coproduced each SmCB5 mutant variant with AtF5H1 or SmF5H in *E. coli*, for a whole-cell biocatalytic assay. We first validated whether the *E. coli* endogenous redox systems could reduce SmCB5s. We thus isolated total proteins from the *E. coli* strain BL21 (DE3) and then incubated them with purified WT SmCB5s and the SmCB5-1-H5 variant in the presence of reductant NADPH or NADH. All 3 examined CB5 proteins displayed a spectral shift toward their reduced states, although the levels of their reduction differed (Supplementary Fig. S13). SmCB5-1-H5 showed a more effective reduction than the parental WT protein (Supplementary Fig. S13E). These data demonstrate that the *E. coli* endogenous redox system(s) can at least partially reduce SmCB5s.

We then supplied the phenolic substrate coniferyl alcohol to the *E. coli* cell cultures producing the AtF5H1-SmCB5 or SmF5H-SmCB5 fusion proteins. Compared to the parental SmCB5-1, coproduction of SmCB5-1-H5 with AtF5H1 increased P450 catalytic activity by over 9-fold, suggesting that the mutations with acidic amino acid residues in H5 confer CB5D activity (Fig. 8F). Similarly, coproduction of SmCB5-2-H5 with AtF5H1 also slightly elevated P450 activity by 1.5-fold compared to its parental protein, the non-CB5D-like SmCB5-2 (Fig. 8F). Nevertheless, neither SmCB5 variants enhanced SmF5H catalytic activity when

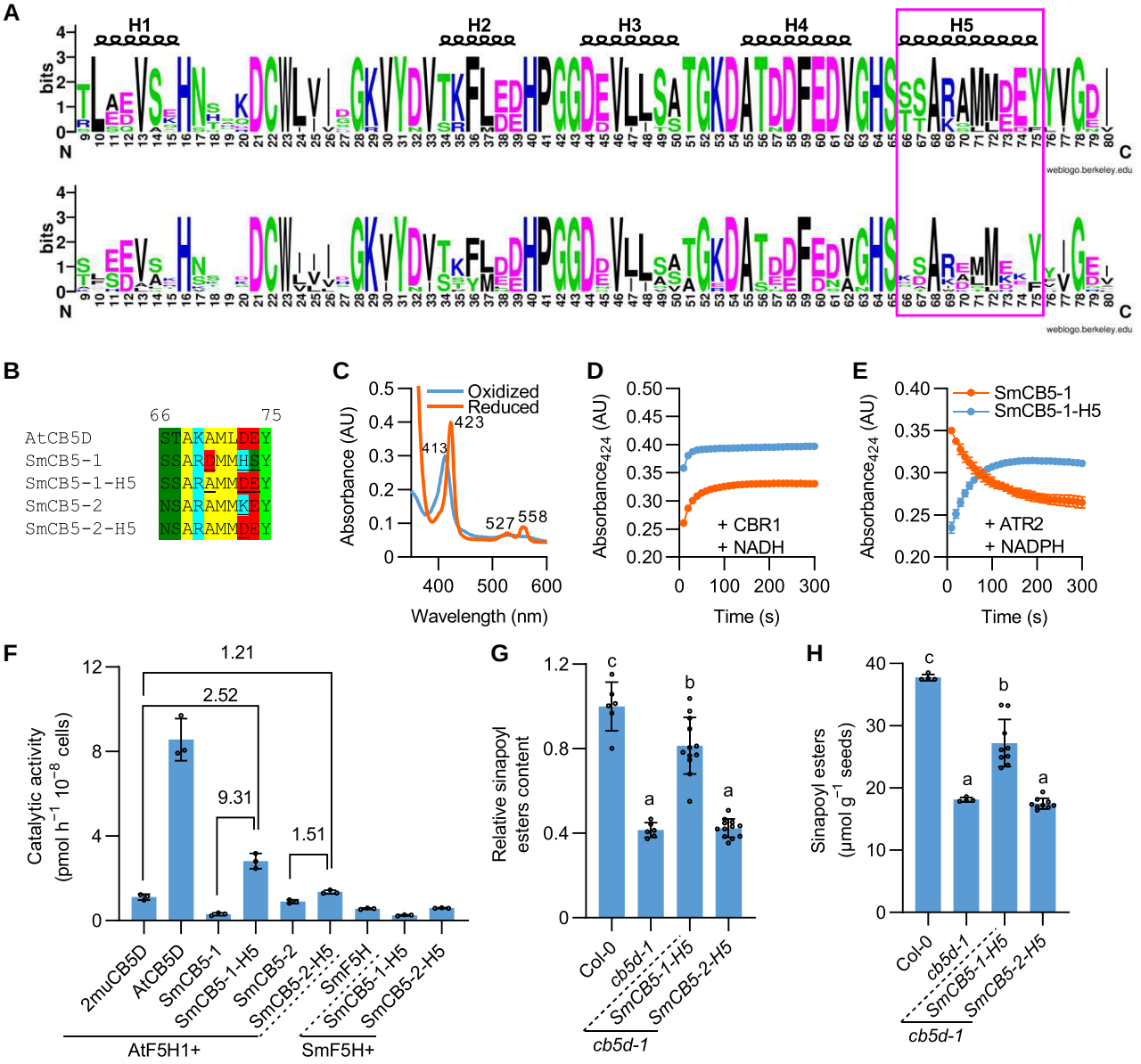

Figure 8. Characterization of the functional influence from Helix 5 of SmCB5. **A)** WebLogo presentation of the cytochrome b_5 domain of group I (upper) and group II (lower) CB5s. The Helix 5 (H5) portion is highlighted with a rectangle. **B)** Amino acid sequences of the H5 region from AtCB5D, SmCB5s, and generated SmCB5 mutant variants. **C)** UV-Vis absorbance spectra of the mutant variant SmCB5-1-H5 in its oxidized and sodium dithionate–reduced states. **D and E)** Reduction of recombinant SmCB5-1 and SmCB5-1-H5 by CBR1 or ATR2 in the presence of NADH **D)** or NADPH **E)**. Changes in absorbance at 424 nm of SmCB5-1 (7 μM) and SmCB5-1-H5 (7 μM) over the indicated duration of reaction with 7.2 nM CBR1 in the presence of 40 μM NADH **D)** or 3.6 μM ATR2 in the presence of 40 μM NADPH **E)**. The absorption was recorded for 30 cycles in 10-s intervals. Data are means ± SD from 3 independent experiments. **F)** Catalytic activity of different F5H-CB5 fusion proteins in *E. coli* whole-cell biocatalytic assays. Coniferyl alcohol was used as substrate, and the 5-hydroxyconiferyl alcohol product was monitored by HPLC. The fold changes are shown above the horizontal lines compared to the indicated control cells. Data are means ± SD from 3 independent experiments. **G and H)** Relative contents of sinapoyl esters in the leaves of 4-wk-old primary transgenic plants harboring *SmCB5-1-H5* or *SmCB5-2-H5* **G)** and content of sinapoyl esters in their T2 mature seeds **H)**. One rosette leaf from one 4-wk-old individual primary transgenic plant was sampled, and 2 leaves were pooled to represent 1 biology replicate for analysis of sinapoyl esters. Seeds from individual lines were collected, representing 1 biology replicate for the analysis of seed sinapoyl esters. Data are means ± SD from the indicated biological replicates. Different lowercase letters indicate significant differences ($P < 0.05$, 1-way ANOVA test followed by Tukey's test).

coproduced in *E. coli*. These results further confirm that SmF5H does not necessarily require CB5 as redox partner for its catalysis, despite the SmCB5 variant being the functional electron carrier both in vitro and in *E. coli*.

Next, we introduced the genes encoding each SmCB5 mutant variant into the Arabidopsis *cb5d-1* mutant, under the control of the *AtC4H* promoter (Supplementary Fig. S15). Measuring the contents of sinapoyl esters in the leaves and seeds of transgenic lines, we found that *SmCB5-1-H5* but not *SmCB5-2-H5* substantially rescued *cb5d-1* defects (Fig. 8, G and H). These data suggest that evolving and preserving the H5-D/E motif in the CB5D-like protein are vital to conferring its electron donor function in supporting F5H catalysis. Additionally, these data also suggest that for non-CB5D-like proteins such as SmCB5-2, apart from the H5-D/E motif, additional sequence element(s) contributes to their functional specialization.

