PCP
PLANT & CELL PHYSIOLOGY

How Embryophytic is the Biosynthesis of Phenylpropanoids and their Derivatives in Streptophyte Algae?

Jan de Vries[1,5,*], Sophie de Vries[1,2,5], Claudio H. Slamovits[1,3], Laura E. Rose[2,4] and John M. Archibald[1,3,*]

[1]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada
[2]Population Genetics, Heinrich-Heine University Düsseldorf, D-40225 Düsseldorf, Germany
[3]Canadian Institute for Advanced Research, Program in Integrated Microbial Biodiversity, Toronto, ON, Canada
[4]CEPLAS – Cluster of Excellence in Plant Sciences, Heinrich-Heine University Düsseldorf, D-40225 Düsseldorf, Germany
[5]These authors contributed equally to this work.
*Corresponding authors: Jan de Vries, E-mail, jan.devries@dal.ca; John M. Archibald, E-mail, john.archibald@dal.ca, Fax, 1-902-494-1355.
(Received December 27, 2016; Accepted March 08, 2017)

Regular Paper

The origin of land plants from algae is a long-standing question in evolutionary biology. It is becoming increasingly clear that many characters that were once assumed to be 'embryophyte specific' can in fact be found in their closest algal relatives, the streptophyte algae. One such case is the phenylpropanoid pathway. While biochemical data indicate that streptophyte algae harbor lignin-like components, the phenylpropanoid core pathway, which serves as the backbone of lignin biosynthesis, has been proposed to have arisen at the base of the land plants. Here we revisit this hypothesis using a wealth of new sequence data from streptophyte algae. Tracing the biochemical pathway towards lignin biogenesis, we show that most of the genes required for phenylpropanoid synthesis and the precursors for lignin production were already present in streptophyte algae. Nevertheless, phylogenetic analyses and protein structure predictions of one of the key enzyme classes in lignin production, cinnamyl alcohol dehydrogenase (CAD), suggest that CADs of streptophyte algae are more similar to sinapyl alcohol dehydrogenases (SADs). This suggests that the end-products of the pathway leading to lignin biosynthesis in streptophyte algae may facilitate the production of lignin-like compounds and defense molecules. We hypothesize that streptophyte algae already possessed the genetic toolkit from which the capacity to produce lignin later evolved in vascular plants.

Keywords: Charophytes • Lignin • Phenylpropanoids • Plant evolution • Plant–microbe interaction • Secondary metabolism.

Abbreviations: AAE, acyl-activating enzyme; ACoS, acyl-CoA synthetase; CAD, cinnamyl alcohol dehydrogenase; CCR, cinnamoyl-CoA reductase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; CYP450, cytochrome P450; DFR, dihydroflavonol 4-reductase; DIR, dirigent; EGT, endosymbiotic gene transfer; F5H, ferulate 5-hydroxylase; HAL, histidine ammonia lyase; I-TASSER, Iterative Threading ASSEmbly Refinement; KEGG, Kyoto Encyclopedia of Genes and Genomes; LGT, lateral gene transfer; MAST, motif alignment and search tool; MEME, multiple Em for motif elicitation; ML, maximum likelihood; PAL, phenylalanine ammonia lyase; PP, phenylpropanoid; SAD, sinapyl alcohol dehydrogenase.

Introduction

Land plants evolved from within the higher streptophyte algae (Lewis and McCourt 2004, Wickett et al. 2014, Delwiche and Cooper 2015, de Vries et al. 2016). Molecular characterization of streptophyte algae has accelerated over the last few years thanks to transcriptome sequencing projects (Ju et al. 2015, Van de Poel et al. 2016) and a genome project for *Klebsormidium flaccidum* (Hori et al. 2014)—and more are underway (Rensing 2017). It is becoming more and more clear that embryophytic features such as 'communication' with soil microbes (Delaux et al. 2015), complex phytohormone signaling (Hori et al. 2014, Holzinger and Becker 2015, Ju et al. 2015), phragmoplast formation during cell division (Fowke and Pickett-Heaps 1969, Buschmann and Zachgo 2016), tighter plastid–nucleus communication and plastid properties (de Vries et al. 2016, de Vries et al. 2017a), and intricate stress responses (Holzinger et al. 2014, Holzinger and Pichrtová 2016) were in fact already present in streptophyte algae. These features now underpin the concept that streptophyte algae were 'pre-adapted' to the terrestrial environment (see Becker and Marin 2009).

One of the remarkable features of land plant biology is their production of a broad range of secondary metabolites. A major source of plant secondary metabolites is the phenylpropanoid (PP) pathway (Vogt 2010). The PP pathway generates precursors for land plant-associated compounds such as lignin (the 'wood molecule'), cinnamaldehyde (a major constituent of the flavor of cinnamon; Ito 2008) and flavonoids, which include antioxidants such as quercetin (Boots et al. 2008). Hence, these molecules display a wide functional diversity. In the case of substances such as suberin and lignin, they provide physical stability to plant tissues (see Whetten and Sederoff 1995). Many PPs (including, but not limited to, flavonoids and stilbenes) are also known for their important role in stress response, especially to phytopathogens (Woodward and Pearce 1988, Dixon et al. 2002, Hammerbacher et al. 2011, Oliva et al. 2015).

PP biosynthesis is generally thought to have originated at the base of land plants (Emiliani et al. 2009). Evolving the ability to synthesize these metabolites has been considered a key adaptation of land plants, helping them to achieve their present-day complexity (Weng 2013). The evolution of the PP core pathway

Editor-in-Chief's Choice

Plant Cell Physiol. 58(5): 934–945 (2017) doi:10.1093/pcp/pcx037, Advance Access publication on 13 March 2017, available FREE online at www.pcp.oxfordjournals.org

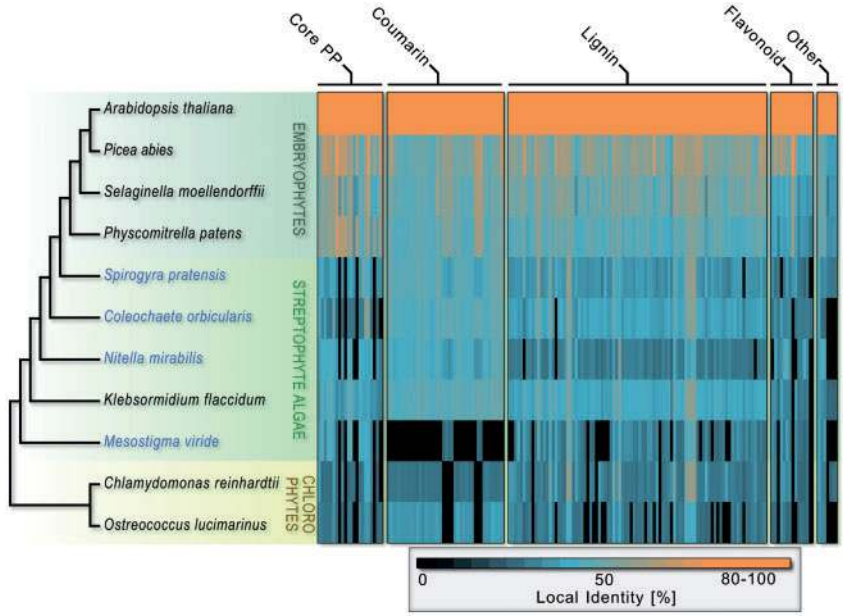

Fig. 1 Similarity matrix of *Arabidopsis thaliana* phenylpropanoid (PP) biosynthesis enzymes across Chloroplastida. A protein sequence data set was assembled from the genomes (black font) of four land plants (embryophytes), the streptophyte alga *Klebsormidium flaccidum* and two chlorophytes, as well as from four RNaseq data sets (blue font) of streptophyte algae. As queries for our BLASTp/tBLASTn survey, we used sequences of well-characterized *Arabidopsis* enzymes, including 22 PP core, 40 coumarin, 89 lignin and 14 flavonoid biosynthesis enzymes. Detected homologs are depicted by ticks colored based on their local identity (blue, low, to orange, high; relative to the respective *A. thaliana* sequences). Black ticks indicate no detection of a homolog (e-value cutoff $<10^{-7}$).

was proposed to have benefited from lateral gene transfer of a key enzyme from bacteria to the base of the land plants (Emiliani et al. 2009). Yet these studies were performed in the absence of genome-scale data for streptophyte algae. Here we present a survey of available genome and transcriptome data from streptophyte algae with regard to genes coding for key enzymes of the PP core, lignin, coumarin and flavonoid biosynthesis pathways. Our data show that an ortholog of the cornerstone enzyme phenylalanine ammonia lyase (PAL) is present in the basal branching streptophyte alga *K. flaccidum* and that streptophyte algae have the 'genetic toolkit' for the biosynthesis of PPs and their derivatives.

