The evolution of the phenylpropanoid pathway entailed pronounced radiations and divergences of enzyme families

Sophie de Vries[1,2*], Janine MR Fürst-Jansen[2*], Iker Irisarri[2,8*], Amra Dhabalia Ashok[2], Till Ischebeck[3,4,5], Kirstin Feussner[3,4], Ilka N Abreu[3], Maike Petersen[6], Ivo Feussner[3,4,5], Jan de Vries[2,7,8#]

1 — Heinrich-Heine University Düsseldorf, Population Genetics, Universitätsstr. 1, 40225 Düsseldorf, Germany

2 — University of Goettingen, Institute for Microbiology and Genetics, Department of Applied Bioinformatics, Goldschmidtstr. 1, 37077 Goettingen, Germany

3 — University of Goettingen, Albrecht-von-Haller-Institute for Plant Sciences, Department of Plant Biochemistry, Justus-von-Liebig Weg 11, 37077 Goettingen, Germany

4 — University of Goettingen, Goettingen Center for Molecular Biosciences (GZMB), Goettingen Metabolomics and Lipidomics Laboratory, Justus-von-Liebig Weg 11, 37077 Goettingen, Germany

5 — University of Goettingen, Goettingen Center for Molecular Biosciences (GZMB), Department of Plant Biochemistry, Justus-von-Liebig Weg 11, 37077 Goettingen, Germany

6 — Philipps-Universität Marburg, Institut für Pharmazeutische Biologie und Biotechnologie, Robert-Koch-Str. 4, 35037, Marburg, Germany

7 — University of Goettingen, Goettingen Center for Molecular Biosciences (GZMB), Department of Applied Bioinformatics, Goldschmidtsr. 1, 37077 Goettingen, Germany

8 — University of Goettingen, Campus Institute Data Science (CIDAS), Goldschmidstr. 1, 37077 Goettingen, Germany

*these authors contributed equally

author for correspondence:

Jan de Vries

Georg-August University Göttingen, Institute for Microbiology and Genetics, Department of Applied Bioinformatics, Goldschmidtstr. 1, 37077 Göttingen, Germany, phone: +49-551-39-23755

devries.jan@uni-goettingen.de

ORCIDs: Sophie de Vries 0000-0002-5267-8935, Janine MR Fürst-Jansen 0000-0002-5269-8725, Iker Irisarri 0000-0002-3628-1137, Amra Dhabalia Ashok 0000-0001-5787-6941, Till Ischebeck 0000-0003-0737-3822, Ilka N Abreu 0000-0003-4728-0161, Kirstin Feussner 0000-0003-1634-8258, Maike Petersen 0000-0001-7769-8556, Ivo Feussner 0000-0002-9888-7003, Jan de Vries, 0000-0003-3507-5195

AUTHOR EMAILS: sophie.devries@uni-goettingen.de; Janine.fuerst-jansen@uni-goettingen.de; iker.irisarriaedo@uni-goettingen.de; amradhabalia.ashok@uni-goettingen.de; tischeb@uni-goettingen.de; kfeussn@uni-goettingen.de; ilkanacif.deabreu@uni-goettingen.de; petersen@mailer.uni-marburg.de; ifeussn@uni-goettingen.de; devries.jan@uni-goettingen.de

KEYWORDS: phenylpropanoid biosynthesis; plant evolution; evolution of gene families; evo-physio

RUNNING HEAD: Phenylpropanoid evolution: divergence & radiation

ABSTRACT

Land plants constantly respond to fluctuations in their environment. Part of their response is the production of a diverse repertoire of specialized metabolites. One of the foremost sources for metabolites relevant to environmental responses is the phenylpropanoid pathway, which was long thought to be a land plant-specific adaptation shaped by selective forces in the terrestrial habitat. Recent data have however revealed that streptophyte algae, the algal relatives of land plants, have candidates for the genetic toolkit for phenylpropanoid biosynthesis and produce phenylpropanoid-derived metabolites. Using phylogenetic and sequence analyses, we here show that the enzyme families that orchestrate pivotal steps in phenylpropanoid biosynthesis have independently undergone pronounced radiations and divergence in multiple lineages of major groups of land plants; sister to many of these radiated gene families are streptophyte algal candidates for these enzymes. These radiations suggest a high evolutionary versatility in the enzyme families involved in the phenylpropanoid-derived metabolism across embryophytes. We suggest that this versatility likely translates into functional divergence and may explain the key to one of the defining traits of embryophytes: a rich specialized metabolism.

INTRODUCTION

A diverse profile of specialized metabolites is one of the characteristics of land plants (embryophytes). Almost any aspect of the biology of land plants is underpinned by specialized metabolites—be it the phytohormones that are major modulators upstream in various regulatory hierarchies (Scheres and van der Putten, 2017; Berens et al., 2017; Blázquez et al., 2020) or pigments that give land plants their color and attune photochemical properties (Jahns and Holzwarth, 2012).

A key aspect of the biological relevance of most specialized metabolites is their use under challenging environmental conditions. Indeed, the elaboration of their specialized metabolism is considered one of the drivers for the massive radiation of embryophytes on land (Weng 2013). Moreover, a diversity of specialized metabolism likely played a key role during the earliest steps of plants on land—allowing for the production of compounds that protected land plants against the challenges of the terrestrial environment such as drought and increased UV radiation (Rensing, 2017; de Vries and Archibald, 2018; Fürst-Jansen et al., 2020; Jiao et al., 2020). A major pathway giving rise to a variety of specialized metabolites that act in warding off environmental stressors is the biosynthesis of phenylpropanoids (Dixon and Paiva, 1995; Dixon et al., 2002; Vogt, 2010).

The phenylpropanoid pathway is the source of precursors for thousands of metabolites with multifaceted functions, and accounts for about 40% of organic carbon on earth (Vogt, 2010). One facet of these functions is that phenylpropanoid-derived compounds act as structural polymers, foremost among which are the different types of lignin (Ralph et al., 2004; Vanholme et al., 2012; Vanholme et al., 2019). Another prominent facet is that these metabolites act as UV-protecting substances. While some of the best-known UV screens are flavonoids, various other compounds stemming from the phenylpropanoid pathway are equally potent UV protectants (Sheahan, 1996; Booij James et al., 2000; Sytar et al. 2018; Xue et al., 2020). The list of links between phenylpropanoid-derived compounds and the response to environmental challenges could be continued; in fact, the response to almost any abiotic stressor that plants face in the terrestrial habitat involves the action of phenylpropanoid-derived compounds (for comprehensive reviews see, e.g., Dixon and Paiva, 1995; Vogt, 2010). Furthermore, phenylpropanoid-derived compounds are involved in the defense responses against plant pathogens in many land plant lineages (Danielsson et al. 2011; Ponce de Leon et al., 2012; König et al., 2014; Overdijk et al., 2016; Carella et al., 2019).

All embryophytes make use of the enzymatic routes in the phenylpropanoid pathway. For example, the utilization of flavonoids under UV stress appears to be a conserved response across Embryophyta (Wolf et al., 2010; Clayton et al., 2018). However, not all embryophytes produce the same compounds under the same stress conditions—in contrary, the diversity of compounds is immense. Major differences in the biosynthesis of phenylpropanoid-derived compounds occur in distinct lineages of land plants. This includes specialized roles such as the flower coloration determining anthocyanins that attract pollinators (Miller et al., 2011; Sheehan et al., 2012); such a role of anthocyanins is obviously limited to flowering plants and can vary even among closely related species (Saito and Harborne, 1992). That said, Piatkowski et al. (2020) phylogenetically inferred that orthologs for the entire anthocyanin biosynthesis pathway are already present in the ancestor of seed plants and more than half of the important orthogroups were already present in the most recent common ancestor of all land plants. An important recent insight into the deep evolutionary roots of flavonoid biosynthesis was the discovery of auronidins—a novel class of red flavonoid pigments that are synthesized in the bryophyte *Marchantia polymorpha* (Berland et al. 2019). Further, for example, Renault et al. (2017a) reported on the enrichment of the *Physcomitrium patens* (moss) cuticle in phenolic compounds—an enrichment that hinges on the action of a cytochrome P450 enzyme that is orthologous to enzymes that act in lignin biosynthesis; the

production of lignin might trace its evolutionary roots back to an ancient set of enzymes acting in the production of complex, phenol-enriched polymers (Renault et al., 2019). Carella et al. (2019) showed that the liverwort model plant *Marchantia polymorpha* triggers phenylpropanoid biosynthesis upon attack by the oomycete phytopathogen *Phytophthora palmivora*. Similar responses towards phytopathogens are known from gymnosperms (Oliva et al., 2015) and angiosperms (Dixon and Paiva, 1995; Bednarek et al., 2005; Kaur et al., 2010; Chezem et al., 2017; Carella et al., 2019). Thus, all land plants use the core framework of the phenylpropanoid pathway to produce—often lineage-specific—variations of phenylpropanoid derivatives that aid in response to biotic and abiotic stressors.

The production of the chemical repertoire of land plants is often catalyzed by members of large enzyme-coding gene families (Shockey et al., 2003; Nelson and Werck-Reichhart, 2011; Renault et al., 2017b), and this also seems to be the case for the enzymes involved in phenylpropanoid biosynthesis (Hamberger et al., 2007; Xu et al., 2009; Vogt, 2010). It is thus conceivable that various adaptive forces have shaped the families of enzymes that act in the phenylpropanoid pathway, leading to multiple independent cases of sub- and neofunctionalization (see also Rensing, 2010). An inference of the common (minimal) set of enzymes that were present in the last common ancestors (LCA) of (i) streptophytes, (ii) land plants and their closest streptophyte algal relatives, and (iii) land plants can thus shed light on which enzymatic building blocks evolution acted upon to give rise to the elaborate chassis of the phenylpropanoid pathway.

The phenylpropanoid pathway has long been considered to be specific to Embryophyta. However, homologs of the genes coding for the enzymes that constitute the embryophytic phenylpropanoid pathway can be found in extant algal relatives of land plants, suggesting that they were already present in a common ancestor shared by streptophyte algae and land plants (de Vries et al., 2017; Renault et al., 2019; Maeda and Fernie, 2021). Since the beginning of 2020, we have genome data from all major lineages of Streptophyta—except Coleochaetophyceae (Szövényi et al., 2021); only using this extended repertoire of species and sequences allows us to pinpoint which subfamilies and/or which ancestral enzyme of multiple subfamilies were present in the aforementioned LCAs. Compounds that, in land plants, emerge from the phenylpropanoid pathway are indeed found in algae; these include flavonoids and lignin-like compounds in streptophyte algae (Delwiche et al., 1989; Sørensen et al., 2011; Jiao et al., 2020) and core phenylpropanoid building blocks as well as flavonoids in a phylodiverse set of algae (Goiris et al. 2014). Interestingly, lignin-like compounds were even found in distantly-related red macroalgae (Martone et al., 2009)—although this is likely

a case of convergence that builds on an unknown enzymatic framework. However, even within the green lineage (Chloroplastida), the question of the deep evolutionary roots of the phenylpropanoid pathway is wide open.