By searching and comparing the occurrence of Helix 5 sequences in green lineages with available genomic or transcriptomic

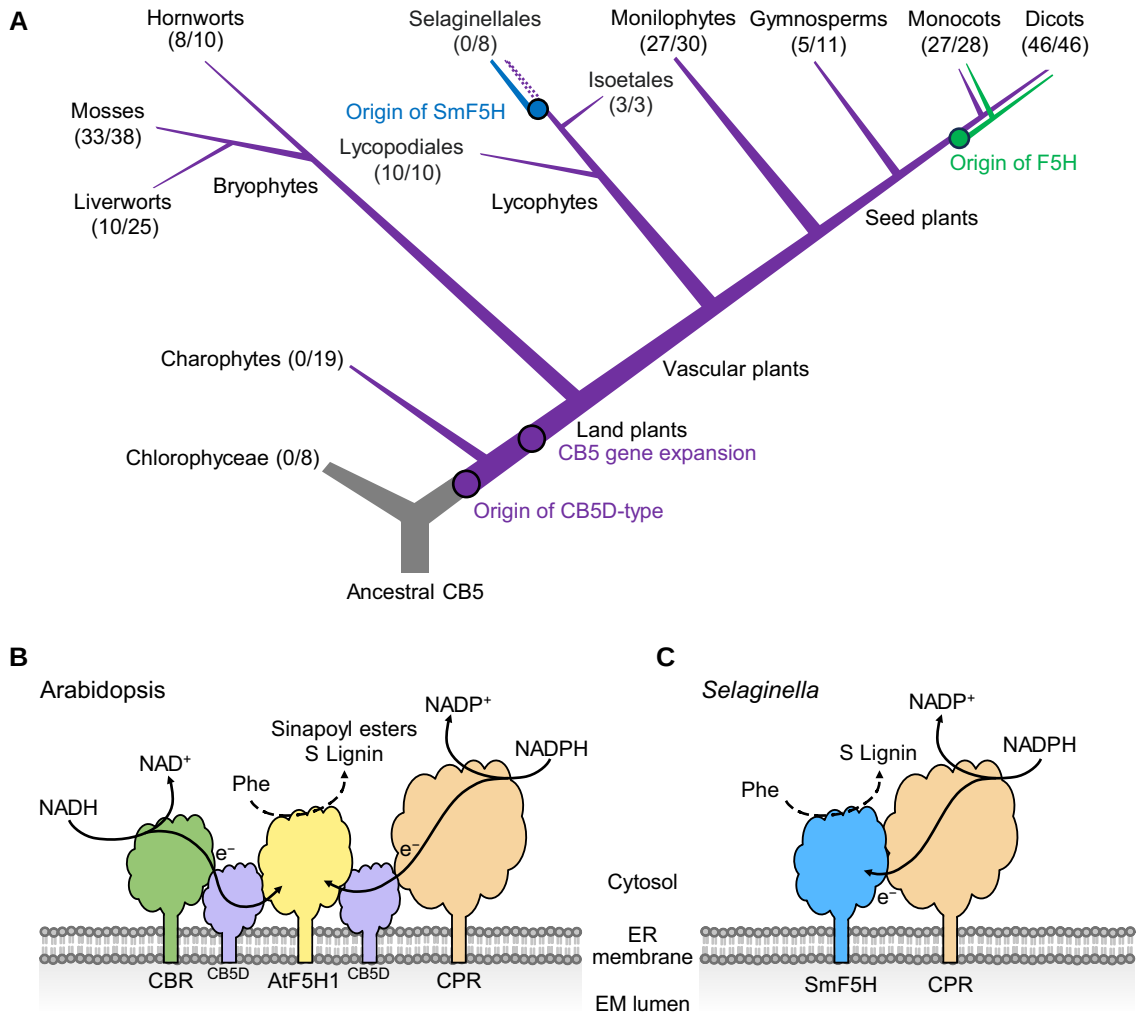

Figure 9. Evolutionary scheme of CB5D-type and F5H proteins and models of S-lignin biosynthetic cytochrome P450 systems in green lineages. **A)** Evolutionary history of CB5D- type proteins and F5H in green lineages. The simplified plant phylogenetic tree illustrates the origin of CB5D-type proteins and their phylogenetic relationship with F5H. The numbers in brackets indicate the conservation of the H5-D/E motif in the surveyed species. A species is counted if it possesses at least 1 CB5 protein containing the H5-D/E motif; otherwise, the count for that plant species was 0 (see detailed information in Supplementary Data Set 3). **B and C)** Diagrams of the cytochrome P450 system and the electron transfer pathway(s) for F5H-catalyzed phenolic biosynthesis in Arabidopsis **B)** and *Selaginella* **C)**. AtF5H1 employs both NADPH–CPR and NADH–CBR pathways, with CB5D acting as an electron shuttle hub, delivering electrons from NADPH–CPR and/or NADH–CBR to AtF5H1 for biosynthesis of sinapoyl esters and S-lignin. SmF5H exclusively requires CPR as the sole redox partner for its catalytic activities.

data, we counted the numbers of plant species that possess at least one of the recognized H5-D/E motifs. Notably, all examined dicots and 27 out of 28 monocots retained the sequence containing the H5-D/E motif (Fig. 9A, Supplementary Data Set 3). The absence of this motif in the monocot species *Phalaenopsis equestris* might be attributed to the currently low quality of its genome sequence. Within bryophytes, monilophytes, and gymnosperms, not all species possessed the H5-D/E motif, suggesting potential loss of the early evolved *CB5D* gene in certain lineages during evolution. Interestingly, while none of Selaginellales species had the H5-D/E motif, all species in 2 other orders of lycophytes, Isoetales and Lycopodiales, retained this motif. These findings underscore the significance of Helix 5 in conferring and retaining the functional evolution of CB5D function in land plants. Interestingly, none of the algal species had a characteristic H5-D/E motif (Supplementary Fig. S14 and Data Set 3), even in *C. braunii* CB5 that displayed typical CB5D function in supporting AtF5H (Fig. 4). This finding suggests that additional sequence or structural features govern the functional diversification and specialization of the ancient CB5 proteins. Nevertheless, CbCB5 emerged with a Glu residue at the corresponding Glu/Asp-74 of the H5 helix sequence, indicating that the characteristic X7-D/E-E/D-X motif of land species might have arisen from the sequences of primitive CB5s from charophyte algae and further developed within land plants.

Discussion

CB5D-type function evolved prior to plant terrestrialization

The emergence of vascular plants on Earth occurred approximately 450 million yr ago (Renault et al. 2019). A hallmark of vascular plants is the presence of the phenolic lignin heteropolymer in xylem and other sclerified cell types (Renault et al. 2019). Unlike H-lignin and G-lignin subunits that are present in all vascular plants, S-lignin is taxonomically restricted to certain lineages, such as angiosperms and Selaginellaceae. Although some species of ferns and gymnosperms have been reported to produce

a trace amount of S-lignin in old biochemical studies (Towers and Gibbs 1953; Faix et al. 1977), this assertion remains unverified (Weng et al. 2008b). Accordingly, the emergence of F5H (CYP84A1), along with the rise of S-lignin, has been regarded as recent evolutionary events, occurring approximately along with the emergence of angiosperms in the later Devonian period, about 350 million yr ago (Weng and Chapple 2010; Renault et al. 2019). In stark contrast to the anciently evolved cytochrome P450 C4H and C3′H in the phenylpropanoid biosynthetic pathway (Renault et al. 2017; Alber et al. 2019), the recently evolved F5H uniquely requires CB5 as an indispensable redox partner for its catalysis (Zhao et al. 2023). CB5D is a member of the CB5 family in Arabidopsis. It is physically and functionally associated with F5H, supporting S-lignin biosynthesis (Gou et al. 2019). Therefore, this specialized CB5 redox partner might have co-evolved with F5H (CYP84A1) and emerged with the advent of F5H-catalyzed S-lignin biosynthesis. Contrary to this intuitive view, our present study reveals that the CB5D type of function first arose in the land plant ancestors, charophytic algae, much earlier than the emergence of angiosperm F5H and S-lignin biosynthesis.

Validated with *E. coli* whole-cell biocatalytic assays for enhancing F5H catalytic activity, CB5 from *C. braunii*, one of the most morphologically complex extant Charophyta (Nishiyama et al. 2018), surprisingly displayed a similar function as AtCB5D (Fig. 4). It is widely believed that charophytic algae are the closest living relatives of land plants. The ancestral lineage of charophytes emerged onto and colonized land 450 to 500 million yr ago. These organisms adapted to terrestrial conditions, becoming capable of surviving and reproducing when fully exposed to the atmosphere, and some members ultimately evolved into land plants (Domozych et al. 2016). The emergence of CB5D function in these algal species undergoing habitat transition suggests its crucial evolutionary role(s) for plant terrestrialization. The importance of this anciently evolved CB5D function in plant evolution is further evidenced by its high conservation across a broad range of embryophytes. Validated with genetic rescue of the Arabidopsis *cb5d-1* mutant deficient in the biosynthesis of syringyl-type phenolics (Fig. 2) and/or with *E. coli* whole-cell biocatalytic assays (Fig. 3), we confirmed that the CB5D-type function was retained within primitive bryophytes (liverwort, hornwort, and moss species) and in monilophytes (*S. cucullata*, *A. evecta*, and *A. aleuticum*), gymnosperms (*P. glauca* and *P. abies*), and angiosperms (Arabidopsis, poplar, petunia, and rice; Fig. 2).