Results and Discussion

Enzymes for the biosynthesis of phenylpropanoids and their derivatives across Chloroplastida

The PP biosynthetic pathway is the backbone for multifaceted secondary metabolites. It can generate many clade-specific secondary metabolites such as sinapine, which is a signature metabolite of Brassicaceae (Milkowski and Strack 2010). It can also lead to capsaicin in pepper fruits (Díaz et al. 2004). Nevertheless, underlying this complexity is a common gene suite. Here, we sought to determine the extent to which this and its downstream pathways are present in two chlorophytes, five streptophyte algae and four land plants. Based on Vanholme et al. (2012), we categorized the pathway into four major groups: the PP core pathway, coumarin biosynthesis, lignin biosynthesis and flavonoid biosynthesis (**Supplementary Table S1**). We began by extracting all 22 PP core pathway enzymes, as well as 40 coumarin, 89 lignin and 14 flavonoid biosynthesis pathway enzymes from the *Arabidopsis thaliana* KEGG (Kyoto Encyclopedia of Genes and Genomes) entry and using them as BLAST queries against our data set. Six enzymes listed in the KEGG pathway but not considered by Vanholme et al. (2012) were assigned to the category 'other'. For the 22 PP core pathway enzymes, we identified a variety of homologs within the streptophyte algae, ranging from 16 in *Mesostigma viride* to 22 in *K. flaccidum*. The homologs from streptophyte algae show an average of 37.2% protein sequence identity (i.e. in the local alignment) relative to the *A. thaliana* sequences, while land plants show a higher average sequence identity (56.0%) and chlorophytes show a lower sequence identity (29.0%) compared with *A. thaliana* (**Fig. 1**). For the coumarin biosynthesis proteins, we identified homologs for all 40 *A. thaliana* sequences in all streptophyte algae, except *M. viride* (six out of 40). Similar to the PP core pathway, we found an increase in sequence identity from chlorophytes to streptophyte algae to land plants for the coumarin biosynthesis proteins (31.1% in chlorophytes $<$ 43.7% in streptophyte algae $<$ 50.3% in land plants; **Fig. 1**). Of the 89 lignin biosynthesis proteins in *A. thaliana*, streptophyte algae possess homologs for 74 (*M. viride*) to 89 (*K. flaccidum*) proteins. Even though less pronounced, the sequence identities of the lignin biosynthesis sequences increased in the same manner as the aforementioned PP core and coumarin biosynthesis enzymes (**Fig. 1**): 31.1% in chlorophytes $<$ 36.1% in streptophyte algae $<$ 51.2% in land plants. For the flavonoid pathway, the detected number of homologs ranged from 10 in *Nitella mirabilis* to 13 in

Coleochaete orbicularis (**Fig. 1**). However, in contrast to the other three major pathways, this pathway showed a similar degree of sequence identity in chlorophytes and streptophyte algae compared with *A. thaliana* (28.3% and 31.1%, respectively), while land plants showed the greatest identity (44.2%). These data suggest a distinct progression of PP, coumarin and lignin biosynthesis protein identity in streptophyte algae, setting them apart from the pattern observed for chlorophytes. However, a high degree of sequence redundancy is likely for these proteins and it is therefore necessary to identify the true orthologs with a more rigorous screening.

Orthologous enzymes for the biosynthesis of phenylpropanoids and their derivatives across streptophyte diversity

A number of algae are known to produce PPs and PP-derived metabolites (Goiris et al. 2014). To address the question of to what degree the PP pathway enzymes of embryophytes have orthologs in streptophyte algae, we used a reciprocal best BLAST hit to screen our assembled data set for genes and transcripts coding for proteins orthologous to those PP biosynthesis enzymes from *A. thaliana* (**Fig. 2A; Supplementary Table S2**). Of 157 *Arabidopsis* proteins (i.e. 22 PP core, 40 coumarin, 89 lignin and six others), we found between four (*M. viride*) and 11 (*N. mirabilis*) proteins fulfilling this reciprocal best BLAST hit criterion in the streptophyte algae. An even smaller number of putative orthologs was detected in chlorophytes: two in *Ostreococcus lucimarinus* and four in *Chlamydomonas reinhardtii*. Among land plants, we found only 28 orthologs in *Physcomitrella patens* and 39 in *Picea abies*. These small numbers of putative orthologs may reflect the gene family expansion of many PP core, lignin and coumarin biosynthesis genes in *Arabidopsis* (Hamberger et al. 2007; see Xu et al. 2009).

We next attempted to infer the routing of streptophyte algal PP biosynthesis (**Fig. 2B**). Based on the orthologs detected, reactions are present which connect phenylalanine (the point of entry into the PP pathway) with the metabolite *p*-coumaroyl-CoA, which links the PP core pathway to the flavonoid biosynthesis pathway (Vogt 2010; **Supplementary Fig. S1**). The only enzyme missing from the thread is cinnamate 4-hydroxylase (C4H). However, it has been noted previously that many photosynthetic eukaryotes [both Archaeplastida and meta-algae (i.e. a polyphyletic group encompassing eukaryotes that gained plastids through secondary/tertiary endosymbiosis)] apparently lack the C4H enzyme (Labeeuw et al. 2015). Yet, Goiris et al. (2014) found high levels of *p*-coumaric acid across a broad range of Archaeplastida and meta-algae, including *Tetraselmis*, which belongs to a basal branching group of chlorophytes (Leliaert et al. 2012). This suggests a non-canonical, thus far unknown, mechanism for the production of *p*-coumaric acid in these algae. We therefore assume that the means by which *Tetraselmis* produces *p*-coumaric acid might also be used by streptophyte algae. From *p*-coumaric acid, the streptophyte algal 4-coumarate:CoA ligase (4CL) is predicted to catalyze the synthesis of *p*-coumaryl-CoA (see Gross and Zenk 1974), a metabolite that can be funneled into flavonoid, H-lignin or further along into G- and S-lignin biosynthesis (Vogt 2010). Some of the orthologs predicted to play a role in generating flavonoids, and most of the orthologs for H-lignin and G-lignin were detected among streptophyte alga in our screen (**Fig. 2**; **Supplementary Fig. S1**).

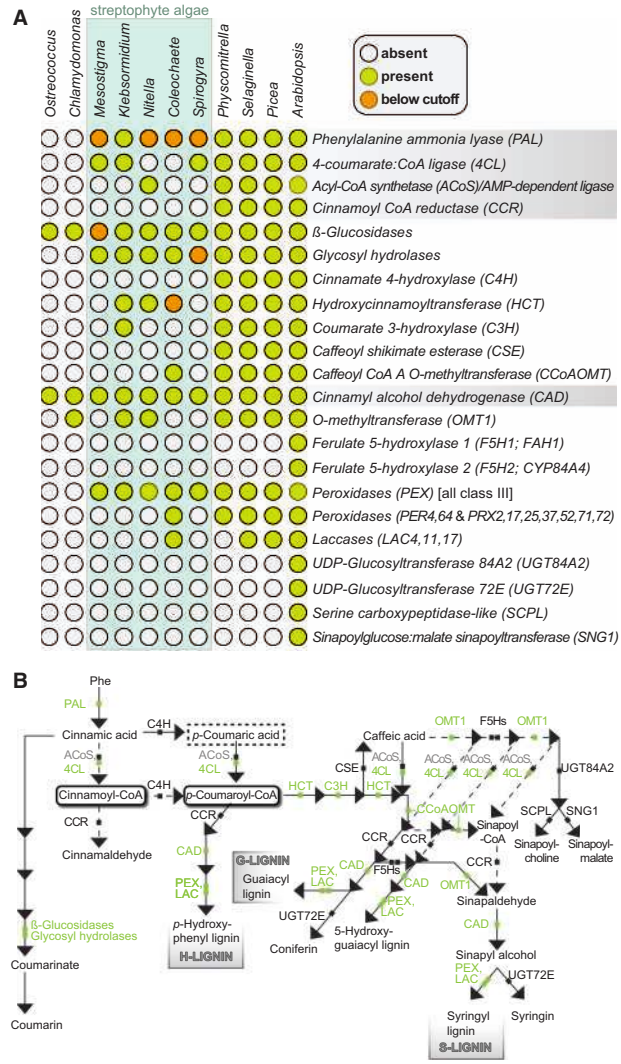

Fig. 2 Orthologs of core phenylpropanoid (PP), coumarin and lignin biosynthesis enzymes in streptophyte algae. (A) Orthologs were detected based on a reciprocal BLASTp and BLASTx search querying the detected homologs in **Fig. 1** against the *A. thaliana* genome protein data (e-value cut-off $\leq 10^{-7}$). Gene families have been reduced to one functional category. Gray boxes highlight those gene families for which in-depth analyses have been performed. (B) A simplified diagram of the PP, coumarin and lignin biochemical pathway based on KEGG. The dashed arrows indicate steps that are only suggestive or are not included in the core PP and lignin pathway (see Vanholme et al. 2012, Vanholme et al. 2013). Green font and flags (on the arrows) highlight those enzymes present in streptophyte algae; black font and flags label enzymes not detected in streptophyte algae. Gray font labels indicate that the function of this enzyme in the PP and lignin pathway could not be demonstrated experimentally (de Azevedo Souza et al. 2009). The dotted box around *p*-coumaric acid refers to its UHPLC–MS/MS-based detection among many Archaeplastida in the absence of detectable C4H homologs (see Goiris et al. 2014, Labeeuw et al. 2015). Cinnamoyl-CoA and *p*-coumaroyl-CoA are precursors of the flavonoid pathway.

For the production of S-lignin, the required ferulate 5-hydroxylase (F5H; see Meyer et al. 1998) was not detected. One needs to bear in mind that our approach is based on sequence identity and is very strict; it has been shown that the lycophyte *Selaginella moellendorffii* possesses another F5H gene, which arose independently from the F5H gene in angiosperms—although both belong to the cytochrome P450 (CYP P450) protein family (Weng et al. 2008). As a result of convergent evolution, the *Sm*F5H shows little sequence identity with the angiosperm F5H counterparts (Weng et al. 2008) and is functionally distinct (Weng et al. 2010. Weng et al. 2011). Because *Sm*F5H and *At*F5H are not orthologous, we were unable to detect this gene (i.e. *Sm*F5H) in our analyses. The evolution of F5H function is characterized by convergent evolution of the CYP P450 protein family and an absence of F5H in many lineages (see Weng et al. 2010). Nonetheless, this does not exclude the possibility that orthologs of *Sm*F5H could be present in lineages missing an *At*F5H ortholog. To test if the streptophyte algae possess a protein orthologous to *Sm*F5H, we conducted an additional reciprocal BLAST search using the *Sm*F5H sequence published by Weng et al. (2008) (Protein-IDs: ABV80343.1; 166299); we detected only for *P. patens* (Protein-ID: Pp3c17_14010V3.1.p), but not for any other organism, a protein orthologous to *Sm*F5H. Due to the aforementioned convergence and versatility of CYP P450 protein evolution, it is conceivable that streptophyte algae might also have (convergently) evolved another CYP P450 protein catalyzing a reaction that allows the plants to produce S-lignin.

Evolutionary history of phenylpropanoid biosynthesis in plants and algae

Labeeuw et al. (2015) found homologs of the enzymes catalyzing the reactions leading to the biosynthesis of monolignol in many archaeplastidians. However, biosynthesis of G- and S-lignin seemed limited to land plants. Here we show that orthologs of these enzymes are in fact present in streptophyte algae. The same restricted distribution was thought to be the case for the key enzyme acting upon phenylalanine at the very beginning of the PP pathway: only land plants were thought to encode PAL (see Labeeuw et al. 2015). Emiliani et al. (2009) in fact proposed that the PAL gene was transferred from soil bacteria to the base of land plants by lateral gene transfer (LGT). However, we identified a clear ortholog of PAL in the streptophyte alga *K. flaccidum*, which we call *Kf*PAL (**Fig. 2**). Reciprocal best BLAST hits to PAL were obtained for the four other streptophyte algae for which transcriptome data are available, although these genes are very fragmented, suggesting low expression/low coverage, and must be considered with caution. We therefore focused our efforts on *Kf*PAL and what it might reveal about the evolutionary history of PAL in the green lineage.