Investigations of the algal relatives of land plants have strongly benefitted from recent progress in phylogenomics on plants and algae. A major outcome of these recent phylogenomic analyses was that the Zygnematophyceae have been pinpointed as the class of algae most closely related to land plants (Wodniok et al., 2011; Wickett et al., 2014; Leebens-Mack et al., 2019). Hand in hand with these phylogenomic efforts went the generation of genomic (Hori et al., 2014; Nishiyama et al., 2018; Cheng et al., 2019; Wang et al., 2020; Jiao et al., 2020) and transcriptomic data on streptophyte algae (Ju et al., 2015; Rippin et al., 2017; de Vries et al., 2018; de Vries et al., 2020). Additionally, critical gaps in the land plant tree of life have been filled; this includes recent publications of the first genomes of liverworts (Bowman et al., 2017), ferns (Li et al., 2018) and hornworts (Szövenyi et al., 2015; Li et al., 2020; Zhang et al., 2020). These data allow for the fine-grained tracing of the evolution of key plant enzyme families across the green tree of life. Recent studies have illuminated the diversity of enzymes in the routes towards flavonoids and anthocyanins as well as the PAL-dependent pathway of salicylic acid biosynthesis via benzoic acid (Piatkowski et al., 2020; Güngör et al., 2021, de Vries et al., 2021).

In this study, we infer the evolutionary history of eleven critical enzyme families known to be woven into the mesh of routes from phenylpropanoids to lignin biosynthesis in land plants; we have paid particular attention to the routes leading to the biosynthesis of lignin. We use the new diversity of genomic and transcriptomic data from land plants as well as streptophyte and chlorophyte algae to infer the origin of these large gene families. The datasets were chosen in a manner that they cover the breadth of streptophyte diversity while providing a balanced sampling; the latter is especially relevant in light of the high number of genomes of flowering plants. We aimed to include at least one representative of each of the major lineages of streptophytes in the datasets we surveyed. Our data pinpoint deep homologs of candidate enzymes in streptophyte algae for L-**p**henylalanine **a**mmonia-**l**yase (PAL), **4**-**c**oumarate–CoA **l**igase (4CL), **c**affeoyl-**CoA** ***O*-m**ethyltransferase (CCoAOMT); further, for streptophyte *and* chlorophyte algae, we pinpoint homologs for **c**innamoyl-**C**oA **r**eductase (CCR), **c**innamyl **a**lcohol **d**ehydrogenase (CAD), and potentially relevant **m**ono**a**cyl**g**lycerol **l**ipases (MAGLs). Further, we find that often the functionally characterized enzymes of the core phenylpropanoid and lignin biosynthesis routes derive from lineage-specific radiations, limiting the inference of function outside the model system. Nonetheless we could infer

which subfamilies were present in LCAs along the trajectory of streptophyte evolution, even though ancestral functional inference was limited. That said, for enzyme families with deep homologs, we approximated the function through domain prediction and the conservation (or lack thereof) of key residues of known functional importance. We found that all enzyme families underwent several lineage-specific expansions and losses as well as bursts in growth of enzyme families that occurred early during the radiation of land plants. We hypothesize that lineage-specific expansions in these enzyme families is linked with the diversity of lineage-specific phenylpropanoid derivatives and functions that occur in the species analyzed here.

RESULTS AND DISCUSSION

The checkered occurrence of phenylalanine ammonia-lyase among streptophyte algae

The conversion of the aromatic amino acid phenylalanine and/or tyrosine into cinnamate and/or *p*-coumarate is the first step of the plant phenylpropanoid pathway (Figure 1). This first committed step is catalyzed by PAL and the bifunctional L-phenylalanine/ L-tyrosine ammonia-lyase (PTAL; Barros and Dixon, 2020). For a long time, it was thought that among Chloroplastida, PAL/PTAL were limited to land plants; their gain was considered to have occurred via a lateral gene transfer event that has occurred at the base of the land plant clade (Emiliani et al., 2009). Recently, however, genes coding for putative PAL-like enzymes were detected in streptophyte algae, such as the filamentous streptophyte alga *Klebsormidium nitens* (de Vries et al., 2017). In light of the recent surge in available genomes from across the green tree of life, we set out to explore the evolutionary history of PAL.

Using *At*PAL1 as a bait sequence, we screened protein data from diverse land plants and all streptophyte algal genomes available. Hits among streptophyte algae fell into three categories: (i) proteins of between 480 (*Klebormidium nitens* PAL, kfl00104_0290_v1.1) and 527 amino acids (two homologs in *Chara braunii*; g57646_t1 and g34530_t1); (ii) short proteins such as ME1156409C09523 of *Mesotaenium endlicherianum*, which is 184 amino acid in length; (iii) long proteins of between 991 (*Chlorykbus atmophyticus* Chrsp482S06115) and 1115 amino acids (*Klebsormidium nitens* kfl00024_0250_v1.1). Proteins falling into the third category are fusions of an aromatic amino acid lyase domain and a putative tRNA synthetase; homologs of such also occur in land plants (e.g., AT3G02760 and Os05g0150900). The short proteins of category '(ii)' are found in some Zygnematophyceae and are proteins of unknown function with a putative HAL domain. Genomic artefacts leading to this result can be excluded given that these types of protein-

encoding genes have been recovered for independent Zygnematophyceae that are likely >500 million years divergent from one another. The proteins in category '(i)' are those with the highest identity to *bona fide* land plant PALs. This category includes the promising PAL candidate kfl00104_0290_v1.1 (see de Vries et al., 2017). We therefore set out to further explore these PAL-like candidates, which noteworthily were only found in the genome of *Klebsormidium nitens* and *Chara braunii* and no other streptophyte algal genome data.

To understand the evolutionary history of PAL in streptophytes, we computed a maximum likelihood phylogeny (Figure 2). The phylogenetic analysis included the aforementioned PAL homologs from diverse Streptophyta as well as bacterial and fungal PALs. In agreement with previous studies (Emiliani et al., 2009; de Vries et al., 2017), the fungal and bacterial PAL sequences are closely related (bacterial clade: bootstrap support 100; fungal clade: bootstrap support 85) to the clade of land plant PALs (bootstrap support: 96). Additionally, we included diverse eukaryotic and prokaryotic histidine ammonia lyases (HALs) based on the set obtained from de Vries et al. (2017), which, as in this latter study, form two clades with eukaryotic and prokaryotic HALs (bootstrap support 83 and 97). All putative PAL-like candidate sequences from *Chara braunii* clustered with HAL sequences, one (*Chara braunii* g66119_t1) with a cyanobacterium (bootstrap support 100), and two (*Chara braunii* g34530_t1 and *Chara braunii* g57646_t1) sister to an entire bacterial HAL clade (bootstrap support 100). Both sequences retrieved for *Klebsormidium nitens*, in contrast, clustered with the PAL clades, one *K. nitens* kfl00024_0250_v1.1 with low support (boostrap level 61) and a rather long branch as sister to plant and fungal PALs, showing that its placement is not fully resolved and that further analyses are required to identify its true identity. The second *Klebsormidium nitens* sequence, kfl00104_0290_v1.1, clustered within the clade of bacterial PALs (bootstrap support 100), of which some were already functionally characterized—for example, the characterized PAL of *Nostoc punctiforme* (Moffit et al., 2007). This is in agreement with the placement of this protein sequence in de Vries et al. 2017 and supports it as a putative PAL sequence.

In sum, the evolutionary origin of streptophyte PAL appears to be complex and remains obscure: it may be that PAL had a distinct origin in streptophyte algae and land plants, yet the pattern may also be explained by an origin via an endosymbiotic gene transfer from the cyanobacterial plastid progenitor that was retained by streptophytes with a gain of an extra C-terminal domain later in the evolution of embryophytes, resulting in the two distinct PAL clades. It is however important to note that the 3' region of the genomic locus that codes for the shorter *K. nitens* protein kfl00104_0290_v1.1 contains sequence

information that resembles code for the missing C-terminal stretch; thus, the C-terminal stretch might have simply been secondarily lost in *K. nitens*. Independently, the presence of PALs in fungi further complicates the evolutionary scenario. Rampant gene losses during eukaryotic evolution or convergent domain acquisitions, as well as horizontal gene transfer from plants (as hypothesized by Emiliani et al., 2009) are other scenarios that can explain the evolutionary origin of PALs in streptophytes and thus ultimately in land plants.

Streptophyte algae have an expanded and divergent repertoire of cytochrome P450 monooxygenases with no clear C4H orthologs

After the synthesis of cinnamate by PAL, two routes open up (Figure 1). One of them is the conversion of cinnamate into *p*-coumarate, which is catalyzed by **c**innamate **4**-**h**ydroxylase (C4H). C4H belongs to the large class of CYP450 enzymes, which are present among all domains of life (Omura, 1999). Among the CYP450 enzymes C4H belongs to the CYP450 subfamily 73 (CYP73). In land plants, CYP450s have undergone massive duplication and subfunctionalization, underpinning the specialized metabolic capabilities of embryophytes (Nelson and Werck-Reichhart, 2011); for example, the CYP73 subfamily belongs to the larger CYP71 clan. The specific CYP450 monoxygenases that fall into the group of C4H appear to be limited to land plants: clear orthologs can be found in bryophytes and tracheophytes (Emiliani et al., 2009; de Vries et al., 2017). That said, the product of the reaction carried out by C4H in land plants (*p*-coumarate) has been detected via UHPLC-MS/MS in phylodiverse algae (Goiris et al., 2014). Therefore, there appears to exist a route towards *p*-coumarate that is either independent of C4H via direct transformation of tyrosine by PTAL or carried out by a highly divergent C4H homolog. PTALs have so far however been observed in monocots (Barros et al., 2016; Barros and Dixon, 2020), suggesting a different CYP73 subfamily enzyme that may carry out the reaction. Owing to the recent increase in genomic data available for streptophyte algae, we revisited the question of when C4H-based *p*-coumarate might have emerged and explored CYP450 evolution.