Sequence analysis and phylogenetic reconstruction revealed that embryophyte lineages, unlike their aquatic progenitors, contain multiple copies of *CB5* genes, indicating substantial gene expansion, probably achieved through WGD (Qiao et al. 2022), during land plant evolution. Accompanying *CB5* sequence expansion is their functional divergence and specialization. For instance, 2 liverwort CB5s displayed distinct subcellular localizations, with MpCB5-1 found in the outer envelope of chloroplast, while MpCB5-2 that exhibits CB5D-type function localized at the ER membrane, indicating their functional differentiation (Figs. 2 and 3). Such sequence expansion and functional divergence similarly occurred in moss species and were inherited in the descendant tracheophytes. Arabidopsis maintains a chloroplast-localized CB5 (AtCB5A) and 4 ER-resident members (AtCB5B to AtCB5E; Maggio et al. 2007), including the S-lignin biosynthesis–required bona fide AtCB5D.

Charophytic algae constitute a paraphyletic group of freshwater and terrestrial green algae frequently exposed to naturally occurring abiotic stresses such as desiccation, freezing, and high photosynthetic activity, as well as UV irradiation (Holzinger and Pichrtová 2016). This group of species has evolved profound morphological properties and metabolic capacity that enable them to synthesize, for example, flexible primary cell walls, and the ethylene-like phytohormones (Nishiyama et al. 2018). Moreover, some charophycean species produce and accumulate UV protection compounds, such as mycosporine-like amino acids, purpurogallin, and gallic acid phenolic derivates, as well as other unusual soluble phenolics and pigments to protect themselves in conditions of high irradiation (Holzinger and Pichrtová 2016). Similarly, bryophytes also synthesize soluble phenolics, such as flavonoids, to deal with UV irradiation and oxidative stresses, but they lack true lignin (Basile et al. 1999; Umezawa 2003). For example, the model moss *P. patens* produces significant amounts of caffeate-derived monomers in its cuticle (Buda et al. 2013), and a CYP98 ortholog catalyzes the phenolic ring *meta*-hydroxylation to yield caffeate derivatives (Renault et al. 2017). It is interesting to explore whether the early emerged CB5D function in green algae and bryophytes is related to the evolution of adaptive metabolism, such as the biosynthesis of UV-protective soluble phenolics, for stress tolerance. Additionally, as a conserved electron transfer component, CB5 is well documented to associate with cytochrome b_5 reductase for lipid biosynthesis in yeast and mammals (Schenkman and Jansson 2003). During land colonization, pioneering plants had to deal with desiccation by innovating more complex biochemical capacities, such as lipidic cuticle polymers as surface barrier to control water and gas movements. Expanding and diversifying CB5 functionalities might have been instrumental to meet the biosynthetic needs for the formation of more complicated cuticle polyesters and cuticular wax, in addition to the conventional structural lipids.

In summary, it remains to be determined what metabolic roles and biological significance are associated with the early emergence of CB5D function in the ancestors of land plants, including charophytic algae, and in basal land plants lacking the evolved F5H enzyme and S-lignin biosynthesis. Of particular interest is the investigation of whether early-existing CB5D-like proteins can interact with enzymes similar to F5H. Such interactions and functional associations may have played essential roles in the early colonization of plants and their subsequent adaptation to harsh terrestrial environments. Additionally, it would be intriguing to explore whether mutations in the recently evolved angiosperm F5H led to specific interactions with the CB5D protein.

CB5D underwent further species-specific divergence after its innovation

Phylogenic analysis indicated that *CB5D*-like gene sequences are present across all the evolutionarily characteristic embryophytes, suggesting a highly conserved evolution. Nevertheless, the CB5D-type function that supports F5H catalysis for the biosynthesis of syringyl-type phenolics is not always conferred by all CB5D-like proteins. This is particularly exemplified with lycophyte *Selaginella* CB5 proteins. In well-defined *Selaginella* genomic sequences, only 2 *CB5* members can be identified, with *SmCB5-1* clustering within the *CB5B/D*-like clade (Fig. 1). Both SmCB5-1 and SmCB5-2 proteins localized to the ER membrane. However, neither rescued the *cb5d* defects when they were expressed in Arabidopsis, nor did they supported AtF5H1 function or native SmF5H activity in whole-cell biocatalytic assays (Figs. 5 and 6), although both proteins exhibited CB5 redox protein characteristics and can be reduced with reductases CBR1 and ATR2 or the *E. coli* redox systems (Fig. 7, Supplementary Fig. S13). Similarly, the aquatic floating fern *S. cucullata* has 4 *CB5B/D*-like genes. Of the 2 being tested, *S. cucullata* CB5-1 but not *S. cucullata* CB5-2 exhibited CB5D function (Fig. 2). These data suggest that *CB5D*,

after its early emergence and expansion, underwent further evolution and divergence within the more advanced land lineages; in certain lineages such as Selaginellales, its function might have been lost, probably due to the lack of evolutionary pressure for its retention.

The acidic amino acid residues of Helix 5 are crucial for sustaining CB5D function

Examining sequence deviations of CB5D-like proteins revealed substantial variability in the presence and position of acidic amino acid residues within Helix 5. The characteristic H5-D/E motif is conserved in the land plant proteins exhibiting CB5D-type function. The presence or absence of acidic amino acid residues at the defined positions, such as Asp-73 and Glu-74 in AtCB5D, within Helix 5 significantly influences the functional specialization of CB5D-like proteins. SmCB5-1, classified within the CB5B/D-like clade but lacking these crucial acidic amino acid residues in Helix 5, does not possess CB5D function. By contrast, introducing Asp and Glu at their corresponding positions in CB5 proteins that normally do not possess them yielded protein variants with substantial CB5D functionality, enhancing AtF5H1 activity and rescuing *cb5d* deficiency (Fig. 8). Two CB5 homologs, *G. biloba* and *P. taeda* CB5-1s, that failed to rescue the *cb5d-1* defects appear to have retained D/E residues in the defined positions of Helix 5 (Supplementary Fig. S12). Nevertheless, these proteins possess additional acidic amino acid residues in the defined X7 sequence of the H5-D/E motif (X7-D/E-E/D-X), which might affect their electrochemical or structural properties, resulting in the loss of their CB5D function. The exact molecular mechanism remains to be determined. Therefore, the H5-D/E motif can represent a defining sequence feature and an evolutionary indicator for sustaining CB5D function in land plant species. Notably, this motif is ubiquitously found within nearly all land plant lineages except Selaginellales; it is particularly frequent in angiosperms (Fig. 9A), suggesting that CB5D function is largely conserved and flourished in the land plants. However, the current study cannot exclude the possibility that additional sequence/structural features also contribute to the determination of CB5D function, and the CB5 proteins that lack a typical H5-D/E motif may also possess CB5D function. The ancient charophyte CB5 emerged with a CB5D function but does not have a typical H5-D/E motif (Supplementary Fig. S14), suggesting its ancestral role in the evolution of land plant CB5D-type proteins. Presumably, the recognized H5-D/E motif might have evolved and been selected along with CB5 family expansion and diversification to meet the needs of a particular set of functionally associated partners in land plants.

Charged acidic amino acid residues are often implicated in intra/intermolecular electron/proton transfer (Kawano et al. 1998). Presumably, the distinctly positioned acidic charged amino acid residues within Helix 5 might affect the redox potential or other electrochemical and structural properties of the CB5 protein, thus defining its ability as an electron acceptor to obtain electrons from its redox partners, or as an electron donor for its associated oxidative enzyme. This hypothesis is supported by our in vitro reduction assay of the variant SmCB5-1-H5. The variant with its His and Ser replaced with Asp and Glu, respectively, in its Helix 5 demonstrated a more effective and steady reduction when it was exposed to NADPH-dependent ATR2 (Fig. 8) or the total proteins from *E. coli* (Supplementary Fig. S13).