*Kf*PAL is 445 amino acids in length and is significantly shorter than the PAL proteins of Arabidopsis (e.g. *At*PAL1 with 725 amino acids) and the 758 amino acid *Physcomitrella* PAL (Pp3c14_11870V3.1.p) (**Fig. 3A**). Nevertheless, the main structural domain of all of these proteins is the Lyase 'class I' domain, predicted to span 428 amino acids in *Kf*PAL (486 amino acids in *At*PAL1; **Fig. 3A**). One of the best predictions of the function of PAL is an F residue followed by an L residue in its active site (position 144/145 in *At*PAL1; Watts et al. 2006). When the first of these aforementioned residues is an F, the substrate is phenylalanine and the product is cinnamic acid.

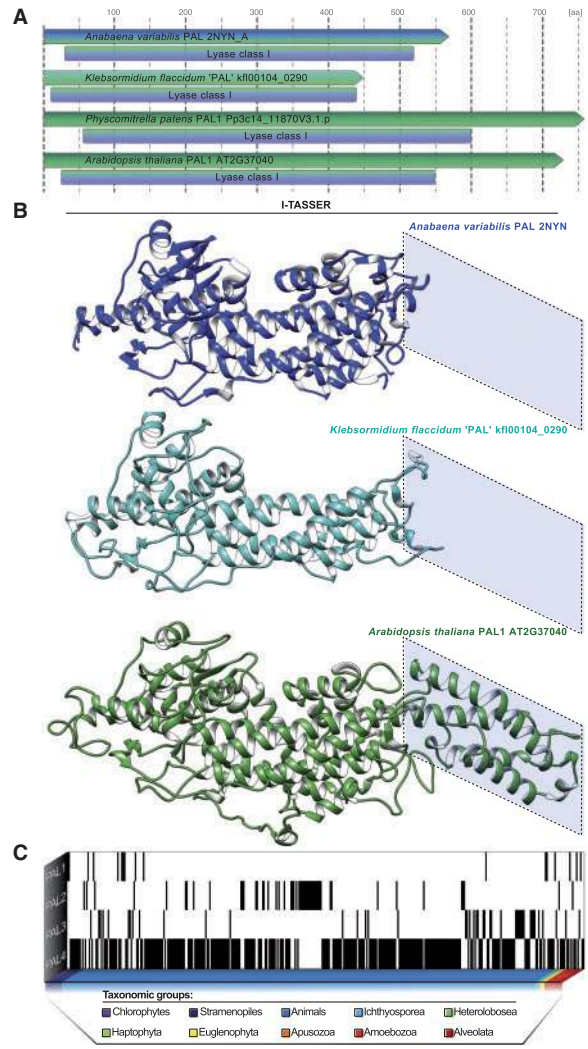

Fig. 3 *Klebsormidium* and the evolution of the phenylalanine ammonia lyases (PALs). (A) PALs are drawn to scale based on their length in amino acids (aa). Lyase class I domain prediction is based on NCBI CD searches. (B) *Kf*PAL resembles cyanobacterial PALs. Best protein structure model predictions for *Kf*PAL and *At*PAL1 were computed by I-TASSER; the structure of *Av*PAL (chain A) was downloaded from the Protein Data Bank (PDB). The blue box highlights the extra residues/domains characteristic of *bona fide* eukaryotic PALs, which are organized in the quaternary structure in a fan-like manner. (C) Orthologs of eukaryotic PALs can be found outside of fungi and land plants. *At*PAL1 was used as a query in a BLASTp search against all eukaryotic sequences in RefSeq, excluding land plants and fungi, returning 381 hits. All 381 retrieved hits (from various groups of eukaryotes, sorted, color labeled and shown below the matrix) were used as a query in a reciprocal BLASTp search against the *A. thaliana* genome peptide release, returning in all 381 cases one of the four PAL proteins encoded by the *A. thaliana* genome. The black ticks in the matrix denote which of the four *At*PAL paralogs was returned as the best reciprocal hit.

However, when this position corresponds to H, tyrosine is the substrate and *p*-coumaric acid is the product (Watts et al. 2006). To evaluate the putative functionality of *Kf*PAL, we (i) investigated the corresponding region in *Kf*PAL (position 80/81) and (ii) attempted to gain further insight into its functionality by using Iterative Threading ASSEmbly Refinement (I-TASSER; Zhang 2008) to model its tertiary structure. At its corresponding active site (position 80/81), *Kf*PAL has neither an F nor H residue, but rather a G residue followed by L, as in *At*PAL1. This variation of the active site residues of *Kf*PAL is inconclusive in terms of what substrate *Kf*PAL will convert. *Kf*PAL's best structure model was found to be closest to the PAL of the Gram-negative bacterium *Pantoea agglomerans* (PDB accession: 3UNV). The most confidently predicted ligand-binding sites for *Kf*PAL were for *S*-β-phenylalanine (C-score 0.94) and *trans*-cinnamic acid (C-score 0.74), representing the usual substrate and product of PAL, respectively. Hence, even though its active site does not show the canonical residues, the I-TASSER analysis suggests that *Kf*PAL has its canonical function using phenylalanine as a substrate. However, the predicted *Kf*PAL structure resembles the histidine ammonia lyases (HALs; see Schwede et al. 1999). One of the features thought to distinguish HAL and PAL is the presence of extra residues in land plant PALs (**Fig. 3B**, blue box), which are arranged in a fan-like manner in the quaternary structure of eukaryotic PALs (Calabrese et al. 2004). Moffit et al. (2007) found the absence of this domain to be a characteristic of cyanobacterial PALs and proposed that the structure of the cyanobacterial PAL represents an evolutionary intermediate from which the eukaryotic PAL arose. Given that *Kf*PAL also lacks this domain, it seems reasonable to conclude that the *K. flaccidum* ortholog has retained the ancestral (i.e. cyanobacterial) state. Streptophyte algae thus provide an opportunity to study the evolutionary trajectory of plant PALs.

What does the presence of a PAL ortholog in streptophytes mean for the bacterium to land plant LGT proposed by Emiliani et al. (2009)? Using *Kf*PAL as a query, the first BLASTp hit against the non-redundant GenBank CDS translations + PDB + SwissProt + PIR + PRF bacterial sequences (NCBI Resource Coordinators 2016) is to HALs of soil-dwelling methylobacteria (WP_012330172.1), while the second hit is to the cyanobacterium *Anabaena variabilis* PAL (Q3M5Z3.1). *Anabaena variabilis* belongs to the section IV cyanobacteria, whose proteins show high similarity to plastid endosymbiosis-derived nuclear genes (Dagan et al. 2013). Since phenylalanine is synthesized in the plastid (Rippert et al. 2009) and PAL is present in cyanobacteria (Moffit et al. 2007), it is conceivable that eukaryotic PAL was inherited from the cyanobacterial endosymbiont via endosymbiotic gene transfer (EGT; see Timmis et al. 2004), not from bacteria by LGT. Yet, fungi have the eukaryotic PAL (Emiliani et al. 2009), making the scenario of cyanobacterial/plastid EGT less likely. To explore this further, we generated a maximum likelihood (ML) phylogeny of 228 PALs and HALs from a wide range of Bacteria and Eukarya (**Supplementary Figs. S2, S3**). Similar to Emiliani et al. (2009), we retrieved a well-supported clade with soil bacteria, cyanobacteria, fungi and land plants; *Kf*PAL clustered with this soil bacterial/cyanobacterial clade and not with the land plants. This further underscores the strong resemblance to bacterial PALs deduced from the *Kf*PAL structure. In contrast to Emiliani et al. (2009), the clade harboring fungi and land plant PALs also included animals (such as the recently sequenced brachiopod *Lingula anatina*; Luo et al. 2015). The results suggest that the evolutionary history of PAL is indeed not easily explained by an LGT event at the base of land plants, but is rather complex. This was further supported when we searched the eukaryotic RefSeq database (excluding fungi and land plants) via BLASTp using *At*PAL1 as query. We detected 381 sequences from all eukaryotic supergroups (see Adl et al. 2012) except Excavata. Furthermore, when these 381 sequences are used as BLASTp queries against the *A. thaliana* genome, all of them return one of the four *At*PALs as best hit (**Fig. 3C**). Hence all eukaryotic PALs and HALs might share a common ancestry. From this common ancestor onwards, sequence divergence and convergence and even possible replacements of nuclear PAL/HAL by organellar versions via EGT may have shaped the evolutionary history of PAL; we therefore think it likely that *Kf*PAL's affinity for cyanobacterial sequences is indicative of such an EGT-driven replacement. Regardless of the implications for LGT, our data show that the streptophyte alga *Klebsormidium* possesses a PAL, the key entry point into the PP pathway.

A phylogenetic framework for lignin biosynthesis in streptophyte algae

Lignin is a major component of (vascular) land plants. It adds strength to their cell walls, supporting upward growth (Cowles et al. 1984). Naturally, streptophyte algae have been investigated as potential models for early plant cell wall evolution (Sørensen et al. 2011, Domozych et al. 2014, Harholt et al. 2016) and found to have lignins or lignin-like compounds (Delwiche et al. 1989, Kroken et al. 1996, Ligrone et al. 2008, Sørensen et al. 2011); these cell wall-reinforcing metabolites are considered to be involved in streptophyte algal stress responses (Holzinger and Pichrtová 2016). As described above, synthesis of these compounds starts with PAL (Cowles et al. 1984). The product of PAL is cinnamic acid, which is converted to *p*-coumaric acid by C4H. Yet, the presence of C4H is not a requirement for *p*-coumaric acid biosynthesis in numerous algae (Goiris et al. 2014), which might apply to streptophyte algae, too. The conversion of *p*-coumaric acid to *p*-coumaroyl-CoA is catalyzed by 4CL (for an overview, see Shockey et al. 2003). Within the streptophyte algae, we found potential orthologs for at least one representative of the 4CL enzyme class, with the exception of the transcriptome of *Coleochaete* and *Nitella* (**Fig. 2A**).