We sampled C4H homologs from seven land plant genomes that had a BLAST bit score (a normalized alignment score) of at least 200, as well as seven streptophyte algal and five chlorophyte algal genomes that had a bit score of at least 100. We aligned all C4H homologs and computed a maximum likelihood phylogeny (Figure S1). The well-characterized C4H of *Arabidopsis* fell into a clade with full (100 out of 100) bootstrap support; this clade included at least one C4H homolog from each of the other six land plants genomes, corroborating the notion that all land plants have C4H orthologs, which appear

conserved in their function from bryophytes to tracheophytes (Russel and Conn, 1967; Urban et al., 1994; Ro et al., 2001; Wohl and Petersen, 2020) and thus since the LCA of land plants. However, no algal sequences fell into this clade. That said, we observed four well-supported clades of streptophyte algal CYP450 enzymes (Figure S1). Investigating the genetic distances, we find that some of the streptophyte algal sequences have a closer genetic distance to the C4H-like clade than to sequences from land plants (including *Arabidopsis thaliana*) from other CYP450 subfamilies (Table S1). While these sequences remain of unknown function, they are candidates for the CYP450 enzyme family that catalyzes the C4H-function in algae.

A deep split of streptophyte 4CL/ACS

The second route that opens up after the PAL-dependent step is the conversion of cinnamate into cinnamoyl-CoA. This is carried out by the AMP-forming synthetase/ligase 4CL and potentially other enzymes annotated as acyl-CoA synthetases (ACS/ACoS) (Shockey et al., 2003; Figure 1). Altogether, these enzymes belong to a large family of distantly related acyl-activating enzymes (AAEs), such as the long-chain acyl-CoA synthetases (LACS) and many more (Shockey et al., 2003; Figure S2). Homologs with affinity to 4CL appear to occur across chlorophytes and streptophytes (Labeeuw et al., 2015). At least in *Arabidopsis*, the family of 4CLs has expanded and includes four canonical ("4CL") and nine additional 4CL-like ("4CLL") members, falling into AAE clade IV and V as defined by Shockey et al. (2003). We thus set out to understand what the 4CL repertoire of the last common ancestor of land plants and the one shared with algae might have looked like.

In order to trace the radiation of 4CLs across the green tree of life, we sampled 4CL homologs from genomes of nine land plants, seven streptophyte algae (plus four transcriptomes of streptophyte algae) and five chlorophyte algae that had a minimum of 400 and a maximum of 1150 amino acids in length and showed affinity to the 4CL clade in a larger phylogenetic survey (Figure S2). We recovered a large clade (bootstrap support 85) that included all *bona fide* 4CL paralogs and ACOS5 of *Arabidopsis thaliana* (Figure 3); ACOS5 has been previously associated with the *bona fide* 4CL clade (Shockey et al., 2003) but it did show only inconsistent activity on typical substrates of 4CL (Costa et al., 2005) and appears to have a very specific function in sporopollenin biosynthesis of pollen (de Azevedo Souza et al., 2009). These observations agree with the presence of different (but conserved) amino acids at sites that bind hydroxycinnamate in typical 4CLs (Figure 3), which might suggest a different natural substrate for ACOS given that affinity is mostly determined by the

binding pocket size (Hu et al. 2010). We further recovered the angiosperm-specific separation defined by Ehlting et al. (1999) into class I and class II 4CLs. Additional lineage-specific radiations occurred, for example in *Physcomitrium patens*, which fell into a clade of bryophyte sequences (bootstrap support of 70) and in *Selaginella moellendorffii* (spread out over the fully-supported clade of 4CLs). The common ancestor of land plants appears to have possessed an ACOS5-like and one 4CL-like gene, all other 4CL paralogs in this clade likely emerged later during land plant evolution. Clustering with AAE clade IV (including *At*ACOS5, and *At*4CL1,2,3 and 5 (bootstrap support 85)) are sequences from five streptophyte algae. Each of the five streptophyte algae possesses one homolog to these five types of AMP-forming ligases with 4-coumarate–CoA synthesizing activity. When we predicted the tertiary structure of *Chlorokybus atmophyticus* Chrsp175S02417 and *Penium margaritaceum* 006213.t1 via Iterative Threading ASSEmbly Refinement (I-TASSER; Zhang 2008), we recovered, in both cases, firefly luciferases as best match (TM-scores 0.852 and 0.918; 1BA3 and 2D1S; Franks et al., 1998; Nakatsu et al., 2006). Investigating the putative structure of the other streptophyte algal sequences (*Zygnema circumcarinatum* DN42558_c0_g1_i1, *Spirogyra pratensis* 3442_c2_g1_i6, and *Klebsormidium nitens* 00016_0470_v1.1), however, always recovered *Populus tomentosa* 4CL (3A9U; Hu et al., 2010) as their closest structural analog (TM scores of 0.957, 0.969, and 0.961, respectively). Hence, we hypothesize that a 4CL/ACOS5-like encoding gene was present in the last common ancestor of all streptophytes. Underpinning this hypothesis is that the sequence of the amino acids in the binding pocket in the streptophyte algal 4CL homologs is consistent with that of 4CL homologs from other land plants (including that of *Arabidopsis thaliana*; Figure 3). Further, the amino acids relevant for the enzymatic function (i.e., the residues KQK involved in adenylation, nucleophilic substitution and coumaroyl-AMP cleavage) are also conserved across most 4CL/ACOS5 sequences, including those of the streptophyte algae. Variation in these residues is already apparent in 4CLL homologs and outside of the 4CL/ACOS5/4CLL clade these residues show high variability (Figure 3).

A similar pattern was observed when we investigated the domain structure of all recovered sequences. Most of the 4CL/ACOS5-like sequences contained four domains: Phosphopantetheine binding ACP domain (IPR025110), AMP-binding, conserved site (IPR020845), AMP-dependent synthetase-like superfamily (IPR042099), AMP-dependent synthetase/ligase (IPR000873; Figure S3). There were four exceptions to this pattern. They include one sequence from the water fern *Azolla filiculoides* (Azfis0013.g013344) and two hornwort sequences from *Anthoceros agrestis* BONN (Sc2ySwM344.2803.3 and

Sc2ySwM344.2803.4), which all missed the Phosphopantetheine binding ACP domain (IPR025110). The other exception was *At*ACOS5, which is the only sequence in this clade that missed the conserved AMP-binding site (IPR020845). The domain pattern is similar across the 4CLL-like clade, too. Yet, more sequences miss either the IPR025110 and/or the IPR020845 domain. The streptophyte algal sequences within the ACOS5/4CL clade contained all four domains, whereas algal sequences outside of this clade missed at least one—but recovered several other domains. These additional domains are not conserved within the phylogenetic subclades of these algal sequences and only exceptionally occur in the *4CL/ACOS4/4CLL* (two sequences) or *Other AMP-dependent synthetase and ligase family protein* clades (two sequences).

We recovered a second clade of spermatophyte sequences (bootstrap support 89) representing AAE clade V enzymes, which contains 4CLL8 and several other ATP-ligases of *Arabidopsis thaliana* with predicted 4CL activity, including OPCL1 (matching 4CL-like 5 with a 100% amino acid identity according to Uniprot). Sequences in this clade, however, diverge in the amino acids that are involved in the formation of the binding pocket in the canonical 4CLs (Figure 3), which might point to a different substrate preference of the enzymes in this clade. In fact, OPCL1 and many of these "4CLLs" (e.g., AT5G63380, AT1G20500, AT4G05160) showed higher activity on fatty acids and fatty acid-derived precursors for the phytohormone jasmonic acid than cinnamate-derived compounds in an *in vitro* substrate survey carried out by Kienow et al. (2008). It is thus questionable that the enzymes of this clade act as *bona fide* 4CLs. Homologs to these sequences are found in *Brachypodium distachyon* and *Picea abies*, suggesting an origin in the last common ancestor of seed plants followed by two duplication events with either (a) both taking place in the last common ancestor of angiosperms or (b) one early on in the LCA of seed plants and the second in the LCA of angiosperms. Bootstrap support to include the *Picea abies* sequences in the clade containing AT5G63380 (4CLL9) is however low (bootstrap 54). Each duplication event was followed by independent lineage-specific radiations giving rise to a whole plethora of possible candidates for 4CL, but also a large evolutionary potential with regard to substrate specificity and flexibility. The 4CLL clade of spermatophytes is nested within a larger, lowly supported clade (bootstrap 65) that included sequences from across the land plant tree of life. Here, pronounced and independent expansion occurred in most of the major lineages of land plants, leading to large clades of, for example, proteins of the hornwort *Anthoceros* and the lycophyte *Selaginella*. As noted above, most of the *Selaginella* and some of the *Anthoceros* homologs retained the conserved KQK residues required for the catalytic activity but others

did not (Figure 3), which suggests the presence of species-specific functions. Outside of the entire 4CL-ACOS and 4CLL clade ('Streptophyte 4CL/ACOS/4CLL-likes'; bootstrap 84) clustered various highly divergent ATP-dependent synthetases and ligases that exist throughout the green tree of life including sequences from chlorophytes. None of these synthetases and ligases retained the catalytic triad KQK.

Altogether, our phylogenetic data indicated that a 4CL/ACOS5-like encoding gene was present at the base of Streptophyta. Domain annotation and the analysis of amino acid patterns in the binding pocket and functional sites support this idea. Further, the similarity of these residues between the candidates of streptophyte algal homologs for 4CLs and the sequences of 4CL proteins with high activity on cinnamate derivatives as substrates (see also Costa et al., 2005), indicates that 4CL activity may evolved more than 700 million years ago in streptophytes.

Patchy distribution of CCR-likes in streptophyte algae and massive independent radiations in land plants

En route to the production of different lignin monomers is the NADPH-dependent reduction of the activated acyl-group of the phenylpropanoid backbone molecules. This first step towards an aldehyde functionality is carried out by CCR (Figure 1), which falls into a larger family of NADPH-dependent reductases, including dihydroflavonol reductases (DFRs) and DFR-likes (DFL) (Lacombe et al., 1997; Devic et al., 1999). We previously reported the presence of CCR-like protein sequences in streptophyte algae (de Vries et al. 2017). Since these previous analyses, however, genome data on additional major lineages of land plants and streptophyte algae have become available.