Selaginella F5H does not require CB5D for its catalysis

S. moellendorffii, a basal vascular plant, is capable of synthesizing S-lignin (White and Towers 1967; Weng et al. 2008b). An independently evolved lineage-specific P450, SmF5H, functions as a dual *meta*-hydroxylase, which rescues the AtF5H1-deficient *fah1* mutant (Weng et al. 2008b). Strikingly, when *SmF5H* was expressed in the *fah1-2 cb5d-2* double mutant under the control of the *AtC4H* promoter, it acted more effectively in restoring the biosynthesis of syringyl-type phenolics than when *AtF5H1* was expressed from the same promoter and in the same genetic background (Fig. 6). These different functional effects do not appear to be caused by the different enzymatic properties of SmF5H and AtF5H1, since both enzymes kinetically work on coniferyl alcohol and coniferaldehyde substrates equally well (Weng et al. 2010). Rather, these results suggest differences between the 2 enzymes in their requirements for redox partners. The absence of CB5D notably affected AtF5H1 more than SmF5H (Fig. 6). With the reconstruction of a P450 electron donor system in *E. coli*, and consistent with our previous report (Gou et al. 2019; Zhao et al. 2023), AtF5H1 strictly required CB5D for its catalysis, while SmF5H activity was not affected by either AtCB5D or SmCB5s; instead, the activity was strongly enhanced by NADPH-dependent CPR (Fig. 7). These lines of evidence unambiguously suggest that the independently evolved SmF5H, more like anciently emerged C4H and C3′H and other conventional P450s, relies on the NADPH–CPR electron transfer system (Fig. 9B), differing sharply from its angiosperm F5H counterparts.

SmF5H possesses a significantly distinct protein sequence compared to other F5H members, representing a unique *Selaginella* P450 clade (Weng et al. 2008b). Although the predicted overall structures and key structural elements, such as the E-R-R triad (in K-helix and PERF domain) and heme-binding domain (Bak et al. 2011), appear similar between AtF5H1 and SmF5H, differences arise in the lengths of their α-helices 8 and 10, along with large numbers of different amino acids (Supplementary Fig. S16). Such structural differences might alter the interfaces or the local conformations of the proteins that mediate the physical association with electron donor proteins or change redox potentials/electrochemical properties of the proteins, thus influencing their recruitment of electron transfer systems. Solving the structures of P450–redox partner complexes and further determining their electrochemical properties will shed light on their unique redox catalytic mechanisms. Besides the recognized functional association of CB5D with F5H-catalyzed reactions, sporadic evidence also suggests the potential involvement of CB5 electron donor in other plant P450 monooxygenase systems; for instance, the AtF5H2 (CYP84A4)–catalyzed α-pyrone biosynthesis (Gou et al. 2019), the F3′5′H (CYP75A)–driven delphinidin-based anthocyanin formation (de Vetten et al. 1999), and the CYP71AV1-catalyzed artemisinin biosynthesis (Paddon et al. 2013). Conservation of CB5D-like proteins in the plants lacking the biosynthesis of S-lignin suggests that the evolved CB5Ds might interact with other cognate P450 enzymes in different biochemical reactions, warranting further exploration.

Taken together, our study reveals the early emergence and functional divergence of CB5D in green lineages, which is sharply in contrast to the recent emergence of angiosperm CYP84A1 F5H. These findings suggest that angiosperm F5H utilizes the anciently evolved electron transfer component to assemble a modern biochemical machinery for the biosynthesis of reemerged S-lignin. The F5H in the basal vascular species *Selaginella* represents a

conventional cytochrome P450 monooxygenase system, employing the general NADPH–CPR electron transfer chain for its independently evolved S-lignin biosynthesis. These data hint that the F5H-CB5D catalytic system in angiosperms might confer an evolutionary advantage.

Materials and methods

Plant materials and growth conditions

The plant materials used in this study were *N. benthamiana* and *A. thaliana*. All mutants and transgenic lines were in the Col-0 background. Mutant lines used were *cb5d-1* (SALK_045010), *cb5d-2* (GABI_328H06), and *fah1-2* (CS6172). Transgenic lines used are listed in Supplementary Data Set 4. Homozygous T-DNA insertion mutants were obtained by genotyping genomic DNA as a template. Higher-order mutants were created via crossing the single mutants; the double mutants were identified from the F2 progenies via genotyping through PCR or sequencing, using the primers listed in Supplementary Data Set 5. Seeds were surface sterilized with 70% (v/v) ethanol and sown onto half-strength MS medium (PhytoTech, M524), containing 1% (w/v) agar and 1% (w/v) sucrose for seed germination. After stratification for 3 d at 4 °C, the seeds were germinated and maintained at 22 °C under a 16-h light/8-h dark photoperiod in a growth chamber (BioChambers) equipped with combined LED T5 tubes (PLT-90020, 25W 4000K, PLT, Garland, TX, USA) and incandescent bulbs (PLT-12125S, 40W, PLT, Garland TX, USA) with light intensity of 130 μmol/m^2/s. Seven-day-old seedlings were transferred to moist soil (BM2 seed germination mix, Berger, QC, Canada) and grown until maturity under the same conditions.

Phylogenetic analysis

All CB5 homologous sequences were retrieved by BLASTP search using AtCB5D (At5g48810) as the initial query from the UCSC Genome Browser (https://genome.ucsc.edu/, *Homo sapiens*, *Mus musculus*, and *Danio rerio*), the Saccharomyces Genome Database (https://www.yeastgenome.org/, *S. cerevisiae*), Phytozome 13 (https://phytozome-next.jgi.doe.gov/, *O. sativa*, *S. bicolor*, *P. trichocarpa*, *A. trichopoda*, *S. moellendorffii*, *M. polymorpha*, *P. patens*, *V. carteri*, and *C. reinhardtii*), ORCAE (https://bioinformatics.psb.ugent.be/orcae/, *C. braunii*), the Solanaceae Genomics Network (https://solgenomics.net/, *P. axillaris*), Hornworts (https://www.hornworts.uzh.ch/en.html, *A. agrestis*), PLAZA (https://bioinformatics.psb.ugent.be/plaza/, *P. glauca*, *P. abies*, *P. taeda*, and *G. biloba*), ONEKP (https://db.cngb.org/onekp/, *C. vulgaris*, *A. evecta*, and *A. aleuticum*), and FernBase (https://www.fernbase.org/, *S. cucullata*). All hits were then reciprocally blasted (via BLASTP) against the Arabidopsis protein sequences and discarded if the best hits were not AtCB5A, AtCB5B, AtCB5C, AtCB5D, or AtCB5E. Finally, the obtained hits were manually examined for an intact heme-binding domain and the existence of a C-terminal transmembrane domain, as criteria for a CB5 protein. The full-length protein-coding nucleotide sequences of the obtained 87 *CB5*s (Supplementary Data Set 1) were used for the phylogenetic analysis. The polypeptide coding cDNA sequences were aligned using MUSCLE algorithm integrated within the MEGA X program with default settings (Kumar et al. 2018). The CB5 phylogenetic tree was then inferred using the maximum likelihood method implemented in IQ-TREE-2.2.2.7 program (Hoang et al. 2018; Minh et al. 2020). Phylogenetic tree reconstruction employed the following command line: "bin\iqtree2 -s CDS-full.FAS -alrt 1000 -B 1000". The SYM+I+R5 model was used for the phylogenetic reconstruction, which was selected by the integrated ModelFinder algorithm. One thousand ultrafast bootstraps and 1,000 SH-aLRT tests were performed to determine branch support (Supplementary Files 1 and 2). For phylogenetic analysis with protein sequences, the ClustalW algorithm integrated within MEGA X was used for multiple sequence alignment. The tree was also reconstructed with IQ-TREE-2.2.2.7 program as for the cDNA-based tree above but with the command line "bin\iqtree2 -s Protein.FAS -m JTT+G4 -alrt 1000 -B 1000" (Supplementary Files 3 and 4). The visualization and annotation of the trees were performed in TVBOT (Xie et al. 2023).

Plant transformation

The full-length coding sequences of *CB5* and *F5H* genes from representative plant taxa were synthesized by GenScript (https://www.genscript.com) and ligated into the pMDC32-*ProC4H* vector via *KpnI/SalI* restriction enzyme sites. The full list of synthesized genes is given in Supplementary Data Set 2. All constructs were transformed into the corresponding homozygous *cb5d-1*, *fah1-2*, or *fah1-2 cb5d-2* mutants using the Agrobacterium GV3101 (*Agrobacterium tumefaciens*)–mediated floral dip method (Clough and Bent 1998). The primary transformants were selected on half-strength MS plates containing 15 μg/mL hygromycin B (Gold Biotechnology). After 2 wk, the hygromycin-resistant seedlings were transferred to soil and grown to maturity in a growth chamber with the above-described conditions. For the multigene expression of *SmF5H* and *SmCB5*s, the sequence encoding the T2A self-cleaving peptide was cloned between the *SmF5H* and *SmCB5* sequences. The resulting DNA fragments were ligated into pMDC32-*ProC4H* vector via *KpnI/SalI* sites. The *T2A* coding sequence was included within the primers together with the gene-specific sequences (Supplementary Data Set 5). The *SmCB5-1-H5* and *SmCB5-2-H5* variants were obtained by site-directed mutagenesis using the Gibson assembly method (Gibson et al. 2009).