While most streptophyte sequences had their best hit to one of the *At*4CLs, a sequence obtained for *N. mirabilis* retrieved the *A. thaliana* AMP-dependent ligase AT4G05160, a closely related enzyme, as a reciprocal hit. In an attempt to resolve this ambiguity, we inferred a phylogeny on 4CL, acyl-CoA synthetase (ACoS), other AMP-dependent ligases (AT4G05160 and AT4G19010) in the pathway and acyl-activating enzymes (AAEs), which are related AMP-dependent ligases not involved in PP biosynthesis (see Shockey et al. 2003, Gao et al. 2015; **Fig. 4A**). All streptophyte algal sequences clustered within the

group of 4CLs, ACoS and the other AMP-dependent ligases (bootstrap value = 99) and with a slightly weaker support more closely to 4CLs and ACoS (bootstrap value = 65). It was shown that *At*ACoS5 and the other two 4CL-like AMP-dependent ligases have only poor affinity for the usual 4CL substrates (Costa et al. 2005). I-TASSER predictions for the 4CL ortholog of *K. flaccidum* and the *N. mirabilis* enzyme show that the closest structural analog is the *Nicotiana tabacum* 4CL (PDB ID 5BSR; Li and Nair 2015; TM-scores 0.967 and 0.891).

We further analyzed the active site and CoA-binding pocket of the 4CL sequences based on the data on *N. tabacum* 4CL2 in its thioester conformation (Li and Nair 2015; **Supplementary File S1**). The active site (H237, T336, D420, K437, K441, Q446; positions based on *Nt*4CL2) is conserved across all sequences shown in the alignment (which includes all sequences shown in the phylogeny except for the AAEs), including all streptophyte algae. The CoA-binding pocket, however, shows distinct differences between canonical 4CLs, ACoS and those sequences that we classified as other AMP-dependent ligases (all positions based on the *N. tabacum* sequence): while residues K443 and F445 are conserved in all sequences, F506 is only conserved among canonical 4CLs and substituted by P or H in all other sequences, including those of streptophyte algae. Residues L233 and F261 group the streptophyte algae sequences with the canonical 4CLs, while ACoS sequences in our alignment have I or M at position 233 and Y at position 261 (with the sole exception being *Pa*ACoS). Residue K260 is conserved between the streptophyte algal sequences of *Mesostigma* and *Klebsormidium* and the canonical 4CLs, while *Nitella* possesses an R. These data indicate that while similar to ACoS in some respects, streptophyte algal sequences show a distinct progression in sequence similarity towards canonical 4CLs, supporting their potential 4CL-like function in the PP pathway (**Supplementary Fig. S4**).

Labeeuw et al. (2015) reported no 4CLs for *C. reinhardtii* and *O. lucimarinus* but found a potential 4CL in *Chlorella variabilis*.

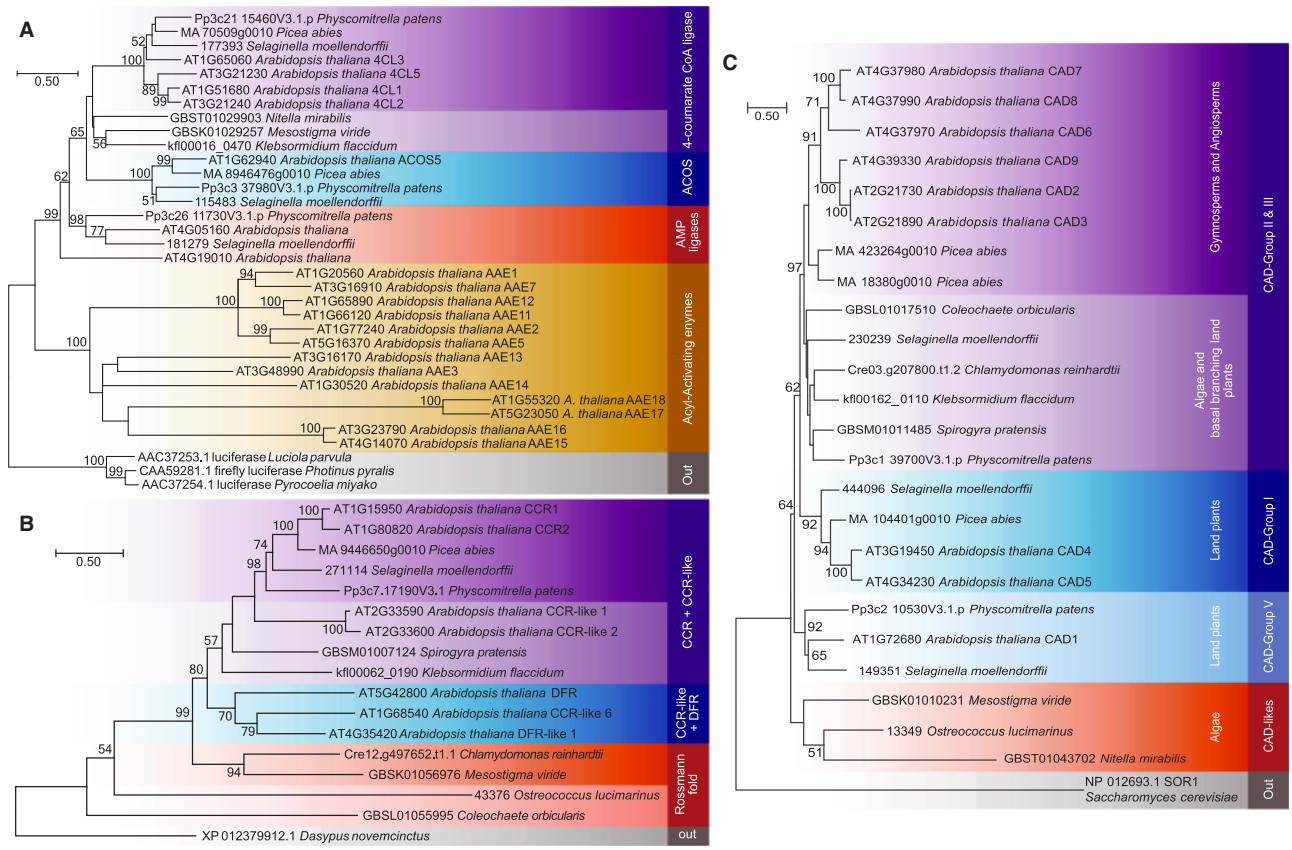

Fig. 4 Phylogenies of three key enzymes of the core phenylpropanoid (PP) and lignin biosynthesis pathways. All phylogenies were computed based on a maximum likelihood (ML) LG + G + I method (bootstrap values <50 are not shown in the figure) and rooted with their respective outgroups ('out'); RNaseq-derived protein sequences too short for building a proper alignment were excluded from the phylogenetic analyses. (A) Phylogeny of the core PP pathway AMP ligases catalyzing the conversion of hydroxycinnamic acids to their respective CoA thioester. All ingroup sequences were retrieved via a reciprocal best BLAST hit with the *Arabidopsis* sequences as an initial query. Only *Arabidopsis* sequences given in the KEGG database for this reaction were used. (B) Phylogeny of CCR, CCR-like and DFR sequences identifies CCR and CCR-like streptophyte algal sequences. Land plant sequences were recovered via a reciprocal best BLAST hit based on the *Arabidopsis* CCR1 and CCR2 sequence. All algal sequences were recovered via a one-way best BLAST hit with *At*CCR1 and *At*CCR2 as query sequences. (C) Phylogeny of CAD sequences based on reciprocal best BLAST hits using the *Arabidopsis* CAD sequences as query. Only *Arabidopsis* CAD sequences given by the KEGG database were used. CAD groups were defined based on Saballos et al. (2009). Outgroups were chosen (i) based on Labeeuw et al. (2015), (ii) because there are no plant sequences and (iii) because of their sequence similarities to the respective enzyme families.

We therefore performed a BLASTp using *At*4CL1 as a query against *C. variabilis* that returned the protein with the identifier 'CHLNCDRAFT_48296'. Performing a BLASTp using 'CHLNCDRAFT_48296' as a query against *A. thaliana* retrieved *At*AAE3 as a best hit. *At*AAE3 clusters outside of the AMP-dependent ligases involved in PP biosynthesis (**Fig. 4A**). This suggests that chlorophytes may only possess some AAEs, which are closely related to the PP pathway AMP-dependent ligases, but are functionally distinct. Hence, our data support the presence of 4CL among streptophyte, but not chlorophyte algae.

The next important enzyme *en route* to lignin—next to caffeoyl-CoA *O*-methyltransferase (CCoAOMT)—is cinnamoyl-CoA reductase (CCR). We did not detect a reciprocal hit for CCR in our BLAST analyses of our algal data sets (**Fig. 2A**). However, CCRs share high sequence similarity with dihydroflavonol 4-reductases (DFRs) (Lacombe et al. 1997), which could have influenced the BLAST results. We therefore constructed an ML phylogeny with all *A. thaliana* CCRs, CCR-like sequences and DFRs [as indicated by The Arabidopsis Information Resource (TAIR)] to see where the streptophyte algal homologs, i.e. sequences with a unidirectional hit to a CCR sequence in the BLAST analyses, cluster (**Fig. 4B**). All Chloroplastida sequences, with the exception of *O. lucimarinus* and *C. orbicularis*, were clustered with CCR, CCR-like and DFR sequences (**Fig. 4B**). This is because all Chloroplastida sequences possess a very large stretch containing the Rossmann fold and therefore, like CCR and DFR, belong to the Rossman-fold superfamily.