With these new data at hand, covering most major lineages of streptophyte algae and all major lineages of land plants, it is now possible to infer the evolutionary history of CCR-like and DFR-like sequences. We computed a maximum likelihood phylogeny of CCR homologs with a minimum of 220 amino acids that we detected in genomes of 15 land plants, seven streptophyte algae, and five chlorophytes; additionally, we included sequences found in the transcriptomes of the Zygnematophyceae *Spirogyra pratensis* (de Vries et al., 2020), *Zygnema circumcarinatum* (de Vries et al., 2018), and the Coleochaetophyceae *Coleochaete orbicularis* (Ju et al., 2015; Figure 4).

The CCR homologs were distributed over several major clades. This included the CCRL/DRL-like sequences described as TETRAKETIDE α-PYRONE REDUCTASE (TKPR) by Grienenberger et al. (2010), which is an important enzyme acting in the

production of sporopollenin. We recovered a well-supported clade of TKPR1 homologs (bootstrap support of 84) and fully supported clade of TKPR2 homologs. Both clades of TKPRs contained homologs from across the diversity of land plants, bolstering the idea that TKPR1 and TKPR2 split early during plant evolution (Grienenberger et al., 2010)—before the most recent common ancestor of land plants came about. Our domain structure analyses found that TKPR1 possessed the NAD-dependent epimerase/dehydratase (IPR001509) and NAD(P)-binding domain superfamily (IPR036291) domains, which appear to be present in most sequences included in the phylogeny as well as the Tetraketide alpha-pyrone reductase 1 (IPR033267) domain (Figure S4). In contrast, TKPR2 only encoded the first two domains, which is more similar to what is found in the CCR clade.

The *bona fide* CCRs and CCR-likes were spread out over two clades. These two clades were nested in a weakly-supported monophylum (bootstrap support of 53), which was sub-divided into four medium to fully supported clades. A fully supported clade of CCR-likes (including AT4G30470 and AT2G23910) included sequences from across embryophytes; we coined this monophylum CRL-A. *At*CCRL1 and *At*CCRL2 appear to be co-orthologs to one sequence in the Brassicaceae *Capsella grandiflora* (0380s0077.1.p), thus our data suggest a limited distribution of direct orthologs to CCRL1 and CCRL2. Yet the two sequences together fall into a large clade, here coined CRL-B (bootstrap support 73), that contained sequences from all major lineages of tracheophytes. Another medium-supported clade (bootstrap support 73) included the *bona fide* CCRs, CCR1 and CCR2, of *Arabidopsis*. The duplication that resulted in these two CCRs occurred earliest in the common ancestor of all rosids and latest in the common ancestor of Brassicaceae, yet the CCR clade included homologs from across tracheophytes. Many of these lineages appear to have expanded their own repertoire from one CCR1/2 homolog that was present in the last common ancestor of tracheophytes. Interestingly, a clade of divergent monocot CCRs display several replacements in key amino acids involved in the binding of the substrate's phenolic ring, in particular a replacement of non-polar aliphatic Ile to aromatic Tyr/Phe (Figure S5), which might reflect a difference in substrate affinity. The catalytic triad SYK (Figure S5; Pan et al. 2014) is required for enzymatic activity and is overall conserved across CCR/DFR-likes and FLDHs. Interestingly, several CRL-As possess non-conservative amino acid replacements from large phenolic (Tyr/Phe) to smaller (His, Leu, Ser, Gly) amino acids, which might suggest divergent substrate affinities for CRL-As. This is consistent with the domain structure of many CRL-A sequences, which often lack the NAD-dependent epimerase/dehydratase (IPR001509) domain, but possess additional domains such as 3-β-

hydroxysteroid dehydrogenase/isomerase (IPR002225) or match an additional NAD(P)-binding domain (IPR016040; Figure S4). This pattern is only occasionally occurring in sequences from the CCR or other CRL clades.

Finally, there is a third clade with a bootstrap support of 69 that included tracheophyte sequences (forming a sub-clade with a bootstrap value of 90) and a single *Marchantia polymorpha* homolog; we coined this clade CRL-C. Altogether, this suggests that the LCA of all land plants had two homologs of CCRs/CRLs: one CRL-A and one CRL-B, CRL-C or CCR homolog. In vascular plants, duplications have resulted in sub-clades of the CRL-B/CRL-C/CCR homologs.

Within the larger clade that encases the DFRs, DFRLs, CCRs and CCRLs ("Chloroplastida CCR/DFR-likes"; bootstrap 86), one supported clade of Zygnematophyceae (bootstrap support 77) and one supported clade of chlorophyte and streptophyte algae (bootstrap support 76) exists. This points to a distinct DFR/DFRL/CCR/CCRL clade that arose in the ancestor of Zygnematophyceae, yet its placement within the phylogeny other than it belonging to the larger DFR/DFRL/CCR/CCRL clade is uncertain. The divergent pattern of amino acids, which perform substrate and cofactor binding in land plants, suggest that these algal homologs might vary in the substrate and enzymatic activity compared to plant CCR/DFRs (Figure S5) It is, however, certain that within land plants, a pronounced radiation of CCRs occurred.

CADs are present across the green lineage

The second reduction step of the activated acyl-group of the phenylpropanoid backbone and one of the last steps in lignin biosynthesis is the production of phenylpropanoid-derived alcohols from the corresponding aldehydes. An example is the conversion of *p*-coumaroyl aldehyde into *p*-coumaryl alcohol (Kim et al., 2004; Pan et al., 2014). The required reduction is catalyzed by CAD (Figure 1), which is the rate determining enzyme by which lignin is produced (Gross et al., 1973; Mansell et al., 1974). In *Arabidopsis thaliana*, there are at least eleven enzymes belonging to the CAD family. Enzymes of the CAD family have been divided into five major groups, of which group IV was described as monocot-specific (Saballos et al., 2009). In our previous studies many of the chlorophyte and streptophyte potential CAD homologs, identified mostly from transcriptomes and few genomes of algae, were described as CAD-like or CAD group II/III-affiliated (de Vries et al. 2017; de Vries et al. 2020). Sequences clustering with those of CAD group II have been characterized as **s**inapyl **a**lcohol **d**ehydrogenase (SAD) or show predicted structural similarity to SAD

484 enzymes (Guo et al. 2010; de Vries et al. 2017). Additionally, some SADs appear involved
485 not in the synthesis of lignin but defense compounds such as lignans (Suzuki and Umezawa,
486 2007; Guo et al., 2010; Saleem et al., 2010; Barakate et al., 2011), and it may thus be that
487 CAD group II is functionally versatile.

488 Here, we used the 11 canonical CAD sequences to understand the diversity in CAD
489 homologs across streptophytes. This includes also CAD group II sequences, for which
490 homologs in other species may have other substrate specificities and thus are involved in
491 different steps of the phenylpropanoid pathway (Barakat et al., 2009; Guo et al., 2010). We
492 computed a phylogeny of CAD homologs (Figure 5) detected in phylodiverse Chloroplastida.
493 While the resolution of the backbone is weak, we recovered all five CAD groups defined by
494 Saballos et al. (2009). All the CAD groups were resolved as land plant-specific clades of
495 CAD homologs with robust support. Each clade contained a varying set of major land plant
496 lineages (described below); the clades of putative streptophyte algal CAD homologs
497 contained both fewer proteins and fall in-between the five CAD-groups. Hence, this more
498 phylodiverse dataset tells a more complicated evolutionary history for CAD homologs than
499 the less-diverse data from de Vries et al. (2017).

500 Our data suggest that the common ancestor of Zygnematophyceae and land plants
501 may have possessed two CAD-like genes, which was followed by lineage-specific radiations.
502 While it appears—based on the overall topology of the tree—appealing to suggest that one
503 gene gave rise to CAD-group V and the other ancestral gene was the basis of CAD-group I to
504 IV, the low statistical support for the backbone of the phylogeny does not allow to confirm
505 such hypothesis (Figure 5). We can infer that the earliest land plants likely inherited a few (or
506 just one) CAD homolog from their algal progenitors. Most of the radiation of CADs has
507 occurred in plants dwelling on land. Of the canonical CAD group—and based on those part
508 of the topology with good bootstrap support—CAD-group V (containing *At*CAD1) is the
509 only group present in all major land plant lineages (Figure 5). CAD-group I was likely
510 present in the LCA of tracheophytes, as it includes sequences from angiosperms,
511 gymnosperms, ferns, and the lycophyte *Selaginella moellendorffii*. CAD-groups II, III, and
512 IV include only angiosperm sequences—but note that with very weak bootstrap support (53)
513 a sequence from the gymnosperm *Gnetum montanum* associates with the group
514 CADII/III/IV; likely, expansion resulted in the ancestral gene of CAD-group II and III, which
515 diverged into CAD group III genes in angiosperms and after another expansion CAD-group
516 II originated in the LCA of dicots. This is in contrast to de Vries et al. (2017), where the
517 streptophyte algal CAD-like sequences were clustering with CAD group II/III sequences, but

resembles the placement of transcriptomic CAD-like sequences from *Spirogyra pratensis* and *Mougotia* sp. in an already more diverse phylogenetic analysis (de Vries et al., 2020). This is a clear case where including a larger diversity of streptophyte sequences to the analysis enables us to better understand the complexity of the evolution of highly radiated gene families. An analysis of the residues salient to CAD function showed a general conservation of residues involved in zinc and $NADP^+$ binding across Chloroplastida CAD-like sequences (cf. Youn et al. 2006), whereas the residues in the binding pocket are generally less conserved in these sequences (Figure 5). However, within a canonical CAD-groups or a CAD-like clade we see a general conservation of residues in the binding pocket. The binding pockets of each CAD-like clade appear different than those of the canonical CAD-groups. Domain analyses suggest that some of the CAD-like sequences of streptophyte algae do not encode all of the five domains present in most canonical CAD sequences, yet many of the Zygnematophyceaen CAD-like sequences encode all of these five domains. Similar patterns emerge for other CAD-like sequences (Figure S6).

All CAD groups are shaped by multiple lineage-specific duplications and losses. This hampers the inference of function and substrate specificity of the diverse CAD-like sequences. Additionally, several lineages have originated a variety of CAD homologs that are not yet designated to previous groups. In the absence of functional data, we however will not give them a group designation but rather designate them as lineage-specific CAD-homologs of unknown SAD or CAD function. CAD-like homologs found in streptophyte algae show similar functional residues to other CAD-likes of land plants (not included in any of the five clades of canonical CADs), including at the binding pockets—the pattern of residues is similar to what is observed for other land plant sequences in-between the *bona fide* CAD groups. The *bona fide* CAD groups showed a more homogeneous pattern of functional residues.