Quantification of soluble phenolics and lignin composition

Soluble phenolics and lignin composition were quantified following the reported method (Zhao et al. 2023). Briefly, for the analysis of soluble phenolics from leaves or seedlings, about 50 mg of rosette leaves from 4-wk-old plants or 1-wk-old seedlings were collected and placed into a 1.5-mL centrifuge tube and weighed. To each tube, 300 μL of 80% (v/v) methanol containing 80 μM chrysin (as an internal standard, C80105, Sigma) and 2 steel balls were added. The samples were then ball milled for 2 min at 30 Hz using a CryoMill (Retsch). An additional 200 μL of 80% (v/v) methanol containing 80 μM chrysin was then added to each tube. After incubation at 4 °C for 4 h, the samples were centrifuged at 20,000 × *g*, twice. Five microliters of the supernatant was injected into a UHPLC system (Thermo Fisher Scientific) or a HPLC (Agilent) system with the same settings as described (Zhao et al. 2023). Briefly, for both systems, the mobile phase was a mixture of 0.1% acetic acid aqueous solution (A) and 0.1% acetic acid in acetonitrile (B). For the Thermo Fisher UHPLC system, the samples were resolved with a reverse phase C18 column (Luna Omega 1.6 μm C18 100 Å, 150 × 2.1 mm, Phenomenex) using a gradient program: 5% B (0 min), 30% B (1 min), 80% B (19 min), 99% B (20 min), holding for 2 min, then to 5% B (23 min) at a flow rate of 0.1 mL/min. For the Agilent 1100 HPLC system, the samples were analyzed with a reverse phase C18 column (Luna 5 μm C18 [2] 100 Å, 250 × 4.6 mm, Phenomenex) using a gradient program: 15% B (0 min), 15% B (2 min), 50% B (15 min), 100% B (18 min), holding for 4 min, then to 15% B (23 min) at a flow rate of 1 mL/min. The

compounds were detected with UV-Vis diode array detector at 265, 280, 330, and 510 nm wavelengths.

For analysis of soluble phenolics in seeds, about 8 mg of dry seeds was mixed with 300 μL 50% (v/v) methanol containing 1.5% (v/v) acetic acid and 100 μM chrysin (as an internal standard). The mixture was then subjected to ball milling using a CryoMill (Retsch) for 2 min at 30 Hz, followed by the addition of 500 μL of 50% (v/v) methanol containing 1.5% (v/v) acetic acid and 100 μM chrysin to each sample. The samples were incubated at 4°C for 1 h before being centrifuged at room temperature, 20,000 × *g* for 5 min, twice using an Eppendorf 5424R centrifuge. From each supernatant, a 5-μL aliquot was injected into an HPLC (Agilent) system with the same settings as described (Zhao et al. 2023).

For stem lignin analysis, the 10-cm basal stems from 11-wk-old Arabidopsis plants were harvested, lyophilized, and ball milled. The obtained fine powders were extracted with 70% (v/v) ethanol at 65°C for 3 h and then centrifuged to obtain pellets. The pellets were resuspended in 70% (v/v) ethanol, and the extraction was repeated twice. The obtained pellets were then extracted with chloroform/methanol (1:1, v/v) at room temperature for 1 h as did as the ethanol extraction. The process was repeated 3 times, and the pellets were finally extracted with acetone at room temperature overnight. After centrifugation, the residues were collected and dried at room temperature to obtain extractive-free cell wall residues. Approximately 10 mg of extractant-free residues was weighed and placed into a glass vial, followed by addition of a freshly prepared reaction mixture composed of 2.5% boron trifluoride etherate and 10% ethanethiol in dioxane (v/v). The vials were purged with nitrogen gas before being sealed then placed in a 95°C heat block for 4 h with intermittent shaking. After cooling, 0.3 mL of 0.4 M sodium bicarbonate, 2 mL of water, and 1 mL of methylene chloride (containing 1 mg/mL tetracosane) were sequentially added. The vials were vortexed for 1 min and centrifuged at 3,000 × *g* for 5 min (Beckman Coulter Avanti J-15R with JS-4.750 rotor) to separate the phases. The organic phase (approximately 1 mL) was collected and dried in a heat block overnight at 50°C. Subsequently, 0.5 mL of methylene chloride was added to resuspend the dried sample, and 50 μL of the sample was transferred to a new centrifuge tube and dried under the same condition. The sample was then derivatized by adding 50 μL of pyridine and 50 μL of *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide, followed by incubation at room temperature for 5 h. A 1-μL aliquot was injected into an Agilent 7890A gas chromatography-flame ionization detector with the same settings as described (Zhao et al. 2023).

Subcellular localization imaging

The synthesized full-length *CB5* coding sequences of *M. polymorpha* and *S. moellendorffii* were amplified with the primers listed in Supplementary Data Set 5, subcloned into the pDONR207 vector, and confirmed by Sanger sequencing. The *AtCB5A* and *AtCB5D* coding sequences were obtained from the previous cDNA clones (Zhao et al. 2023). The *CB5* coding sequences were individually subcloned into the pEarleyGate 104 vector, resulting in the constructs *35S:YFP-AtCB5A*, *35S:YFP-AtCB5D*, *35S:YFP-SmCB5-1*, *35S:YFP-SmCB5-2*, *35S:YFP-MpCB5-1*, and *35S:YFP-MpCB5-2*. All constructs were individually introduced into Agrobacterium strain GV3101. The overnight culture of each strain was pelleted at 4,000 × *g* for 15 min and resuspended in infiltration buffer (10 mM MES [pH 5.6], 10 mM $MgCl_2$, and 150 μM acetosyringone) to an optical density at 600 nm of 0.5. After incubation at room temperature for 2 h, the strains were infiltrated into the leaves of *N. benthamiana* plants, individually. After 3 d of incubation in a growth chamber at 22°C under a 16-h light/8-h dark photoperiod, the fluorescence images were captured with a TCS SP5 laser scanning confocal microscope (Leica) with excitation at 514 nm and an emission wavelength of 520 to 535 nm for YFP signals. Chloroplast autofluorescence was observed at 636 to 725 nm.

RT-qPCR analysis of gene expression

Total RNAs were extracted from plant materials in triplicate using TRIzol reagent (Thermo Fisher Scientific) following the manufacturer's instructions. Reverse transcription was conducted using 0.5 μg of total RNA and 2 μL of iScript Reverse Transcription Supermix (Bio-Rad) in 10-μL reaction volumes. The reaction mixture was incubated in a thermal cycler at 25°C for 5 min, 46°C for 20 min, and 95°C for 1 min. The first-strand cDNA solution was diluted 5 times with water, and 2 μL of the diluted solution was used as template in a 15-μL reaction with SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). qPCR was performed using a CFX96 Real-Time System (Bio-Rad), and cycle threshold (Ct) values were calculated by the CFX Manager Software v.3.3 (Bio-Rad). Primers used for qPCR are listed in Supplementary Data Set 5. Arabidopsis *PP2A* was used as the housekeeping reference gene, and the data were calculated using the delta-Ct method, i.e. $2^{-\Delta Ct}$ method, where ΔCt = (Ct gene of interest – Ct internal control; Schmittgen and Livak 2008).

Protein production, purification, and reduction assay

For SmCB5s and their variant and MpCB5s, their full-length coding sequences without the sequence encoding the transmembrane domain were PCR amplified and inserted into the pET28a(+) vector via *EcoRI* and *XhoI* restriction enzyme sites. AtCBR1 and ATR2 expression constructs were from previous study (Zhao et al. 2023). The production and purification of AtCBR1, ATR2, and CB5s and the reduction assays were performed as previously described (Zhao et al. 2023). Briefly, the Sanger sequence-verified constructs were transformed into *E. coli* BL21(DE3) cells; positive colonies were used to inoculate 20 mL Terrific Broth and cultured overnight at 37°C. The precultures were used to inoculate 600 mL Terrific Broth and grown at 37°C to an optical density of around 1 at 600 nm. After placing the culture at 15°C for 30 min, 0.5 mM IPTG (final concentration) was added for induction of protein production. Overnight cultures were harvested by centrifugation and stored at −80°C. For protein purification, the pellets were resuspended in lysis buffer (20 mM Tris-HCl [pH7.5], 200 mM NaCl, 10 mM imidazole, 0.1% Triton X-100, 1 mM phenylmethylsulfonyl fluoride [PMSF], 5 mM β-mercaptoethanol, and 1× cocktail protease inhibitors), sonicated in an ice bath with 1 s on/4 s off circle (Fisher Scientific Sonic Dismembrator Model 500 with 1.27 cm diameter tips, 50% amplitude), and spined at 9,682 × *g* (Sorvall RC 5C Plus centrifuge, SS-34 rotor) for 1 h. The recombinant proteins were purified with Ni^{2+}-nitrilotriacetic acid agarose beads (Qiagen) and desalted with Bio-Gel P-6DG gel (Bio-Rad). A reduced state for the CB5 samples was achieved by introducing a small quantity of sodium dithionite. CB5 protein concentration was quantified from the absolute spectrum using extinction coefficient $\epsilon_{413} = 117\ mM^{-1}\ cm^{-1}$ or from the different spectra of the cytochrome catalyzed by sodium dithionite using the extinction coefficient of $\Delta\epsilon$(reduced – oxidized)$_{424-409} = 185\ mM^{-1}\ cm^{-1}$ (Sergeev et al. 2006). The purified protein was aliquoted and stored at −80°C until use. The

recombinant Arabidopsis CBR1 and ATR2 enzymes were produced and purified as previously described (Zhao et al 2023).