Key characteristics of CCRs in angiosperms are a conserved N-terminal NAD(P)-binding domain [VTG(A/G)(A/G/S)G(F/Y)(I/V/L)(G/A)] and a catalytic motif (NWYCYGK) (Lacombe et al. 1997). We used MEME (multiple Em for motif elicitation) and MAST (motif alignment and search tool) (Bailey et al. 2009) motif searches to identify common motifs between land plant and streptophyte algae CCRs. We found 12 different amino acid sequence motifs in both of the CCR sequences of *A. thaliana*. Both searches yielded similar results with respect to the presence of these motifs; MAST predicted one additional motif for *M. viride* and *C. orbicularis*, compared with the MEME prediction. In agreement with the phylogeny (**Fig. 4B**), *Spirogyra* and *Klebsormidium* were found to possess most of the land plant motifs (11/12; **Supplementary Fig. S5**). The NAD(P)-binding site-containing motif was present in all algae, while the catalytic site-containing motif was only present in *Spirogyra* and *Klebsormidium*. A closer look at the NAD(P)-binding and catalytic site in the alignment revealed a high degree of conservation of the specific NAD(P)-binding site sequences, but a degeneration of the catalytic site, correlating with phylogenetic distance (**Fig. 5**). I-TASSER protein model prediction retrieved the CCR of *Petunia hydrida* (PDB ID: 4R1S; Pan et al. 2014; TM-score: 0.8) as the closest structure to the *Kf*CCR (kfl00062_0190) (**Supplementary Fig. S6**). The structurally closest protein of CCR from *O. lucimarinus* (protein ID 43376) was the human UDP-glucuronic acid decarboxylase (PDB ID: 2B69; TM-score 0.923); not a single CCR was among the top 10 closest structural

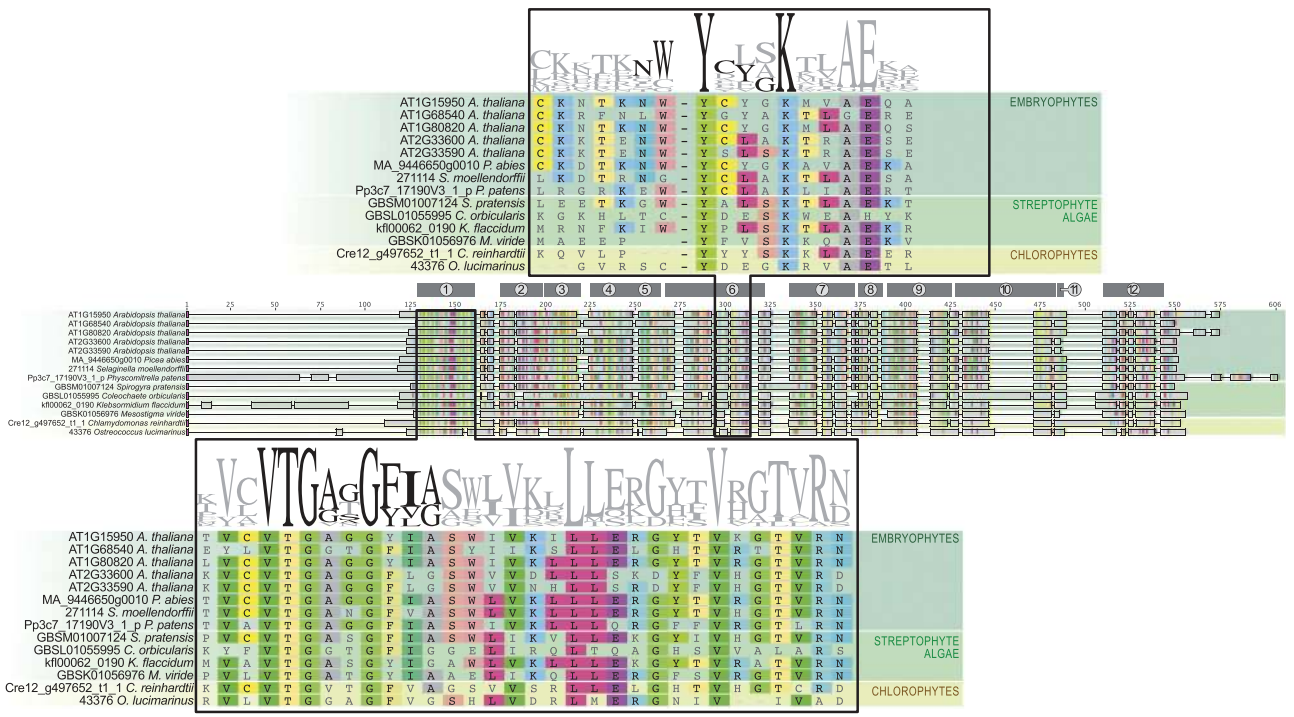

Fig. 5 Conserved motifs in algal and land plant CCRs. L-INS-I alignment of *A. thaliana* CCR and CCR-like sequences with CCR homologs of other land plants (dark green), streptophyte (light green) and chlorophyte (yellow green) algae. Enlargement of the alignment of the entire NAD(P)-binding site-containing motif 1 (bottom) and that part of motif 6 that contains the active site (top) recovered from MEME are shown. The sequence LOGO indicates the relative occurrence of a certain amino acid at a given position in the motifs. The conserved CCR NAD(P)-binding (bottom) and catalytic (top) site are colored in black in the sequence LOGO. All 12 motifs detected by MEME are depicted above the alignment overview and are numbered. The sequence logos for motifs 2–12 can be found in **Supplementary Fig. S5**.

analogs. The I-TASSER protein model of the putative *Chlamydomonas* CCR (Cre12.g497652.t1.1) was structurally closest to CAD2 of *M. truncatula* (PDB ID: 4QTZ; Pan et al. 2014; TM-score: 0.855), which was shown by others to be highly similar to CCR from *P. hybrida* (Pan et al. 2014). Consistent with this, the second best structural match was the CCR of *P. hybrida* (PDB ID: 4R1SA; TM-score 0.842). These results are consistent with the existence of a functional CCR in some streptophyte algae and—despite its phylogenetic position—probably also in the chlorophyte, *C. reinhardtii*.

Cinnamyl alcohol dehydrogenase (CAD) is a key enzyme in lignin biosynthesis. The CAD enzyme family is involved in the synthesis of all three lignin components; its conversion of hydroxycinnamyl aldehydes to the corresponding alcohols is one of the major rate determinants of lignin production (Gross et al. 1973, Mansell et al. 1974). Saballos et al. (2009) divided land plant CADs into five major groups. Groups I and V included all land plants, group II was specific to *A. thaliana*, while groups III and IV were posited to be angiosperm and monocot specific, respectively. These CAD enzymes might be partially functionally redundant but may also have different tissue and functional specificities (Raes et al. 2003). Consistent with Labeeuw et al. (2015), we detected potential CAD orthologs in chlorophytes (**Fig. 2A**). *Arabidopsis thaliana* sequences belonging to group II and III clustered together (**Fig. 4C**). Most of the algal sequences (including *Chlamydomonas*) clustered with group II/III, the moss *P. patens* and the lycophyte *S. moellendorffii*. Groups I and V do indeed appear to be land plant specific (**Fig. 4C**). Three algal sequences (*Mesostigma*, *Ostreococcus* and *Nitella*) remained outside of the CAD clusters, and are thus designated 'CAD-like' sequences. To test whether (i) the algal sequences clustering with CAD group II/III are structurally more similar to CAD-II or CAD-III and (ii) the CAD-like sequences structurally resemble CADs, we performed a structure prediction on one representative of each group. Both the *O. lucimarinus* CAD-like (*Ol*CAD-like) and the *K. flaccidum* CAD (*Kf*CAD) protein models were structurally closest to the sinapyl alcohol dehydrogenase (SAD) of *Populus tremuloides* (*Pt*SAD; Bomati and Noel 2005; **Supplementary Fig. S7**). Based on phylogenetic analyses, the closest homologs of *Pt*SAD in *A. thaliana* are the group II CADs (Guo et al. 2010), suggesting that streptophyte algal CADs may have the less limited functional abilities found in SADs compared with *bona fide* CADs (group I; Youn et al. 2006, Barakate et al. 2011): *Nicotiana* SAD was shown not to be involved in S-lignin synthesis, but rather to play a role in the production of plant defense metabolites (Barakate et al. 2011), e.g. the various lignans derived from the PP pathway (Suzuki and Umezawa 2007, Saleem et al. 2010).

Dirigent (DIR) proteins are an important enzyme family in lignan biosynthesis (Burlat et al. 2001). To analyze the possibility of lignan biosynthesis in streptophyte algae, we queried the transcriptomes and genomes of the streptophyte algae with all sequences of proteins indicated as DIR proteins by TAIR and found that the *K. flaccidum* genome encodes one DIR protein (kfl00751_0020, orthologous to AT3G58090). The absence of DIR proteins from transcriptomes of the other streptophyte algae was, however, not surprising, as many DIR proteins are up-regulated specifically during defense responses (Ralph et al. 2007, Oliva et al. 2015). Similarly, the corresponding CADs (group II) show relaxed constraints in their sequence evolution, increased functional variability and up-regulation upon pathogen elicitor treatments, and were suggested to act in defense responses (Youn et al. 2006, Logemann et al. 1997, Guo et al. 2010). Both *Kf*CAD and *Ol*CAD-like are predicted to have NAD/NADH-binding sites (C-scores 0.92 and 0.89) and CAD activity (EC 1.1.1.195; C-scores^{EC} 0.704 and 0.677). These results suggest that although they do not cluster together in the phylogeny, both CAD and CAD-like sequences harbor a potential CAD function; the products may be lignin-like and/or defense compounds, whose production might involve DIR proteins. This is substantiated by the fact that lignin-like components are present in streptophyte algae (see Sørensen et al. 2011).