Overall, both the topology of the phylogenetic tree and the conservation of key residues point to (i) a deep evolutionary origin of CAD homologs and (ii) independent radiations of CADs—not only in land plants but also in streptophyte algae.

Massive independent radiations of acyltransferases and scattered candidates in streptophyte algae

A versatile group of enzymes that are important for the processes leading up to the lignins but also compounds with antioxidant and antimicrobial properties are the BAHD acyltransferases (named after the first enzymes characterized for this family BEAT, AHCT, HCBT, and DAT;

see also D'Auria, 2006). Most prominent among these are the versatile hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferases (HCT) (Eudes et al., 2016). Recently, Kriegshauser et al. (2020) reported on the functional conservation of the HCT homologs found in bryophytes with those of seed plants.

We computed a phylogeny including phylodiverse acyltransferases (Figure S7). Overall, the topology of the tree corroborates the findings of Kriegshauser et al. (2020) that a clear clade of HCT proteins first emerged in land plants—streptophyte algal sequences were few, divergent from HCT, and scattered over the tree without clear affinity to characterized acyltransferases. Without functional analyses, there is no solid foundation for predicting their function—making them exciting candidates for future studies. What our data, however, clearly reveal is that the last common ancestor of land plants likely had an expanded repertoire of acyltransferases that further diversified during the radiation of plants on land.

A clear clade of C3Hs is limited to land plants

At several steps of the phenylpropanoid pathway, enzymes belonging to the cytochrome P450 family CYP98A, within the large CYP71 clan (Nelson and Werck-Reichhart, 2011), carry out hydroxylations of *p*-coumarate-derived compounds (such as *p*-coumaroyl esters; for function of C3H see Figure 1). In land plants, these hydroxylations are important for the production of lignins, lignans, volatile phenylpropanoids, coumarins, and many more phenylpropanoid-derived compounds. Previously, de Vries et al. (2017) described the detection of C3H in all land plants and one putative C3H ortholog in *Klebsormidium nitens*. Now, with genomic gaps in the streptophyte tree of life filled, we revisited the distribution of C3H.

The number C3H homologs detected in genomes of 15 land plants, 7 streptophyte algae, and five chlorophytes and in the transcriptomes of *Spirogyra pratensis* (de Vries et al., 2020), *Zygnema circumcarinatum* (de Vries et al., 2018), and *Coleochaete orbicularis* (Ju et al., 2015) varied strongly between lineages. When sampling the sequences via BLAST (with AT2G40890 as query sequence), we thus included either (a) all sequences that had a bit score of at least 100, or (b) the top five hits. We aligned all sequences, cropped them to the alignable region and computed a maximum likelihood phylogeny (Figure 6).

A clade of CYP98A included sequences from all major lineages of land plants. This suggests that at least one CYP98A sequence was present in the LCA of all land plants. Based on the lineages included here, it appears that from this single copy gene, radiations occurred in the dicot lineages and *Amborella trichocarpa*. A single copy remained in the bryophytes, lycophytes, ferns, gymnosperms, and monocots. The clade of CYP98A8 and CYP98A9 was

586 in our dataset limited to the Brassicaceae *Arabidopsis thaliana* and *Capsella grandiflora*
587 (bootstrap value 100). These two enzymes function in a route derived from the
588 phenylpropanoid pathway and are involved in the formation of N^1,N^5-di(hydroxyferuloyl)-
589 N^{10}-sinapoylspermidine (Matsuno et al., 2009). The CYP98A8/9 clade falls into the larger
590 CYP98A clade (bootstrap support 100) together with the C3H sequences, suggesting that
591 they are the closest paralogs of C3H in *Arabidopsis thaliana* and *Capsella grandiflora.*
592 Analyses of their substrate recognition sites (SRS; cf. Rupasinghe et al. 2003) support the
593 divergent functional roles between CYP98A8/9 and canonical C3H. In particular, the first
594 two SRS (SRS1 and SRS2) show various distinct amino acid differences between the
595 CYP98A8/9 clade and the C3H clade (Figure S8). SRS1 and 2 are predicted to be involved in
596 binding of the substrate tails, which exhibit strong variation in size (Rupasinghe et al. 2003),
597 and thus may be critical for the substrate specificity of these paralogs. Indeed, SRS1 and 2 are
598 the two SRSs showing the strongest variation across the entire phylogeny, including also
599 other CYP450 subfamilies (Figure 6), corroborating this hypothesis.
600 We recovered additional land plant-specific clades of CYP450 enzymes (Figure 6)
601 such as one containing TRANPARENT TESTA7 (TT7)-like sequences (bootstrap support
602 92); TT7 is a cytochrome P450 75B enzyme that is involved in flavonoid biosynthesis
603 (Tanaka et al., 1997; Schoenbohm et al., 2000). The BLAST search used for sampling C3H
604 homologs further recovered *At*CYP71B and *At*CYP76C members. In our phylogenetic
605 analysis, we inferred that *At*CYP71B34 and *At*CYP71B35 were likely born out of an
606 *Arabidopsis*-specific duplication, while CYP71B enzymes in general are present across
607 angiosperms. In contrast, *At*CYP76C1 and *At*CYP76C4 appear to have originated prior to the
608 split of *Arabidopsis thaliana* and *Capsella grandiflora*; a CYP76C4 ortholog appears to have
609 been lost in the latter plant species. The CYP76C clade may also be represented in other
610 species outside of angiosperms, because we find a sequence from *Gnetum montanum*
611 clustering with these sequences with a bootstrap support of 87. The rather long branch
612 warrants attention and would require a more CYP76C focused phylogenetic analysis, which
613 is not the point of this paper.
614 The algal sequences showed strong divergence to the C3Hs, forming only a larger
615 (fully supported) streptophyte-specific clade with all the recovered and functionally diverse
616 CYP450 enzymes (Figure 6). Domain structures of these streptophyte algal sequences is the
617 same to what is observed for the C3H homologs, but is in general conserved across the
618 phylogeny independent of the CYP450 subfamily assignment (Figure S9). Only few scattered
619 exceptions occur. Given the low support of most of the tree backbone, the role(s) of the

620 streptophyte algal homologs detected here remains elusive. Most of the streptophyte algae
621 show independent radiations of their CYP450 enzymes complicating functional predictions
622 even further. Thus, while there are interesting CYP450 candidates in streptophyte algae, a
623 clear C3H clade likely first arose early during the evolution of embryophytes.
624
625 **Monoacylglycerol lipases: multiple early radiations, independent subfunctionalization,**
626 **and the origin of CSE**
627 The conversion of caffeoyl-5-*O*-shikimate to caffeic acid may be a step along the
628 biosynthetic routes that lead to the production of G- and S-lignins in certain vascular plants
629 (Figure 1). The enzyme responsible for this step is caffeoyl-5-*O*-shikimate esterase (CSE).
630 CSE converts caffeoyl-5-*O*-shikimate to caffeic acid and was hypothesized to act together
631 with 4CL/ACOS5 to circumvent the catalysis of caffeoyl-5-*O*-shikimate to caffeoyl-CoA via
632 HCT (Vanholme et al. 2013). The latter pathway was proposed for tobacco by Hoffmann et
633 al. (2003) and confirmed for *Arabidopsis thaliana in vitro* by Vanholme and colleagues
634 (2013). Yet, based on *cse* mutants in *Arabidopsis thaliana*, they suggested that synthesis of
635 caffeoyl-CoA is more likely to occur via CSE and 4CL/ACOS5 than directly from caffeoyl
636 shikimate by HCT *in planta*. That said, in the model grasses *Brachypodium distachyon* and
637 *Zea mays* no CSE orthologs are present and crude extracts from these species show little
638 signs for the characteristic esterase activity (Ha et al., 2016). On the other hand, non-vascular
639 plants such as the model system *Physcomitrium patens* possess homologs of these enzymes
640 (Renault et al., 2017a), which suggests a secondary loss of CSE in the respective monocots.
641 CSE belongs to the family of putative monoacylglycerol lipases (MAGL). MAGLs are found
642 across eukaryotes and functional analyses in human, yeast and *Arabidopsis* have shown that
643 they possess monoacylglycerol lipase activity (Labar et al., 2010; Aschauer et al., 2016; Kim
644 et al., 2016). In contrast to other MAGLs of *Arabidopsis thaliana*, *At*CSE (MAGL3) was
645 found to exhibit no hydrolytic activity on monoacylglycerols (MAGs) as a substrate (Kim et
646 al., 2016)—which applies to other enzymes of *Arabidopsis thaliana* that belong to the family
647 of MAGLs, too. Indeed, out of the 16 MAGLs that Kim and colleagues (2016) tested, only
648 MAGL6 and 8 showed high activity on MAG as substrate. Given the functional diversity in
649 MAGLs (Kim et al. 2016) and the unequal distribution of caffeoyl-5-*O*-shikimate across
650 embryophytes, functional analyses are required to fully understand how easily MAGLs can
651 lose or gain their MAGL activity. Yet, phylogenetic analyses can pinpoint the diversity of the
652 family across the green lineage.

Here we use phylogenetic analysis to pinpoint the distributions of the diverse MAGL families, including CSE across the green lineage. In total, we recovered all 16 MAGL sequences of *Arabidopsis thaliana* in the similarity search; using maximum likelihood phylogenetics, we recovered clades for all the 16 MAGLs; some MAGL clades are widely distributed throughout streptophytes, while others appear to have originated in embryophytes, where they have again undergone lineage-specific expansions. Essentially, we recovered two large clades: one restricted to streptophytes, containing homologs of MAGL2, 4 and 13; the other has representation in chlorophytes as well and includes homologs of MAGL1, 3, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15 and 16 (Figure 7).

Focusing on the MAGL2/4/13 clade first, we observe that MAGL13 has representation in angiosperms, gymnosperms and ferns, suggesting its origin to be in the LCA of tracheophytes, while MAGL2 and 4 came from a duplication event before the split between *Arabidopsis* and *Capsella*. However, MAGL2/4 orthologs are present in other species including *Picea abies*, suggesting that the common ancestor of seed plants possessed a *MAGL2/4-like* and a *MAGL13* gene. Forming a clade with MAGL2/4/13 are lycophyte, bryophyte and streptophyte algal sequences, which in general branch in an order expected based on their species phylogeny (although within species duplication events have occurred). This suggests that already at the base of streptophytes a *MAGL2/4/13-like* gene was present.