The redox assay was performed in 100 μL of 20 mM Tris-HCl (pH 7.5) in a 96-well microplate. Each well contained 7 μM CB5 and 7.2 nM CBR1 or 720 nM ATR2, mixed, and preincubated at room temperature for 5 min. The reduction assays commenced with the addition of either 40 μM NADH (for CBR1) or 40 μM NADPH (for ATR2), followed by the immediate recording of absorbance at 424 nm at 10-s interval using a Spark microplate reader (Tecan, Männedorf).

For the CB5 reduction assay with *E. coli* total proteins, the overnight-incubated BL21 (DE3) cells were pelleted and resuspended in lysis buffer (20 mM Tris-HCl, pH 7.5, 200 mM NaCl), sonicated in an ice bath with 1 s on/4 s off circle (Fisher Scientific Sonic Dismembrator Model 500 with 1.27 cm diameter tips, 50% amplitude), and centrifuged at 10,000 × *g*, at 4°C, twice with an Eppendorf 5424R centrifuge for 1 min. The clear supernatant was harvested as *E. coli* total protein and stored at −80°C until use. The reduction assay was performed in 200 μL of 20 mM Tris-HCl (pH 7.5) in a 96-well microplate. After preincubation of 2.5 μM SmCB5s with 108 μg *E. coli* total proteins at room temperature for 5 min, the reaction was initiated with the addition of either 125 μM NADH or NADPH. Absorbance at 424 nm was recorded at 5-min intervals using a Spark microplate reader.

Whole-cell biocatalytic assay in *E. coli*

For the whole-cell biocatalytic assays described in the present study, the targeted P450 enzyme and its redox component were truncated by removing their N- or C-terminal transmembrane domains and fused together through a lambda linker GSTSSGSG, using the primers designed and listed in Supplementary Data Set 4. The constructs used for protein production were generated using the Gibson assembly (Gibson et al. 2009). Specifically, the *AtF5H1* and *SmF5H* fragments were PCR amplified with the forward primers that removed the sequence encoding their N-terminal transmembrane domain and with the reverse primers that incorporated the sequence encoding a lambda linker (Supplementary Data Set 5). The resulting PCR amplicons were ligated into the previously reported pET28a(+)-CB5D and pET28a(+)-ATR2 vectors (Zhao et al. 2023), respectively, via *BamHI* restriction enzyme digestion site using the Gibson assembly method, yielding the chimeric constructs *AtF5H1-CB5D*, *SmF5H-CB5D*, and *SmF5H-ATR2*, respectively. These constructs served as templates in subsequent PCR amplifications with the primers listed in Supplementary Data Set 4 to obtain the PCR amplicons *AtF5H1-lambda linker* and *SmF5H-lambda linker*; meanwhile, the *2muCB5D*, *SmCB5s*, *SmCB5-H5s*, *SmCPR*, and *MpCB5s* fragments were PCR amplified with primers bearing a short overlapping lambda linker sequence. After purification with Wizard SV Gel and PCR Clean-Up System (Promega), the amplified *F5H* and *CB5* or *CPR* fragments were mixed with the linearized pET28a(+) vector via *BamHI* and *XhoI* digestion sites and ligated together using the Gibson assembly method to obtain the chimeric constructs *AtF5H1-2muCB5D*, *AtF5H1-SmCB5-1*, *AtF5H1-SmCB5-2*, *AtF5H1-SmCB5-1-H5*, *AtF5H1-SmCB5-2-H5*, *AtF5H1-MpCB5-1*, *AtF5H1-MpCB5-2*, *SmF5H-SmCB5-1*, *SmF5H-SmCB5-2*, *SmF5H-SmCB5-1-H5*, *SmF5H-SmCB5-2-H5*, and *SmF5H-SmCPR*. For the production of SmF5H alone in *E. coli*, the *SmF5H* fragment was PCR amplified and inserted into the pET28a(+) vector via the *EcoRI* and *XhoI* sites. The *CrCB5*, *VcCB5*, and *CbCB5* genes from green algae, and the hornwort *AaCB5* gene was synthesized (GenScript) and ligated into the pET28a(+)-*AtF5H1-CB5D* vector by replacing *CB5D* fragment via the *EcoRI* and *XhoI* sites to obtain the chimeric constructs *AtF5H1-CrCB5*, *AtF5H1-VcCB5*, *AtF5H1-CbCB5*, *AtF5H1-AaCB5-1*, and *AtF5H-AaCB5-2*.

The resulting constructs were transformed into *E. coli* BL21 (DE3) cells. Positive colonies were used to inoculate small overnight cultures in Terrific Broth, from which 500 μL was used to inoculate 10 mL Terrific Broth for growth at 37°C until an optical density of ∼1 at 600 nm. The cells were then incubated at 15°C for 30 min, and protein production was induced with the addition of 1 mM IPTG (final concentration). The cells from the overnight cultures were pelleted by centrifugation at 3,000 × *g* for 5 min (Beckman Coulter Avanti J-15R with JS-4.750 rotor) and resuspended in 3 mL of 50 mM potassium phosphate buffer (pH 7.4), containing 2% (w/v) glucose, before the addition of 60 nmol coniferyl alcohol. The samples were incubated at 28°C under 250 rpm rotation for 4 h. The reactions were stopped by adding 1 mL of ethyl acetate containing 10 μM chrysin as the internal standard. The extracts were vacuum dried and analyzed with an Agilent 1100 HPLC system with the same settings as described above.

Statistical analysis

Statistical analysis in each required experiment was performed using either Student's *t* test in Microsoft Excel (2-tailed distribution and 2-sample unequal variance) or ANOVA test with GraphPad Prism version 4 ($P < 0.05$, 1-way ANOVA and Tukey's test). Details of statistical analyses, including sample sizes and biological replications, are provided in the figure legends. Detailed statistical analysis data are shown in Supplementary Data Set 6.

Accession numbers

DNA and derived protein sequence data from this article are available in the UCSC Genome Browser (https://genome.ucsc.edu/), Saccharomyces Genome Database (*S. cerevisiae*), TAIR database (https://www.arabidopsis.org/), GenBank (https://www.ncbi.nlm.nih.gov/genbank/), Phytozome 13 (https://phytozome-next.jgi.doe.gov/), PLAZA (https://bioinformatics.psb.ugent.be/plaza/), ORCAE (https://bioinformatics.psb.ugent.be/orcae/), Hornworts (https://www.hornworts.uzh.ch/en.html), ONEKP (https://db.cngb.org/onekp/), Solanaceae Genomics Network (https://solgenomics.net/), and FernBase (https://www.fernbase.org/). Their gene identifiers are given in Supplementary Data Set 1. The Arabidopsis genes are under the following TAIR accession numbers: *AtCB5A* (At1g26340); *AtCB5B* (At2g32720); *AtCB5C* (At2g46650); *AtCB5D* (At5g48810); *AtCB5E* (At5g53560); and *AtF5H1* (At4g36220).

Author contributions

C.-J.L. conceived the study. C.-J.L. and X.Z. designed the experiments. X.Z. conducted most of the experiments. Y.Z. initiated in planta complementation assays. Q.-y.Z. participated in phylogenetic analysis. C.-J.L. and X.Z. analyzed and interpreted the data and wrote the manuscript. All authors edited the manuscript.

Supplementary data

The following materials are available in the online version of this article.

Supplementary Figure S1. Simplified diagram of the phenylpropanoid–monolignol biosynthetic pathway leading to the biosynthesis of S-lignin and sinapoyl esters in angiosperms and in the lycophyte *Selaginella*.

Supplementary Figure S2. Phylogeny of CB5 homologs from representative plant taxa, constructed with protein sequences.

Supplementary Figure S3. Confirmation of the CB5 transgenic lines in the Arabidopsis *cb5d-1* background through genotyping PCR and RT-qPCR analysis.

Supplementary Figure S4. SDS–PAGE gel images showing the induced production of F5H-CB5 fusion proteins.