Lignification is facilitated by class III peroxidases (Marjamaa et al. 2009, Fagerstedt et al. 2010, Shigeto and Tsutsumi 2016). Here we found that all streptophyte algae possess at least one class III peroxidase (**Fig. 2**). However, so far only nine of these class III peroxidases (*At*PRX2, *At*PER4, *At*PRX17, *At*PRX25, *At*PRX37, *At*PRX52, *At*PER64, *At*PRX71 and *At*PRX72) were identified as important in lignification in *A. thaliana* (Pedreira et al. 2011, Herrero et al. 2013, Lee et al. 2013, Shigeto et al. 2013, Fernández-Pérez et al. 2015a, Fernández-Pérez et al. 2015b, Fernández-Pérez et al. 2015c, Shigeto and Tsutsumi 2016, Cosio et al. 2017). We therefore surveyed whether streptophyte algae possess orthologs of these specific peroxidases, and found only an ortholog of the PRX52-encoding gene in *Coleochaete*, but no other streptophyte algae (**Fig. 2A**). *Co*PRX52 is slightly longer than *At*PRX52; however, both are predicted to have a secretory peroxidase domain of similar length and spanning almost the entire protein (**Fig. 6A**). We further checked the

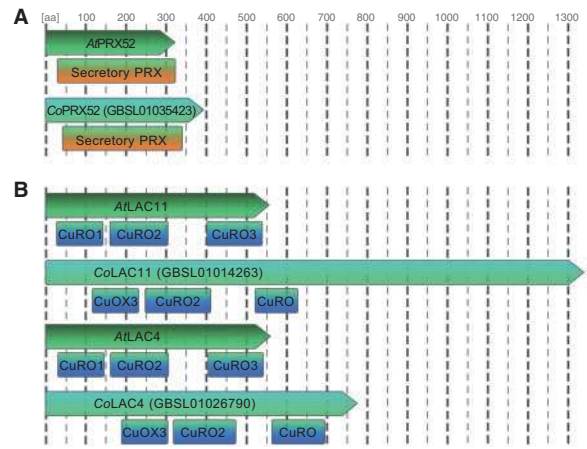

Fig. 6 Lignification-associated peroxidases and laccases of *Coleochaete orbicularis* and *Arabidopsis thaliana*. In our data set, the only orthologs of lignification-associated peroxidases (A) and laccases (B) detected in streptophyte algae were found in *C. orbicularis*. The *C. orbicularis* and *A. thaliana* orthologs are depicted to scale (in amino acids), along with their putative domain structures predicted using CD search (Marchel-Bauer et al. 2015). Domains: Secretory PRX, secretory plant peroxidase; CuRO1,2,3, cupredoxin domain 1, 2 and 3; CuRO, cupredoxin superfamily; CuOX, Cu-oxidase.

six active site residues reported by Welinder et al. (2002) for class III peroxidases including *At*PRX52 (R66, H70, N98, P167, H197 and D269, positions according to the *At*PRX52 sequence). We found that all residues, except N98 (i.e. five out of six), were conserved in *Co*PRX52. This suggests that *Co*PRX52 could aid lignification. It is noteworthy that Harholt et al. (2016) mentioned that *Klebsormidium* possesses a lignin peroxidase-resembling protein, highlighting that there are probably more lignin-related enzymes to be discovered in streptophyte algae sequence data that are currently being generated (see Rensing 2017).

We further surveyed the putative polymerization of monolignols into lignins by looking at the laccase repertoire, using *At*LAC4, *At*LAC11 and *At*LAC17 (Berthet et al. 2011, Zhao et al. 2013). Among streptophyte algae, our bi-directional BLAST analyses only detected laccase orthologs in the data set of *C. orbicularis*. Using CD (conserved domain) search, we found that the putative *Co*LAC4 and *Co*LAC11 probably have three cupredoxin domains, resembling their *A. thaliana* orthologs (**Fig. 6B**). The presence of these laccases and PRX52 in *Coleochaete* warrants attention, as it has long been known to be particularly desiccation tolerant (Graham et al. 2012) and ranks among the best-known reports of lignin in streptophyte algae (Delwiche et al. 1989).

Conclusion

The PP pathway is the source of a variety of signature metabolites of land plants. Even using a very stringent approach, our analyses suggest that the basal branching streptophyte alga *Klebsormidium* has the genes not only for the production of lignin-like precursors for cell wall reinforcement, but also for secondary metabolites that are known for their importance in plant defense (Dixon et al. 2002). The latter warrants attention. Beneficial microbiota have long been considered to have been a factor in streptophyte terrestrialization (Field et al. 2015; see also Knack et al. 2015). In turn, however, early land plants were probably challenged by a whole set of new pathogens. While streptophyte algal immunity has not been studied, our data suggest that it could include PP-derived metabolites. The degree to which all of this genetic potential translates into the biosynthesis of specific metabolites awaits metabolomic investigation.

Materials and Methods

Identification of orthologs associated with the biosynthesis of phenylpropanoids and their derivatives

We assembled a protein sequence data set from the genomes of two chlorophytes [*Ostreococcus lucimarinus* v2.0 (Palenik et al. 2007) and *Chlamydomonas reinhardtii* v5.5 (Merchant et al 2007)] the streptophyte alga *Klebsormidium flaccidum* v1.0 (Hori et al. 2014) and four land plants [*Physcomitrella patens* v3.3 (Rensing et al. 2008), *Selaginella moellendorffii* v1.0 (Banks et al. 2009), *Picea abies* v1.0 (Nystedt et al. 2013) and *Arabidopsis thaliana* v10 (Arabidopsis Genome Initiative 2000)]. The data set was further complemented with transcriptome data from four streptophyte algae [*Mesostigma viride*, *Nitella mirabilis*, *Coleochaete orbicularis* and *Spirogyra pratensis* (Ju et al. 2015)]. We extracted 22 PP core, 40 coumarin, 89 lignin, six 'other' and 14 flavonoid biosynthesis pathway enzymes from the *A. thaliana* KEGG (cf. Kanehisa et al. 2016; see Kanehisa et al. 2017) entry and used them as queries for BLAST analyses against our data set. Additionally, we searched for (i) *At*LAC4 (AT2G38080), *At*LAC11 (AT5G03260) and *At*LAC17 (AT5G60020)—as they were characterized by Zhao et al. (2013) as crucial for lignin polymerization—and (ii) 25 Arabidopsis DIR proteins, which were extracted from TAIR. For more in-depth analyses, a reciprocal best BLAST hit analysis was performed. Protein structure predictions were performed using I-TASSER. Motif searches were performed using a MEME/MAST search. Settings were as described in de Vries et al. (2017b).

Phylogenetic analyses

Phylogenies were created using (i) only the best hits of a reciprocal BLAST search in our data sets (4CL and CAD); (ii) the best BLAST hit of a one-way BLAST of our data sets (CCR); and (iii) the best reciprocal BLAST hit combined with homologous sequences from the NCBI (HAL/PAL). Fungal PAL sequences were chosen based on Emiliani et al. (2009). Plant PAL sequences were retrieved from genome data based on their annotations. The remaining PAL and HAL sequences were retrieved via a BLASTp search against the refseq databases of α-, β-, ε- and γ-Proteobacteria, Firmicutes, Stramenopiles, Alveolata, Cnidaria, Protistoma, Insecta and Mammalia. All selected sequences had $\geq$60% query coverage, an e-value of $\leq 10^{-5}$ and an identity of $\geq$30%. Nucleotide sequences were converted into amino acid sequences using the translate function of the Sequence Manipulation Suite (Stothard 2000). Non-plant outgroup sequences were chosen largely based on Labeeuw et al. (2015). Sequences were aligned using either L-INS-I (4CL, CCR and PAL) or G-INS-I (CAD) with default options in mafft v7.305b (Katoh et al. 2002, Katoh and Standley 2013). ML phylogenies were created using MEGA 7 (Kumar et al. 2016) using the LG-model + G (category = 5) + I, partial deletion cut-off of 95% for sites that contained gaps and 100 (PAL) or 500 (4CL, CCR and CAD) bootstrap replicates.

Supplementary data

Supplementary data are available at PCP online.

Funding

This work was supported by the German Research Foundation (DFG) [Research Fellowship to J.dV. (VR 132/1-1) and grant Ro 2491/6-1 to L.E.R. and S.dV.] and Natural Sciences and Engineering Research Council of Canada [Discovery grant to J.M.A.]. J.M.A. is a Senior Fellow of the Canadian Institute for Advanced Research.

Acknowledgments

We would like to thank two anonymous reviewers for their excellent comments.

Disclosures

The authors have no conflicts of interest to declare.

Note added in proof

After acceptance of this manuscript, Renault et al. (2017) published a study on the basal branching land plant *Physcomitrella patens* showing that *Pp*C3H (CYP98A3; for which we detected

an ortholog in the streptophyte alga *Klebsormidium flaccidum*) directs PP derivatives into another pathway that acts in cuticle biosynthesis. While lignin-like components were detected in streptophyte alga (Delwiche et al. 1989, Kroken et al. 1996, Ligrone et al. 2008, Sørensen et al. 2011), the presence of cuticle-like surface materials has only been suggested (Cook and Graham 1998, Graham et al. 2012). Whether or not C3H is involved in the biosynthesis of cuticle(-like) surface materials in streptophyte algae, too, remains to be investigated. If so, it has important implications for streptophyte algal tolerance of drought/desiccation – a major terrestrial stressor – and hence streptophyte terrestrialization.

References

Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., et al. (2012) The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* 59: 429–514.

Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., et al. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37: W202–W208.

Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., dePamphilis, C., et al. (2011) The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332: 960–963.

Barakate, A., Stephens, J., Goldie, A., Hunter, W.N., Marshall, D., Hancock, R.D., et al. (2011) Syringyl lignin is unaltered by severe sinapyl alcohol dehydrogenase suppression in tobacco. *Plant Cell* 23: 4492–4506.

Becker, B. and Marin, B. (2009) Streptophyte algae and the origin of embryophytes. *Ann. Bot.* 103: 999–1004.

Berthet, S., Demont-Caulet, N., Pollet, B., Bidzinski, P., Cézard, L., Le Bris, P., et al. (2011) Disruption of *LACCASE4* and *17* results in tissue-specific alterations to lignification of *Arabidopsis thaliana* stems. *Plant Cell* 23: 1124–1137.

Bomati, E.K. and Noel, J.P. (2005) Structural and kinetic basis for substrate selectivity in *Populus tremuloides* sinapyl alcohol dehydrogenase. *Plant Cell* 17: 1598–1611.

Boots, A.W., Haenen, G.R.M.M. and Bast, A. (2008) Health effects of quercetin: from antioxidant to nutraceutical. *Eur. J. Pharmacol.* 585: 325–337.

Burlat, V., Kwon, M., Davin, L.B. and Lewis, N.G. (2001) Dirigent proteins and dirigent sites in lignifying tissues. *Phytochemistry* 57: 883–897.

Buschmann, H. and Zachgo, S. (2016) The evolution of cell division: from streptophyte algae to land plants. *Trends Plant Sci.* 21: 872–883.

Calabrese, J.C., Jordan, D.B., Boodhoo, A., Sariaslani, S. and Vannelli, T. (2004) Crystal structure of phenylalanine ammonia lyase: multiple helix dipoles implicated in catalysis. *Biochemistry* 43: 11403–11416.

Cook, M.E. and Graham, L.E. (1998) Structural similarities between surface layers of selected charophycean algae and bryophytes and the cuticles of vascular plants. *Int. J. Plant Sci.* 159: 780–787.