In the second large clade that includes also chlorophyte sequences, we find the clade containing the CSE/MAGL3 orthologs. This clade includes sequences from both vascular and non-vascular plants, pointing to an origin of CSE in the last common ancestor of land plants. This adds support for a secondary loss in those monocots without a CSE ortholog. Despite the origin of CSE in the common ancestor of land plants and a clear CSE ortholog in *Physcomitrium patens* (3c19_14430V3.1.p), the substrate of CSE, caffeoyl-5-*O*-shikimate, was not detected in crude extracts of the moss (Renault et al., 2017a). The HCT-based reaction leading to caffeoyl-CoA has been confirmed *in vitro* using moss HCT (Kriegshauser et al. 2021). Hence, the CSE homologs of *Physcomitrium patens* may have another function. Indeed, the atypical function of *At*CSE, together with the lack of other MAGL family members, to act on MAGs (Kim et al. 2016), suggests that the functional spectrum of the MAGL family is not very limited in land plants.

Members of the MAGL family share several conserved motifs across diverse eukaryotes. One such motif (amino acid positions 132-141 in MAGL6 and 167-176 in CSE/MAGL3) is situated in a region likely involved in substrate binding based on the crystal structure of human MAGL (Labar et al., 2010). Within this motif a leucine in position

number four is found in a diverse set of 249 mammal and Sauria MAGLs investigated here (Figure 7, inset) and most plant MAGLS including MAGL6 and 8. This is followed by another hydrophobic amino acid (I in mammals/Sauria and V or L in most plant MAGLs). It is striking that exactly these highly conserved amino acids are changed to a phenylalanine and a serine in the Arabidopsis CSE and some homologs from other species. These changes from two very hydrophobic amino acids to an aromatic and a hydrophilic one could be one of the reasons for a change in substrate specificity from a substrate with a hydrophobic acyl chain to a more hydrophilic substrate with aromatic properties. Based on this hypothesis, CSEs would be restricted to some of the members of this clade that, however, stem from across the diversity of land plants.

MAGL1 appears to have originated prior to the split of angio- and gymnosperms, while specific MAGL14 and 16 orthologs likely arose after the split of asterids and rosids, but a MAGL14/16 ortholog was likely present in the LCA of land plants. MAGL15, like MAGL1, originated prior to the split of gymno- and angiosperms, and MAGL5 appeared to come from a duplication of MAGL15 later on possibly in the ancestors of dicots. Interestingly, a *MAGL5/15-like* sequence was already encoded in the genome of the ancestor of streptophytes. The same is true for a *MAGL6/7/8/9/10/11/12-like* gene, which similar to the *MAGL5/15-like* genes shows independent paths of radiation in streptophyte algae and land plants. MAGL6,7,8,10 and 11 are only present in the here included Brassicaceae, while a *MAGL6/7/8/10/11-like* gene was already present in the common ancestor of spermatophytes. The same evolutionary history describes the scenario under which MAGL9 and 12 originated.

Finding the MAGL6/7/8/10/11 subclade specifically expanded in seed plants is noteworthy. At least *At*MAGL8 localizes to lipid droplets (Kim et al., 2016), which are structures found in various photosynthetic eukaryotes but are well-known from seeds. Thus, expansion of this clade might be a read-out of spermatophyte-specific additions to the ancient set of proteins relevant to LD formation and function (see de Vries and Ischebeck, 2020).

All in all, the MAGLs have experienced an early radiation in streptophytes. Given that even the Arabidopsis MAGLs without detectable activity on MAG (see Kim et al., 2016) do not form one single monophylum, it is conceivable that subfunctionalization of members of the MAGL family occurred multiple times independently. This may likewise be true for all independent expansions of MAGL-encoding genes observed in any other species included here. The versatility in functional evolution of MAGLs makes it difficult to make robust predictions of putative MAGL functions.

721

722 **COMT: convergence and complexity**

723 In angiosperms, ferulate 5-hydroxylase (F5H) and caffeate *O*-methyltransferase (COMT)
724 carry out important catalytic steps along the route from *p*-coumaroyl-CoA to S-lignin. COMT
725 catalyzes the methylation of caffeic acid or 5-hydroxyferulic acid, the product formed by
726 F5H. Like C3H and C4H, F5H belongs to the large CYP450 clan 71 (Nelson and Werck-
727 Reichhart, 2011). The function of F5H evolved at least twice in P450 enzymes, once in the
728 ancestor of angiosperms and once in the ancestor of lycophytes (Weng et al., 2008; Weng and
729 Chapple, 2010).

730 For angiosperm F5H, no clear putative orthologs were found outside of flowering
731 plants and likewise no clear orthologs were found for the lycophyte F5H (i.e. "SmF5H"),
732 which forms a separate clade from the angiosperm F5H sequences (Figure S10). This
733 corroborates previous results (de Vries et al., 2017) and is in agreement with the hypothesis
734 that F5H function evolved at least twice in the evolution of land plants (Weng et al., 2008).
735 Additionally, the average pairwise identity of the F5H homologs was quite low (18.5%)—
736 hampering robust phylogenetic analyses. We thus did not further delve into the evolution of
737 F5H. COMT however caught our attention.

738 The lycophyte *Selaginella moellendorffii* not only uses a genetically distant F5H
739 enzyme; the same appears to be true for COMT (Weng et al., 2011). This highlights a
740 promiscuity for substrate specificity and activity in P450 enzymes that is yet to be discovered
741 and mere orthology analyses can only go so far as to discover putative candidates. Using
742 phylogenetics, we explored the diversity of methyltransferases by screening for sequences
743 homologous to COMT/OMT1 of *Arabidopsis thaliana* across our phylodiverse dataset. This
744 approach identified not only clear orthologs but can also serve as a backbone to map relevant
745 mutations facilitating in functionally convergence in this group of enzymes and by that may
746 highlight possible candidates for *in vivo* and *in vitro* studies.

747 For most clades of land plant methyltransferases, based on the here recovered
748 topology, predicting a putative function was not straightforward. This applied even more so
749 to the homologs of COMT/OMT found in chlorophyte and streptophyte algae. We recovered
750 a clade of methyltransferases that included chlorophyte and streptophyte algae as well as
751 diverse land plant sequences (coined 'Chloroplastida OMT' in Figure 8); among these
752 clustered *Arabidopsis thaliana* proteins such as COMT, indole glucosinolate
753 methyltransferases (IGMT), and nicotinate *N*-methyltransferase (NANMT; see Li et al.,
754 2017)—hence different methyltransferases that act on a range of aromatic compounds. What

this means for the presence of a putative COMT in algae is obscure. However, it corroborates the previously observed patchy detection of OMT1 across the green lineage based on a reciprocal BLASTp searches (de Vries et al., 2017). Clear orthologs of *At*COMT were only detected for a few angiosperms, notably not including any monocot sequence that we included in our dataset. Of all methyltransferases in our dataset, only NANMT formed a clade of clear orthologs that included more than one major lineage of land plants by encasing sequences from angiosperms and *Picea abies* (bootstrap support 87). All other orthogroups appear, like COMT, to be restricted to only a few of the included angiosperm lineages.

The lycophyte *Selaginella moellendorffii* has a COMT that is distantly related to OMT of angiosperms. It appears to have acquired its OMT activity through convergent evolution and was coined *Sm*COMT (Weng et al., 2011). In agreement with this, *Sm*COMT did not cluster with the *At*OMT1 sequence in our analyses. Instead, it forms its own (weak) clade with only one other sequence from *Selaginella moellendorffii* (bootstrap support 63). The other *Sm*COMT-like sequences (described in Weng et al., 2011) were distributed over the phylogeny and appear to be specific to *Selaginella moellendorffii*. Nonetheless, this pattern highlights a certain versatility in the evolutionary history of substrate specificity of *O*-methyltransferases in land plants. This appears to be only logical, noting the large lineage-specific expansions in the larger clade of *O*-methyltransferases that encompasses all *O*-methyltransferases from *Arabidopsis thaliana*—that is COMT, IGMTs, NANMT and *N*-acetylserotonin *O*-methyltransferase (ASMT; for more on this enzyme see, e.g., Tan et al., 2012; Byeon et al., 2016) (bootstrap-support 99). Within this clade fall also algal sequences from chlorophytes and streptophyte algae. Their position within the clade is undetermined due to low bootstrap support. These sequences appear highly divergent, many of them cluster with rather long branches. Yet, some of the sequences from our previous analysis found a reciprocal BLASTp hit to *At*COMT, including *Klebesormidium nitens* 00158_0100v1.1 that clusters in a fully supported clade of *Klebesormidium* paralogs; these sequences are promising candidates to explore caffeic acid *O*-methyltransferase activity. Indeed, when we modelled the tertiary structure of *Klebesormidium nitens* 00158_0100v1.1 and *Mesotaenium endlicherianum* ME000591S08520 using I-TASSER (Zhang, 2008), we recovered *Medicago sativa* and *Lolium perenne* COMT as its closest structural analogs (1KYZ; Zubieta et al., 2002; 3P9C; Louie et al., 2010; TM-scores 0.865 and 0.956, respectively).

Like the land plant COMTs, also the algal COMT-likes appear to have undergone independent radiations in this large gene family. Given the observed convergent evolution of

COMT activity in *Selaginella moellendorffii*, the question of whether there is COMT activity across Streptophyta remains wide open.