Supplementary Figure S5. UHPLC-MS analysis of *E. coli* whole-cell biocatalytic assays for AtF5H1 catalytic activity.

Supplementary Figure S6. Characterization of hornwort AaCB5s via whole-cell biocatalytic assays.

Supplementary Figure S7. Analysis of lignin composition in the representative fern and gymnosperm species and complementation assay of the *fah1-2* mutant with the selected *F5H* candidates.

Supplementary Figure S8. Verification of *SmCB5* transgenic lines in the *cb5d-1* background.

Supplementary Figure S9. Complementation test of the *fah1-2* mutant with *SmF5H*.

Supplementary Figure S10. Generation and confirmation of the *fah1-2 cb5d-2* double mutant.

Supplementary Figure S11. Gene expression in the putative *fah1-2 cb5d-2* complementation lines.

Supplementary Figure S12. Homology models of AtCB5D and SmCB5-1 and multiple protein sequence alignment of the genetically confirmed CB5s.

Supplementary Figure S13. Reduction assays of SmCB5s with *E. coli* total proteins.

Supplementary Figure S14. Amino acid sequence alignment of the H5 region in land plant CB5s with confirmed CB5D function and the 3 algal CB5s.

Supplementary Figure S15. Verification of transgenic lines harboring *SmCB5* variants in the *cb5d-1* background.

Supplementary Figure S16. Comparison of S-lignin biosynthetic F5H enzymes of Arabidopsis and *Selaginella*.

Supplementary Data Set 1. Genes used for the phylogenetic analysis shown in Fig. 1B.

Supplementary Data Set 2. Synthesized genes used in this study.

Supplementary Data Set 3. Sequences of Helix 5 in searched species used for calculating Helix 5 conservation in Fig. 8A.

Supplementary Data Set 4. Genetic material module.

Supplementary Data Set 5. Primers used in this study.

Supplementary Data Set 6. Summary of statistical analyses.

Supplementary File 1. Multiple DNA sequence alignment used for the tree shown in Fig. 1B.

Supplementary File 2. Newick format of the phylogenetic tree shown in Fig. 1B.

Supplementary File 3. Multiple protein sequence alignment for the tree shown in Supplementary Fig. S2.

Supplementary File 4. Newick format of the phylogenetic tree shown in Supplementary Fig. S2.

Funding

This work was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under contract number DE-SC0012704—specifically through the Physical Biosciences program of the Chemical Sciences, Geosciences and Biosciences Division (to C.-J.L.). This research used the confocal microscope at the Center for Functional Nanomaterials,which is a U.S. DOE Office of Science Facility, at Brookhaven National Laboratory under contract number DE-SC0012704.

Conflict of interest statement. None declared.

References

Alber AV, Renault H, Basilio-Lopes A, Bassard JE, Liu Z, Ullmann P, Lesot A, Bihel F, Schmitt M, Werck-Reichhart D, et al. Evolution of coumaroyl conjugate 3-hydroxylases in land plants: lignin biosynthesis and defense. *Plant J*. 2019:99(5):924–936. https://doi.org/10.1111/tpj.14373

Bak S, Beisson F, Bishop G, Hamberger B, Höfer R, Paquette S, Werck-Reichhart D. Cytochromes p450. *Arabidopsis Book*. 2011:9:e0144. https://doi.org/10.1199/tab.0144

Basile A, Giordano S, Lopez-Saez JA, Cobianchi RC. Antibacterial activity of pure flavonoids isolated from mosses. *Phytochemistry* 1999:52(8):1479–1482. https://doi.org/10.1016/s0031-9422(99)00286-1

Bhatt MR, Khatri Y, Rodgers RJ, Martin LL. Role of cytochrome b5 in the modulation of the enzymatic activities of cytochrome P450 17alpha-hydroxylase/17,20-lyase (P450 17A1). *J Steroid Biochem Mol Biol*. 2017:170:2–18. https://doi.org/10.1016/j.jsbmb.2016.02.033

Boerjan W, Ralph J, Baucher M. Lignin biosynthesis. *Annu Rev Plant Biol*. 2003:54(1):519–546. https://doi.org/10.1146/annurev.arplant.54.031902.134938

Buda GJ, Barnes WJ, Fich EA, Park S, Yeats TH, Zhao L, Domozych DS, Rose JK. An ATP binding cassette transporter is required for cuticular wax deposition and desiccation tolerance in the moss *Physcomitrella patens*. *Plant Cell*. 2013:25(10):4000–4013. https://doi.org/10.1105/tpc.113.117648

Chapple CCS, Vogt T, Ellis BE, Somerville CR. An Arabidopsis mutant defective in the general phenylpropanoid pathway. *Plant Cell*. 1992:4(11):1413–1424. https://doi.org/10.1105/tpc.4.11.1413

Clough SJ, Bent AF. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J*. 1998:16(6):735–743. https://doi.org/10.1046/j.1365-313x.1998.00343.x

Cowley AB, Sun N, Rivera M, Benson DR. Divergence in nonspecific hydrophobic packing interactions in the apo state, and its possible role in functional specialization of mitochondrial and microsomal cytochrome b5. *Biochemistry* 2005:44(44):14606–14615. https://doi.org/10.1021/bi051337m

de Vetten N, ter Horst J, van Schaik HP, de Boer A, Mol J, Koes R. A cytochrome b5 is required for full activity of flavonoid 3′,5′-hydroxylase, a cytochrome P450 involved in the formation of blue flower colors. *Proc Natl Acad Sci U S A*. 1999:96(2):778–783. https://doi.org/10.1073/pnas.96.2.778

Domozych D, Popper Z, Sorensen I. Charophytes: evolutionary giants and emerging model organisms. *Front Plant Sci*. 2016:7:1470. https://doi.org/10.3389/fpls.2016.01470

Espineira JM, Uzal EN, Ros LVG, Carrion JS, Merino F, Barcelo AR, Pomar F. Distribution of lignin monomers and the evolution of lignification among lower plants. *Plant Biol*. 2011:13(1):59–68. https://doi.org/10.1111/j.1438-8677.2010.00345.x

Faix O, Gyzas E, Schweers W. Comparative investigations on different fern lignins. *Holzforschung* 1977:31(5):137–144. https://doi.org/10.1515/hfsg.1977.31.5.137

Franke R, Hemm MR, Denault JW, Ruegger MO, Humphreys JM, Chapple C. Changes in secondary metabolism and deposition of an unusual lignin in the ref8 mutant of Arabidopsis. *Plant J*. 2002:30(1):47–59. https://doi.org/10.1046/j.1365-313x.2002.01267.x

Friedman WE, Cook ME. The origin and early evolution of tracheids in vascular plants: integration of palaeobotanical and

neobotanical data. *Philos Trans R Soc*. 2000:355(1398):857–868. https://doi.org/10.1098/rstb.2000.0620

Fukuchi-Mizutani M, Mizutani M, Tanaka Y, Kusumi T, Ohta D. Microsomal electron transfer in higher plants: cloning and heterologous expression of NADH-cytochrome b5 reductase from Arabidopsis. *Plant Physiol*. 1999:119(1):353–362. https://doi.org/10.1104/pp.119.1.353

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. 2009:6(5):343–345. https://doi.org/10.1038/nmeth.1318

Gou M, Yang X, Zhao Y, Ran X, Song Y, Liu CJ. Cytochrome b5 is an obligate electron shuttle protein for syringyl lignin biosynthesis in Arabidopsis. *Plant Cell*. 2019:31(6):1344–1366. https://doi.org/10.1105/tpc.18.00778

Hannemann F, Bichet A, Ewen KM, Bernhardt R. Cytochrome P450 systems–biological variations of electron transport chains. *Biochim Biophys Acta*. 2007:1770(3):330–344. https://doi.org/10.1016/j.bbagen.2006.07.017

Hatakeyama M, Kitaoka T, Ichinose H. Heterologous expression of fungal cytochromes P450 (CYP5136A1 and CYP5136A3) from the white-rot basidiomycete Phanerochaete chrysosporium: functionalization with cytochrome b5 in *Escherichia coli*. *Enzyme Microb Technol*. 2016:89:7–14. https://doi.org/10.1016/j.enzmictec.2016.03.004

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018:35(2):518–522. https://doi.org/10.1093/molbev/msx281

Holzinger A, Pichrtová M. Abiotic stress tolerance of charophyte green algae: new challenges for omics techniques. *Front Plant Sci*. 2016:7:678. https://doi.org/10.3389/fpls.2016.00678

Humphreys JM, Hemm MR, Chapple C. New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase. *Proc Natl Acad Sci U S A*. 1999:96(18):10045–10050. https://doi.org/10.1073/pnas.96.18.10045