Cosio, C., Ranocha, P., Francoz, E., Burlat V., Zheng, Y., Perry, S.E., et al. (2017) The class III peroxidase PRX17 is a direct target of the MADS-box transcription factor AGAMOUS-LIKE15 (AGL15) and participates in lignified tissue formation. *New Phytol.* 213: 250–263.

Costa, M.A., Bedgar, D.L., Moinuddin, S.G.A., Kim, K.-W., Cardenas, C.L., Cochrane, F.C., et al. (2005) Characterization *in vitro* and *in vivo* of the putative multigene 4-coumarate:CoA ligase network in *Arabidopsis*: syringyl lignin and sinapate/sinapyl alcohol derivative formation. *Phytochemistry* 66: 2072–2091.

Cowles, J.R., Scheld, H.W., Lemay, R. and Peterson, C. (1984) Growth and lignification in seedlings exposed to eight days of microgravity. *Ann. Bot.* 3: 33–48.

Dagan, T., Roettger, M., Stucken, K., Landan, G., Koch, R., Major, P., et al. (2013) Genomes of stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. *Genome Biol. Evol.* 5: 31–44.

de Azevedo Souza, C., Kim, S.S., Koch, S., Kienow, L., Schneider, K., McKim, S.M., et al. (2009) A novel fatty acyl-CoA synthetase is required for pollen development and sporopollenin biosynthesis in *Arabidopsis*. *Plant Cell* 21: 507–525.

Delaux, P.-M., Radhakrishnan, G.V., Jayaraman, D., Cheemab, J., Malbreild, M., Volkening, J.D., et al. (2015) Algal ancestor of land plants was preadapted for symbiosis. *Proc. Natl. Acad. Sci. USA* 112: 13390–13395.

Delwiche, C.F. and Cooper, E.D. (2015) The evolutionary origin of a terrestrial flora. *Curr. Biol.* 25: R899–R910.

Delwiche, C.F., Graham, L.E. and Thomson, N. (1989) Lignin-like compounds and sporopollenin in *Coleochaete*, an algal model for land plant ancestry. *Science* 245: 399–401.

de Vries, J., Archibald, J.M. and Gould, S.B. (2017a) The carboxy terminus of YCF1 contains a motif conserved throughout >500 Myr of streptophyte evolution. *Genome Biol. Evol.* 9: 473–479.

de Vries, J., Stanton, A., Archibald, J.M. and Gould, S.B. (2016) Streptophyte terrestrialization in light of plastid evolution. *Trends Plant Sci.* 21: 467–476.

de Vries, S., von Dahlen, J.K., Uhlmann, C., Schnake, A., Kloesges, T. and Rose, L.E. (2017b) Signatures of selection and host-adapted gene expression of the *Phytophthora infestans* RNA silencing suppressor PSR2. *Mol. Plant Pathol.* 18: 110–124.

Díaz, J., Pomar, F., Bernal, Á. and Merino, F. (2004) Peroxidases and the metabolism of capsaicin in *Capsicum annuum* L. *Phytochem. Rev.* 3: 141–157.

Dixon, R.A., Achnine, L., Kota, P., Liu, C.-J., Reddy, M.S.S. and Wang, L. (2002) The phenylpropanoid pathway and plant defence—a genomics perspective. *Mol. Plant Pathol.* 3: 371–390.Domozych, D.S., Sorensen, I., Popper, Z.A., Ochs, J., Andreas, A., Fangel, J.U., et al. (2014) Pectin metabolism and assembly in the cell wall of the charophyte green alga *Penium margaritaceum*. *Plant Physiol.* 165: 105–118.

Emiliani, G., Fondi, M., Fani, R. and Gribaldo, S. (2009) A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. *Biol. Direct* 4: 7.

Fagerstedt, K.V., Kukkola, E.M., Koistinen, V.V.T., Takahashi, J. and Marjamaa, K. (2010) Cell wall lignin is polymerised by class III secretable plant peroxidases in Norway spruce. *J. Integr. Plant Biol.* 52: 186–194.

Fernández-Pérez, F., Pomar, F., Pedreño, M.A. and Novo-Uzal, E. (2015a) The suppression of *AtPrx52* affects fibers but not xylem lignification in *Arabidopsis* by altering the proportion of syringyl units. *Physiol. Plant.* 154: 395–406.

Fernández-Pérez F., Pomar, F., Pedreño, M.A. and Novo-Uzal, E. (2015b) Suppression of *Arabidopsis* peroxidase 72 alters cell wall and phenylpropanoid metabolism. *Plant Sci.* 239: 192–199.

Fernández-Pérez F., Vivar, T., Pomar, F., Pedreño, M.A. and Novo-Uzal, E. (2015c) Peroxidase 4 is involved in syringyl lignin formation in *Arabidopsis thaliana*. *J. Plant Physiol.* 175: 86–94.

Field, K.J., Pressel, S., Duckett, J.G., Rimington, W.R. and Bidartondo, M.I. (2015) Symbiotic options for the conquest of land. *Trends Ecol. Evol.* 30: 477–486.

Fowke, L.C. and Pickett-Heaps, J.D. (1969) Cell division in *Spirogyra*. II. Cytokinesis. *J. Phycol.* 5: 273–281.

Gao, S., Yu, H.-N., Xu, R.-X., Chen, A.-X. and Lou, H.-X. (2015) Cloning and functional characterization of a 4-coumarate CoA ligase from liverwort *Plagiochasma appendiculatum*. *Phytochemistry* 111: 48–58.

Goiris, K., Muylaert, K., Voorspoels, S., Noten, B., De Paepe, D., Baart, G.J.E., et al. (2014) Detection of flavonoids in microalgae from different evolutionary lineages. *J. Phycol.* 50: 483–492.

Graham, L.E., Arancibia-Avila, P., Taylor, W.A., Strother, P.K. and Cook, M.E. (2012) Aeroterrestrial *Coleochaete* (Streptophyta, Coleochaetales) models early plant adaptation to land. *Amer. J. Bot.* 99: 130–144.

Gross, G.G., Stöckigt, J., Mansell, R.L. and Zenk, M.H. (1973) Three novel enzymes involved in the reduction of ferulic acid to coniferyl alcohol in higher plants: ferulate:CoA ligase, feruloyl-CoA reductase and coniferyl alcohol oxidoreductase. *FEBS Lett.* 31: 283–286.

Gross, G. and Zenk, M. (1974) Isolation and properties of hydroxycinnamate:CoA ligase from lignifying tissue of *Forsythia*. *Eur. J. Biochem.* 42: 453–459.

Guo, D.-M., Ran, J.-H. and Wang, X.-Q. (2010) Evolution of the cinnamyl/sinapyl alcohol dehydrogenase (CAD/SAD) gene family: the emergence of real lignin is associated with the origin of *bona fide CAD*. *J. Mol. Evol.* 71: 202–218.

Hamberger, B., Ellis, M., Friedmann, M., de Azevedo Souza, C., Barbazuk, B. and Douglas, C.J. (2007) Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the *Populus* lignin toolbox and conservation and diversification of angiosperm gene families. *Can. J. Bot.* 85: 1182–1201.

Hammerbacher, A., Ralph, S.G., Bohlmann, J., Fenning, T.M., Gershenzon, J. and Schmidt, A. (2011) Biosynthesis of the major tetrahydroxystilbenes in spruce, astringin and isorhapontin, proceeds via resveratrol and is enhanced by fungal infection. *Plant Physiol.* 157: 876–890.

Harholt, J., Moestrup, Ø. and Ulvskov, P. (2016) Why plants were terrestrial from the beginning. *Trends Plant Sci.* 21: 96–101.

Herrero, J., Fernández-Pérez, F., Yebra, T., Novo-Uzal, E., Pomar, F., Pedreño, M.Á., et al. (2013) Bioinformatic and functional characterization of the basic peroxidase 72 from *Arabidopsis thaliana* involved in lignin biosynthesis. *Planta* 237: 1599–1612.

Holzinger, A. and Becker, B. (2015) Desiccation tolerance in the streptophyte green alga *Klebsormidium*: the role of phytohormones. *Commun. Integr. Biol.* 8: e1059978.

Holzinger, A., Kaplan, F., Blaas, K., Zechmann, B., Komsic-Buchmann, K. and Becker, B. (2014) Transcriptomics of desiccation tolerance in the streptophyte green alga *Klebsormidium* reveal a land plant-like defense reaction. *PLoS One* 9: e110630.

Holzinger, A. and Pichrtová, M. (2016) Abiotic stress tolerance of charophyte green algae: new challenges for omics techniques. *Front. Plant Sci.* 7: 678.

Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., et al. (2014) *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* 5: 3978.

Ito, M. (2008) Studies on perilla, agarwood, and cinnamon through a combination of fieldwork and laboratory work. *J. Nat. Med.* 62: 387–395.

Ju, C., Van de Poel, B., Cooper, E.D., Thierer, J.H., Gibbons, T.R., Delwiche, C.F., et al. (2015) Conservation of ethylene as a plant hormone over 450 million years of evolution. *Nat. Plants* 1: 14004.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45: D353–D361.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanebe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44: D457–D462.

Katoh, K., Misawa, K., Kuma, K.-I. and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30: 3059–3066.

Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30: 772–780.

Knack, J.J., Wilcox, L.W., Delaux, P.M., Ané, J.M., Piotrowski, M.J., Cook, M.E., et al. (2015) Microbiomes of streptophyte algae and bryophytes suggest that a functional suite of microbiota fostered plant colonization of land. *Int. J. Plant Sci.* 176: 405–420.

Kroken, S.B., Graham, L.E. and Cook, M.E. (1996) Occurrence and evolutionary significance of resistant cell walls in charophytes and bryophytes. *Amer. J. Bot.* 83: 1241–1254.

Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33: 1870–1874.

Labeeuw, L., Martone, P.T., Boucher, Y. and Case, R.J. (2015) Ancient origin of the biosynthesis of lignin precursors. *Biol. Direct* 10: 23.

Lacombe, E., Hawkins, S., Doorsselaere, J.V., Piquemal, J., Goffner, D., Poeydomenge, O., et al. (1997) Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: cloning, expression and phylogenetic relationships. *Plant J.* 11: 429–441.