To gain a first insight into whether COMT activity can be expected from other streptophyte lineages, we investigated the conservation of residues relevant for the function of COMT, including those that form the substrate binding pocket (Figure 8). The functional residues were identified from COMT of *Lolium perenne* (Louie et al., 2010). Across our phylogeny, these residues differ between the clades of canonical ASMT, NANMT, COMT and IGMT, while they are conserved within them (Figure 8). The binding pocket of *At*OMT1 and its orthologs consist of the amino acid pattern MSNGGG, whereas the pattern for the residues important for the function of the enzyme is HDE. While HDE appears conserved across the majority of sequences analyzed here, independent of the specific function of the enzyme (e.g. ASMT, IGMTs and COMT all have the pattern HDE), the binding pocket is highly variable among the functionally characterized enzymes. This suggests that the reaction-determining residues are those that form the binding pocket and not those that are catalytically important. This seems logical given that all these enzymes catalyze similar types of reactions. The triple G in the binding pocket is also far more conserved across the entire phylogeny, with only few exceptions occurring, while the first three residues are highly variable. Indeed, the COMT-specific MSN motif is not present in the functionally characterized COMT from *Selaginella moellendorfii*, rather it is MTN, which however is a change between similar amino acids (Ser to Thr). Apart from the COMT orthologs and *Sm*COMT sequence no other sequences from any other lineage encode the binding pocket pattern M(S/T)NGGG, suggesting that none has a canonical preference for binding 5-hydroxyconiferalaldehyde. Yet, several homologs—including those of streptophyte algae—would have the ability to catalyze the reaction based on the conservation of the residue pattern HDE. What is more, all the sequences that could not be properly identified as orthologs to ASMT, NANMT, COMT and IGMT (with the exception of *Sm*COMT) show no similarity in their first three residues of the binding pocket to either of these enzyme families. This would suggest that the enzymes from most streptophyte lineages included in this analysis use different substrates than those functionally characterized in *Arabidopsis thaliana.*

CCoAOMT-like sequences emerged in Phragmoplastophyta

Within the phenylpropanoid pathway, caffeoyl-CoA *O*-methyltransferase (CCoAOMT) and most of its homologs are the enzymes that catalyze the first committed step to many of at

least two types of lignin (S- and G-lignin). These enzymes methylate caffeoyl-CoA and thus give rise to feruloyl-CoA (Ye et al., 1994; Ye and Varner, 1995; Martz et al., 1998; Do et al., 2007; Vanholme et al., 2012); in the past, it was also proposed that after the conversion to 5'-hydroxy-feruloyl-CoA, CCoAOMT can methylate this compound to produce Sinapoyl-CoA (Maury et al., 1999; Ferrer et al., 2005). One of the paralogs that exist in *Arabidopsis thaliana* (tapetum-specific *O*-methyltransferase [TSM1], AT1G67990) however, shows activity towards a coniferyl derivative that is formed at N^{10} by F5H starting from N^{1},N^{5},N^{10}-tris-(hydroxyferuloyl) spermidine (Fellenberg et al., 2009); TSM1 catalyzes the production of N^{1},N^{5}-bis-(hydroyferuloyl)-N^{10}-synapoylspermidine (Fellenberg et al., 2009). Thus, CCoAOMTs appear to be versatile in their substrate specificity and can act on different steps in the phenylpropanoid pathway.

Here we used phylogenetics to disentangle the distribution of these enzymes across the green lineage. A duplication gave rise to the genes encoding the functionally divergent enzymes *At*CCoAMT and *At*TSM1 (bootstrap support 77; Figure 9). These two methyltransferases are embedded in a larger clade containing the other CCoAOMT enzymes *At*CCoAOMT1, *At*CCoAOMT7 and *At*CCoAOMT-like (AT1G24735). The latter appears to be specific to the Brassicaceae included in this dataset, while homologs of *At*CCoAOMT7 occur across dicots and were detected in *Amborella trichopoda*, but were absent from the included monocots. Only *At*CCoAOMT1 had a wider distribution. Its cluster (bootstrap support 79) contains angiosperms, gymnosperms, lycophytes, ferns and bryophytes, excluding the sequenced hornworts from the genus *Anthoceros* (Figure 9). Assuming a monophyly of Bryophyta (Puttick et al., 2018), this suggests a loss of CCoAOMT1 in at least the sequenced *Anthoceros* species, and that CCoAOMT1 was present in the common ancestor of land plants. Lineage-specific duplications of CCoAOMT1 appear to have happened, indicated by the expansions seen in tobacco, spruce, the lycophyte *Selaginella moellendorffii* and the water fern *Azolla filiculoides*. The expanded repertoire of sequences in monocots and *Gnetum* indicate additional lineage-specific duplications outside of the CCoAOMT clade. The case of TSM1 suggests that neo-functionalization can easily occur within this type of methyltransferases. We noted that the residues involved in substrate binding (Ferrer et al. 2005) are identical in *At*CCoAMT and *At*TSM1 (Figure 9). A possible explanation might be that the make-up of binding pocket allows for a certain versatility in substrates. Given these observations, the paralogs within this and the other CCoAOMT clades cannot be assumed to hold the function of CCoAOMTs. Likewise, it cannot be ruled out that their LCA may have had this function. As sister to the methyltransferase clade, including the CCoAOMT

857 homologs lies a cluster of genes encoding putative candidates for streptophyte algal
858 CCoAOMTs. These were limited to representatives of the two streptophyte algal lineages
859 closest to land plants: the Coleochateophyceae *Coleochaete scutata* and *Coleochaete*
860 *orbicularis* as well as the Zygnematophyceae *Spirogloea muscicola*. These algal sequences
861 have the same domain structure as the majority of all *S*-adenosyl-*L*-methionine (SAM)-
862 dependent methyltransferases included in the phylogeny (Figure S11). Only single sequences,
863 scattered across the phylogeny and diversity of species included here, vary in their domain
864 structure showing a loss of a domain loss, a gain of an additional domain, or both. The
865 analyses of the specific functional residues gave more insights into the streptophyte algal
866 sequences within the clade of SAM-dependent methyltransferases in the peripheral routes of
867 the phenylpropanoid pathway. These algal sequences maintain the residues involved in ion
868 and cofactor binding, but differ strongly in the substrate binding site (cf. Ferrer et al. 2005;
869 Figure 9).

870 Altogether, it appears that the family of CCoAOMT-like proteins has its origin in
871 Phragmoplastophyta. Clarifying the function of the putative CCoAOMT-like enzymes in
872 Coleochateophyceae and Zygnematophyceae has the potential to shed light on a
873 synapomorphy with physiological relevance.

874

875 **Conclusion**

876 All genes for enzymes that act in early steps in the chassis of the phenylpropanoid pathway
877 investigated here (Figure 1) can be traced back to the LCA of land plants with the exception
878 of COMT (Figure 10); most of these can even be traced back to some ancestor that land
879 plants shared with streptophyte algae. While most of our knowledge on how these genes
880 work comes from angiosperms, this does not capture the sequence diversity in enzymes—and
881 it underpins the versatility in producing specialized metabolites.

882 Our data pinpoint that most of the enzymes have undergone massive lineage-specific
883 expansions. A lineage-specific expansion is palpable even despite the fact that sampling of
884 sequences across the Streptophyta is still strongly biased towards seed plants. These data
885 offer a framework for pinpointing those candidate genes/enzymes that are bound to shed light
886 on the evolution of key enzymatic steps—and novel ones. Such work is exemplified by
887 studies on *Selaginella* or bryophyte model systems such as *Physcomitrium patens*.
888 Characterizing enzymes that are even more divergent from what we know from angiosperms
889 should yield surprising insights and novel routes in this bountiful pathway.

890

MATERIAL AND METHODS

Dataset of protein sequences and screening for homologs

We downloaded protein data from: (a) genomes of fifteen land plants: *Anthoceros agrestis* as well as *Anthoceros punctatus* (Li et al., 2020), *Amborella trichopoda* (*Amborella* Genome Project, 2013), *Arabidopsis thaliana* (Lamesch et al., 2010), *Azolla filiculoides* (Li et al., 2018), *Brachypodium distachyon* (The International Brachypodium Initiative, 2010), *Capsella grandiflora* (Slotte et al., 2013), *Gnetum montanum* (Wan et al., 2018), *Marchantia polymorpha* (Bowman et al., 2017), *Nicotiana tabacum* (Sierro et al., 2014), *Oryza sativa* (Ouyang et al., 2007), *Picea abies* (Nystedt et al., 2013), *Physcomitrium patens* (Lang et al., 2018), *Salvinia cucullata* (Li et al., 2018), *Selaginella moellendorffii* (Banks et al., 2011), and *Theobroma cacao* (Argout et al., 2011); (b) the genomes of seven streptophyte algae: *Chlorokybus atmophyticus* (Wang et al., 2020), *Chara braunii* (Nishiyama et al., 2018), *Klebsormidium nitens* (Hori et al., 2014), *Mesotaenium endlicherianum* (Cheng et al., 2019), *Mesostigma viride* (Wang et al., 2020), *Penium margaritaceum* (Jiao et al., 2020), *Spirogloea muscicola* (Cheng et al., 2019); (c) the genomes of five chlorophytes: *Bathycoccus prasinos* (Moreau et al., 2012), *Chlamydomonas reinhardtii* (Merchant et al., 2007), *Coccomyxa subellipsoidea* (Blanc et al., 2012), *Micromonas pusilla*, *Micromonas* sp. (Worden et al., 2009), *Ostreococcus lucimarinus* (Palenik et al., 2007), *Ulva mutabilis* (De Clerck et al., 2018), *Volvox carteri* (Prochnik et al., 2010). Additionally, we included sequences found in the transcriptomes of *Spirogyra pratensis* (de Vries et al., 2020), *Zygnema circumcarinatum* (de Vries et al., 2018), and *Coleochaete orbicularis* (Ju et al., 2015).

For each of the protein families we investigated here, the representative *Arabidopsis thaliana* protein was used as a query sequence for a BLASTp against this dataset. Initially, we considered all homologs recovered at a cutoff level of 10^{-7}. However, due to the large size of the protein families (i.e. high number of well-supported homologs obtained), refinement of the datasets was carried out as described in the individual sections for these enzymes in the Results and Discussion section.

Alignments, phylogenetic analysis, and primary sequence analysis

Using the homologs detected based on the above described BLASTp search for a given enzyme, we generated alignments using MAFFT v7.453 (Katoh and Standley, 2013) with a L-INS-I approach. Alignments were cropped, if necessary, to retain conserved domains that were alignable for all homologs; alignments are provided in Supplemental Datasets S1 to S11. We computed maximum likelihood phylogenies using IQ-TREE multicore version 1.5.5

(Nguyen et al., 2015), with 100 bootstrap replicates. To determine the best model, we used ModelFinder (Kalyaanamoorthy et al., 2017) and picked the best models based on the Bayesian Information Criterion. The best models were: LG+G4 (Le and Gascuel, 2008) for 4CL, CCR, CCoAOMT; LG+I+G4 for PAL, CAD, MAGL/CSE, COMT, and for the preliminary phylogeny of 4CL; LG+F+I+G4 for C4H, F5H, and C3H; WAG+F+G4 (Whelan and Goldman, 2001) for HCT.

Protein structure prediction was carried out using the sequences as input in the online Iterative Threading ASSEmbly Refinement (I-TASSER; Zhang, 2008; Yang et al., 2015). Functional residue analyses were based on published structural analyses (Rupasinghe et al., 2003; Ferrer et al., 2005, Youn et al., 2006; Hu et al., 2010; Pan et al., 2014) and alignments were viewed with SeaView v.4 (Gouy et al., 2009) and plotted with ETE3 (Huerta-Cepas et al., 2016).