Hwang YT, Pelitire SM, Henderson MP, Andrews DW, Dyer JM, Mullen RT. Novel targeting signals mediate the sorting of different isoforms of the tail-anchored membrane protein cytochrome b5 to either endoplasmic reticulum or mitochondria. *Plant Cell*. 2004:16(11):3002–3019. https://doi.org/10.1105/tpc.104.026039

Ichinose H, Michizoe J, Maruyama T, Kamiya N, Goto M. Electron-transfer reactions and functionalization of cytochrome P450cam monooxygenase system in reverse micelles. *Langmuir* 2004:20(13):5564–5568. https://doi.org/10.1021/la049752n

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021:596(7873):583–589. https://doi.org/10.1038/s41586-021-03819-2

Kandel SE, Lampe JN. Role of protein-protein interactions in cytochrome P450-mediated drug metabolism and toxicity. *Chem Res Toxicol*. 2014:27(9):1474–1486. https://doi.org/10.1021/tx500203s

Kawano M, Shirabe K, Nagai T, Takeshita M. Role of carboxyl residues surrounding heme of human cytochrome b5 in the electrostatic interaction with NADH-cytochrome b5 reductase. *Biochem Biophys Res Commun*. 1998:245(3):666–669. https://doi.org/10.1006/bbrc.1998.8502

Kumar R, Tran LS, Neelakandan AK, Nguyen HT. Higher plant cytochrome b5 polypeptides modulate fatty acid desaturation. *PLoS One* 2012:7(2):e31370. https://doi.org/10.1371/journal.pone.0031370

Kumar R, Wallis JG, Skidmore C, Browse J. A mutation in Arabidopsis cytochrome b5 reductase identified by high-throughput screening differentially affects hydroxylation and desaturation. *Plant J*. 2006:48(6):920–932. https://doi.org/10.1111/j.1365-313X.2006.02925.x

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018:35(6):1547–1549. https://doi.org/10.1093/molbev/msy096

Liu C-J. Cytochrome b5: a versatile electron carrier and regulator for plant metabolism. *Front Plant Sci*. 2022:13:984174. https://doi.org/10.3389/fpls.2022.984174

Maggio C, Barbante A, Ferro F, Frigerio L, Pedrazzini E. Intracellular sorting of the tail-anchored protein cytochrome b5 in plants: a comparative study using different isoforms from rabbit and Arabidopsis. *J Exp Bot*. 2007:58(6):1365–1379. https://doi.org/10.1093/jxb/erl303

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020:37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015

Napier JA, Smith MA, Stobart AK, Shewry PR. Isolation of a cDNA encoding a cytochrome b5 specifically expressed in developing tobacco seeds. *Planta* 1995:197(1):200–202. https://doi.org/10.1007/BF00239957

Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK, Haas FB, Vanderstraeten L, Becker D, Lang D, et al. The Chara genome: secondary complexity and implications for plant terrestrialization. *Cell* 2018:174(2):448–464.e24. https://doi.org/10.1016/j.cell.2018.06.033

Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, et al. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*2013:496(7446):528–532. https://doi.org/10.1038/nature12051

Parthasarathy S, Altuve A, Terzyan S, Zhang X, Kuczera K, Rivera M, Benson DR. Accommodating a nonconservative internal mutation by water-mediated hydrogen bonding between beta-sheet strands: a comparison of human and rat type B (mitochondrial) cytochrome b5. *Biochemistry* 2011:50(24):5544–5554. https://doi.org/10.1021/bi2004729

Porter TD. The roles of cytochrome b5 in cytochrome P450 reactions. *J Biochem Mol Toxicol*. 2002:16(6):311–316. https://doi.org/10.1002/jbt.10052

Qiao X, Zhang S, Paterson AH. Pervasive genome duplications across the plant tree of life and their links to major evolutionary innovations and transitions. *Comput Struct Biotechnol J*. 2022:20:3248–3256. https://doi.org/10.1016/j.csbj.2022.06.026

Raven JA. Physiological correlates of the morphology of early vascular plants. *Bot J Linn Soc*. 1984:88(1–2):105–126. https://doi.org/10.1111/j.1095-8339.1984.tb01566.x

Renault H, Alber A, Horst NA, Basilio Lopes A, Fich EA, Kriegshauser L, Wiedemann G, Ullmann P, Herrgott L, Erhardt M, et al. A phenol-enriched cuticle is ancestral to lignin evolution in land plants. *Nat Commun*. 2017:8(1):14713. https://doi.org/10.1038/ncomms14713

Renault H, Werck-Reichhart D, Weng JK. Harnessing lignin evolution for biotechnological applications. *Curr Opin Biotechnol*. 2019:56:105–111. https://doi.org/10.1016/j.copbio.2018.10.011

Ruegger M, Chapple C. Mutations that reduce sinapoylmalate accumulation in *Arabidopsis thaliana* define loci with diverse roles in phenylpropanoid metabolism. *Genetics* 2001:159(4):1741–1749. https://doi.org/10.1093/genetics/159.4.1741

Schenkman JB, Jansson I. The many roles of cytochrome b5. *Pharmacol Ther*. 2003:97(2):139–152. https://doi.org/10.1016/s0163-7258(02)00327-3

Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc*. 2008:3(6):1101–1108. https://doi.org/10.1038/nprot.2008.73

Schuckel J, Rylott EL, Grogan G, Bruce NC. A gene-fusion approach to enabling plant cytochromes P450 for biocatalysis. *Chembiochem*. 2012:13(18):2758–2763. https://doi.org/10.1002/cbic.201200572

Sergeev GV, Gilep AA, Estabrook RW, Usanov SA. Expression of outer mitochondrial membrane cytochrome b5 in *Escherichia coli*. purification of the recombinant protein and studies of its interaction with electron-transfer partners. *Biochemistry (Mosc)*. 2006:71(7):790–799. https://doi.org/10.1134/s0006297906070121

Smith MA, Jonsson L, Stymne S, Stobart K. Evidence for cytochrome b5 as an electron donor in ricinoleic acid biosynthesis in microsomal preparations from developing castor bean (*Ricinus communis* L.). *Biochem J*. 1992:287(Pt 1):141–144. https://doi.org/10.1042/bj2870141

Towers GH, Gibbs RD. Lignin chemistry and the taxonomy of higher plants. *Nature* 1953:172(4366):25–26. https://doi.org/10.1038/172025a0

Umezawa T. Diversity in lignan biosynthesis. *Phytochem Rev*. 2003:2(3):371–390. https://doi.org/10.1023/B:PHYT.0000045487.02836.32

Uzal EN, Ros LVG, Pomar F, Bernal MA, Paradela A, Albar JP, Barcelo AR. The presence of sinapyl lignin in *Ginkgo biloba* cell cultures changes our views of the evolution of lignin biosynthesis. *Physiol Plant*. 2009:135(2):196–213. https://doi.org/10.1111/j.1399-3054.2008.01185.x

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. Lignin biosynthesis and structure. *Plant Physiol*. 2010:153(3):895–905. https://doi.org/10.1104/pp.110.155119

Weng JK, Akiyama T, Bonawitz ND, Li X, Ralph J, Chapple C. Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte Selaginella and flowering plants. *Plant Cell*. 2010:22(4):1033–1045. https://doi.org/10.1105/tpc.109.073528

Weng JK, Banks JA, Chapple C. Parallels in lignin biosynthesis: a study in *Selaginella moellendorffii* reveals convergence across 400 million years of evolution. *Commun Integr Biol*. 2008a:1(1):20–22. https://doi.org/10.4161/cib.1.1.6466

Weng JK, Chapple C. The origin and evolution of lignin biosynthesis. *New Phytol*. 2010:187(2):273–285. https://doi.org/10.1111/j.1469-8137.2010.03327.x

Weng JK, Li X, Stout J, Chapple C. Independent origins of syringyl lignin in vascular plants. *Proc Natl Acad Sci USA*. 2008b:105(22):7887–7892. https://doi.org/10.1073/pnas.0801696105

Werck-Reichhart D, Feyereisen R. Cytochromes P450: a success story. *Genome Biol*. 2000:1(6):Reviews3003. https://doi.org/10.1186/gb-2000-1-6-reviews3003

White E, Towers GHN. Comparative biochemistry of the lycopods. *Phytochemistry* 1967:6(5):663–667. https://doi.org/10.1016/S0031-9422(00)86005-7

Xie J, Chen Y, Cai G, Cai R, Hu Z, Wang H. Tree visualization by one table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. *Nucleic Acids Res*. 2023:51(W1):W587–W592. https://doi.org/10.1093/nar/gkad359

Zhao X, Zhao Y, Gou M, Liu CJ. Tissue-preferential recruitment of electron transfer chains for cytochrome P450-catalyzed phenolic biosynthesis. *Sci Adv*. 2023:9(2):eade4389. https://doi.org/10.1126/sciadv.ade4389