Lee, Y., Rubio, M.C., Alassimone, J. and Geldner, N. (2013) A mechanism for localized lignin deposition in the endodermis. *Cell* 153: 402–412.

Leliaert, F., Smith, D.R., Moreau, H., Herron, M.D., Verbruggen, H., Delwiche, C.F., et al. (2012) Phylogeny and molecular evolution of the green algae. *Crit. Rev. Plant Sci.* 31: 1–46.

Lewis, L.A. and McCourt, R.M. (2004) Green algae and the origin of land plants. *Amer. J. Bot.* 91: 1535–1556.

Li, Z. and Nair, S.K. (2015) Structural basis for specificity and flexibility in a plant 4-coumarate:CoA ligase. *Structure* 23: 2032–2042.

Ligrone, R., Carafa, A., Duckett, J.G., Renzaglia, K.S. and Ruel, K. (2008) Immunocytochemical detection of lignin-related epitopes in cell walls in bryophytes and the charalean alga *Nitella*. *Plant Syst. Evol.* 270: 257–272.

Logemann, E., Reinold, S., Somssich, I.E. and Hahlbrock, K. (1997) A novel type of pathogen defense-related cinnamyl alcohol dehydrogenase. *Biol. Chem.* 378: 909–913.

Luo, Y.-J., Takeuchi, T., Koyanagi, R., Yamada, L., Kanda, M., Khalturina, M., et al. (2015) The *Lingula* genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nat. Commun.* 6: 8301.

Mansell, R.L., Gross, G.G., Stöckigt, J., Franke, H. and Zenk, M.H. (1974) Purification and properties of cinnamyl alcohol dehydrogenases from higher plants involved in lignin biosynthesis. *Phytochemistry* 13: 2427–2435.

Marjamaa, K., Kukkola, E.M. and Fagerstedt, K.V. (2009) The role of xylem class III peroxidases in lignification. *J. Exp. Bot.* 60: 367–376.

Marchel-Bauer, A., Derbyshire, M.K., Gonzales N.R., Lu, S., Chitsaz F., Geer L.Y., et al. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43: D222–D226.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpwicz, S.J., Witman, G.B., et al. (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318: 245–250.

Meyer, K., Shirley, A.M., Cusumano, J.C., Bell-Lelong, D.A. and Chapple, C. (1998) Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 95: 6619–6623.

Milkowski, C. and Strack, D. (2010) Sinapate esters in brassicaceous plants: biochemistry, molecular biology, evolution and metabolic engineering. *Planta* 232: 19–35.

Moffitt, M.C., Louie, G.V., Bowman, M.E., Pence, J., Noel, J.P. and Moore, B.S. (2007) Discovery of two cyanobacterial phenylalanine ammonia lyases: kinetic and structural characterization. *Biochemistry* 46: 1004–1012.

NCBI Resource Coordinators (2016) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 44: D7–D19.

Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D.G., et al. (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature* 497: 579–584.

Oliva, J., Rommel, S., Fossdal, C.G., Hietala, A.M., Nemesio-Gorriz, M., Solheim, H., et al. (2015) Transcriptional responses of Norway spruce (*Picea abies*) inner sapwood against *Heterobasidion parviporum*. *Tree Physiol.* 35: 1007–1015.

Palenik, B., Grimwood, J., Aerts, A., Rouze, P., Salamov, A., Putnamn, N., et al. (2007) The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl. Acad. Sci. USA* 104: 7705–7710.

Pan, H., Zhou, R., Louie, G.V., Muhlemann, J.K., Bomati, E.K., Bowman, M.E., et al. (2014) Structural studies of cinnamoyl-CoA reductase and cinnamyl-alcohol dehydrogenase, key enzymes of monolignol biosynthesis. *Plant Cell* 26: 3709–3727.

Pedreira, J., Herrera, M.T., Zarra, I. and Revilla, G. (2011) The overexpression of *AtPrx37*, an apoplastic peroxidase, reduces growth in *Arabidopsis*. *Physiol. Plant.* 141: 177–187.

Raes, J., Rohde, A., Christensen, J.H., Van de Peer, Y. and Boerjan, W. (2003) Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiol.* 133: 1051–1071.

Ralph, S.G., Jancsik, S. and Bohlmann, J. (2007) Dirigent proteins in conifer defense II: extended gene discovery, phylogeny, and constitutive and stress-induced gene expression in spruce (*Picea* spp.). *Phytochemistry* 68: 1975–1991.

Renaults, H., Alber, A., Horst, N.A., Basilio Lopes, A., Fich, E.A., Kriegshauser, L., et al. (2017) A phenol-enriched cuticle is ancestral to lignin evolution in land plants. *Nat. Commun*. 8: 14713.

Rensing, S.A. (2017) Why we need more non-seed plant models. *New Phytol*. (in press).

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., et al. (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319: 64–69.

Rippert, P., Puyaubert, J., Grisollet, D., Derrier, L. and Matringe, M. (2009) Tyrosine and phenylalanine are synthesized within the plastids in *Arabidopsis*. *Plant Physiol*. 149: 1251–1260.

Saballos, A., Ejeta, G., Sanchez, E., Kang, C. and Vermerris, W. (2009) A genomewide analysis of the cinnamyl alcohol dehydrogenase family in sorghum [*Sorghum bicolor* (L.) Moench] identifies SbCAD2 as the brown midrib6 gene. *Genetics* 181: 783–795.

Saleem, M., Nazir, M., Ali, M.S., Hussain, H., Lee, Y.S., Riaz, N., et al. (2010) Antimicrobial natural products: an update on future antibiotic drug candidates. *Nat. Prod. Rep.* 27: 238–254.

Schwede, T.F., Rétey, J. and Schulz, G.E. (1999) Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile. *Biochemistry* 38: 5355–5361.

Shigeto, J., Kiyonaga, Y., Fujita, K., Kondo, R. and Tsutsumi, Y. (2013) Putative cationic cell-wall-bound peroxidase homologues in *Arabidopsis*, AtPrx2, AtPrx25, and AtPrx71, are involved in lignification. *J. Agric. Food Chem.* 61: 3781–3788.

Shigeto, J. and Tsutsumi, Y. (2016) Diverse functions and reactions of class III peroxidases. *New Phytol.* 209: 1395–1402.

Shockey, J.M., Fulda, M.S. and Browse, J. (2003) *Arabidopsis* contains a large superfamily of acyl-activating enzymes. Phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme A synthetases. *Plant Physiol.* 132: 1065–1076.

Sørensen, I., Pettolino, F.A., Bacic, A., Ralph, J., Lu, F., O'Neill, M.A., et al. (2011) The charophycean green algae provide insights into the early origins of plant cell walls. *Plant J.* 68: 201–211.

Stothard, P. (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28: 1102–1104.

Suzuki, S. and Umezawa, T. (2007) Biosynthesis of lignans and norlignans. *J. Wood Sci.* 53: 273–284.

Timmis, J.N., Ayliffe, M.A., Huang, C.Y. and Martin, W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* 5: 123–135.

Van de Poel, B., Cooper, E.D., Van Der Straeten, D., Chang, C. and Delwiche, C.F. (2016) Transcriptome profiling of the green alga *Spirogyra pratensis* (Charophyta) suggests an ancestral role for ethylene in cell wall metabolism, photosynthesis, and abiotic stress responses. *Plant Physiol.* 172: 533–545.

Vanholme, R., Cesarino, I., Rataj, K., Xiao, Y., Sundin, L., Goeminne, G., et al. (2013) Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. *Science* 341: 1103–1106.

Vanholme, R., Storme, V., Vanholme, B., Sundin, L., Christensen, J.H., Goeminne, G., et al. (2012) A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis*. *Plant Cell* 24: 3506–3529.

Vogt, T. (2010) Phenylpropanoid biosynthesis. *Mol. Plant* 3: 2–20.

Watts, K.T., Mijts, B.N., Lee, P.C., Manning, A.J. and Schmidt-Dannert, C. (2006) Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic amino acid lyase family. *Chem. Biol.* 13: 1317–1326.

Welinder, K.G., Jutesen, A.F., Kjærsgård, V.H., Jensen, R.B., Rasmussen, S.K., Jepersen, H.M., et al. (2002) Structural diversity and transcription of class III peroxidases from *Arabidopsis thaliana*. *Eur. J. Biochem.* 269: 6063–6081.

Weng, J.-K. (2013) The evolutionary paths towards complexity: a metabolic perspective. *New Phytol.* 201: 1141–1149.

Weng, J.-K., Akiyama, T., Bonatwitz, N.D., Li, X., Ralph, J. and Chapple, C. (2010) Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte *Selaginella* and flowering plants. *Plant Cell* 22: 1033–1045.

Weng, J.-K., Akiyama, T., Ralph, J. and Chapple, C. (2011) Independent recruitment of an O-methyltransferase for syringyl lignin biosynthesis in *Selaginella moellendorffii*. *Plant Cell* 23: 2708–2724.

Weng, J.-K., Li, X., Stout, J. and Chapple, C. (2008) Independent origins of syringyl lignin in vascular plants. *Proc. Natl. Acad. Sci. USA* 105: 7887–7892.

Whetten, R. and Sederoff, R. (1995) Lignin biosynthesis. *Plant Cell* 7: 1001–1013.

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., et al. (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. USA* 111: E4859–E4868.

Woodward, S. and Pearce, R.B. (1988). The role of stilbenes in resistance of *Sitka spruce* (*Picea sitchensis* (Bong.) Carr.) to entry of fungal pathogens. *Physiol. Mol. Plant Pathol.* 33: 127–149.

Xu, Z., Zhang, D., Hu, J., Zho, X., Ye, X., Reichel, K.L., et al. (2009) Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. *BMC Bioinformatics* 10: S3.

Youn, B., Camacho, R., Moinuddin, S.G.A., Lee, C., Davin, L.B., Lewis, N.G., et al. (2006) Crystal structures and catalytic mechanism of the *Arabidopsis* cinnamyl alcohol dehydrogenases *At*CAD5 and *At*CAD4. *Org. Biomol. Chem.* 4: 1687–1697.

Zhao, Q., Nakashima, J., Chen, F., Yin, Y., Fu, C., Yun, J., et al. (2013) *LACCASE* is necessary and nonredundant with *PEROXIDASE* for lignin polymerization during vascular development in *Arabidopsis*. *Plant Cell* 25: 3976–3987.