Protein domain predictions

Protein domains for all protein sequences for the enzyme families 4CL, CCR, C3H, CAD and CCoAOMT included in the phylogenies were predicted using InterProScan version 5.47-82.0 (Jones et al., 2014). The presence or absence of protein domains were mapped onto the phylogenies of the afore mentioned gene families as presence/absence heatmaps, which were visualized using iTOL v6 (Letunic and Bork, 2019).

ACKNOWLEDGEMENTS

J.M.R.F.-J. is grateful for being supported by the Ph.D. program "Microbiology and Biochemistry" within the framework of the “Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular Biosciences” (GGNB) at the University of Goettingen; A.D.A. is grateful for being supported through the International Max Planck Research School (IMPRS) for Genome Science. J.d.V. thanks the European Research Council for funding under the European Union’s Horizon 2020 research and innovation programme (Grant Agreement No. 852725; ERC-StG “TerreStriAL”). M.P., I.F., and J.d.V. are grateful for support through the German Research Foundation (DFG) within the framework of the Priority Programme “MAdLand – Molecular Adaptation to Land: Plant Evolution to Change” (SPP 2237; VR 132/4-1; PE 360/37-1; FE 446/14-1)

FIGURE LEGENDS

Figure 1. Enzymes involved in the biosynthesis of phenylpropanoid-derived compounds investigated here. A simplified schematic of the phenylpropanoid pathway and its routes to different derivatives are shown. Boxes indicate enzyme families, which are mentioned above each box and color-coded. Their coloration is the same as in Figure 10. Dotted lines indicate putative steps in the pathway.

Figure 2. A phylogenetic framework for the origin of streptophyte PAL. PAL and HAL homologs were screened for in fifteen land plant, seven streptophyte algae, and five chlorophytes. Among Chloroplastida, PAL homologs were only recovered from genomes of land plants and the streptophyte algae *Klebsormidium nitens* and *Chara braunii*. From all detected homologs, a rooted maximum likelihood phylogeny was computed using LG+I+G4 as model for protein evolution (chosen according to BIC). 100 bootstrap replicates were computed; only bootstrap values ≥50 are shown and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogenies.

Figure 3. 4CL homologs occur across Streptophyta. 4CL homologs were sampled from protein data of nine land plant, seven streptophyte algal and five chlorophyte algal genomes. Only protein sequences with a minimum length of 400 and a maximum length of 1150 amino acids were included. From all detected homologs, a rooted maximum likelihood phylogeny was computed using LG+G4 as model for protein evolution (chosen according to BIC). 100 bootstrap replicates were computed; only bootstrap values ≥50 are shown and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogenies. On the right we show key residues for substrate binding and function of canonical 4CL as reported by Hu et al. (2010).

Figure 4. The complex evolutionary history of CCR in Chloroplastida. CCR homologs of a minimum of 220 amino acids were sampled from protein data of 15 land plant, seven streptophyte algal, and five chlorophyte algal genomes; additionally, we included sequences found in the transcriptomes of *Spirogyra pratensis* (de Vries et al., 2020), *Zygnema circumcarinatum* (de Vries et al., 2018), and *Coleochaete orbicularis* (Ju et al., 2015). From all detected homologs, an unrooted maximum likelihood phylogeny was computed using LG+G4 as model for protein evolution (chosen according to BIC). 100 bootstrap replicates

were computed; only bootstrap values ≥50 are shown and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogenies.

Figure 5. Phylogenetic analysis highlights CAD candidates across Chloroplastida. CAD homologs were sampled from protein data from fifteen land plant, seven streptophyte algal, and five chlorophyte algal genomes as well as sequences found in the transcriptomes of *Spirogyra pratensis* (de Vries et al., 2020), *Coleochaete scutata* and *Zygnema circumcarinatum* (de Vries et al., 2018), and *C. orbicularis* (Ju et al., 2015). From all detected homologs, an unrooted maximum likelihood phylogeny was computed using LG+I+G4 as model for protein evolution (chosen according to BIC). 100 bootstrap replicates were computed; only bootstrap values ≥50 are shown and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogenies. The five groups of CADs were named in accordance with Saballos et al. (2009). Next to the sequence labels residues from the binding pocket, $NADP^+$- and Zn^{2+} binding are shown—based on Youn et al. (2016).

Figure 6. A clade of C3H orthologs originated at the base of land plants. C3H homologs were sampled from protein data of genomes of fifteen land plants, seven streptophyte algae, and five chlorophytes; additionally, sequences found in the transcriptomes of *Spirogyra pratensis* (de Vries et al., 2020), *Zygnema circumcarinatum* (de Vries et al., 2018), and *Coleochaete orbicularis* (Ju et al., 2015) were included. For downstream analyses, we used either (a) all sequences that had a bit score of at least 100 or (b) the top five hits. We aligned all sequences, cropped them to the alignable region and computed an unrooted maximum likelihood phylogeny was computed using LG+F+I+G4 as model for protein evolution (chosen according to BIC). 100 bootstrap replicates were computed; only bootstrap values ≥50 are shown and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogenies. Two large clades that contained only (a) *Anthoceros* and (b) chlorophyte and streptophyte algal sequences were collapsed; the full tree is shown in Figure S12.

Figure 7. The occurrence of MAGLs across diverse Streptophyta and a phylogenetic framework for the deep evolutionary roots of CSE. MAGL/CSE homologs were sampled from protein data from 15 land plant, seven streptophyte algal, and five chlorophyte algal

genomes as well as sequences found in the transcriptomes of *Spirogyra pratensis* (de Vries et al., 2020), *Coleochaete scutata* and *Zygnema circumcarinatum* (de Vries et al., 2018), and *Coleochaete orbicularis* (Ju et al., 2015). From all detected homologs, an unrooted maximum likelihood phylogeny was computed using LG+I+G4 as model for protein evolution (chosen according to BIC). 100 bootstrap replicates were computed; only bootstrap values ≥50 are shown and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogenies. Purple font highlights those streptophyte algal sequences that share the conserved alpha helix cap domain with CSE. Logos are based on a motif (amino acids 132-141 in MAGL6 and 167-176 in CSE/MAGL3) that is situated in a region likely involved in substrate binding based on the crystal structure of human MAGL.

Figure 8. Low resolution on the complex evolutionary history of COMT. We explored the diversity of methyltransferases by screening for sequences homologous to Arabidopsis COMT/OMT1 across genome data from fifteen land plant, seven streptophyte algae, and five chlorophytes; additionally, we included sequences found in the transcriptomes of *Spirogyra pratensis* (de Vries et al., 2020), *Zygnema circumcarinatum* (de Vries et al., 2018), and *Coleochaete orbicularis* (Ju et al., 2015). From all detected homologs, an unrooted maximum likelihood phylogeny of 226 sequences was computed using LG+I+G4 as model for protein evolution (chosen according to BIC). 100 bootstrap replicates were computed; only bootstrap values ≥50 are shown and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogenies. Blue font highlights streptophyte algal sequences; bold font pinpoints those, that recovered land plant COMT as closest structural analogs in I-TASSER-based modeling. On the right we show residues important for substrate binding and function of canonical COMT as reported by Louie et al. (2010).

Figure 9. A phylogenetic framework for the evolutionary origin of CCoAOMTs in Phragmoplastophyta. CCoAOMT homologs were sampled from protein data from sixteen land plant, seven streptophyte algal, and five chlorophyte algal genomes as well as sequences found in the transcriptomes of *Spirogyra pratensis* (de Vries et al., 2020), *Coleochaete scutata* and *Zygnema circumcarinatum* (de Vries et al., 2018), and *Coleochaete orbicularis* (Ju et al., 2015). From all detected homologs, an unrooted maximum likelihood phylogeny of 138 sequences was computed using LG+G4 as model for protein evolution

(chosen according to BIC). 100 bootstrap replicates were computed; only bootstrap values ≥50 are shown and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogenies. Purple font highlights those streptophyte algal sequences that share the conserved alpha helix cap domain with CSE. The alignment on the right shows functionally characterized sites involved in substrate, ion and co-factor recognition of CCoAOMT (Ferrer et al., 2005).

Figure 10. A summary of the proposed evolutionary trajectory of key enzymes in the phenylpropanoid pathway across the green lineage. At the bottom is a cladogram of the green lineage. The most recent common ancestors (MRCA) of Chloroplastida, Streptophyta, Phragmoplastophyta, Bryophyta, Embryophyta and Tracheophyta are indicated at their respective nodes. On top the cladogram is the proposed evolutionary trajectory of the enzyme families PAL, C4H, 4CL, CCR, CAD, C3H, CSE/MAGL (CSE), and CCoAOMT. The names of the enzyme families are indicated on the left of the trajectory. The enzyme (sub-)families present in a specific common ancestor have been plotted onto the respective nodes of the cladogram below the evolutionary scenario of the enzyme families involved in the phenylpropanoid pathway and lignin biosynthesis. White dots indicate absence/loss of a gene family one dot indicates the presence of one representative of the gene family and several dots indicate an expansion (two or more members of the gene family) in at least one species of the represented lineages in the cladogram. Colors are chosen to distinguish different enzyme families and subfamilies. Question marks label sequences of ambiguous affiliation.

REFERENCES

Amborella Genome Project (2013) The *Amborella* genome and the evolution of flowering plants. *Science,* 342, 1241089

Argout, X., Salse, J., Aury, J.-M., Guiltinan, M. J., Droc, G., Gouzy, J., … Lanaud, C. (2011). The genome of *Theobroma cacao*. *Nature Genetics*, 43, 101–108.

Aschauer, P., Rengachari, S., Lichtenegger, J., Schittmayer, M., Padmanabha Das, K. M., Mayer, N., … Oberer, M. (2016) Crystal structure of the Saccharomyces cerevisiae monoglyceride lipase Yju3p. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids,* 1861, 462–470.

1092 Banks, J. A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, N., dePamhilis, C., Albert,
1093 V. A., … Grigoriev, I. V. (2011). The Selaginella genome identifies genetic changes
1094 associated with the evolution of vascular plants. *Science*, 332, 960-963.
1095 Barakat, A., Bagniewska-Zadworna, A., Choi, A. Plakkat, U., DiLoreto, D. S., Yellanki, P.,
1096 & Carlson, J. E. (2009). The cinnamyl alcohol dehydrogenase gene family in *Populus*:
1097 phylogeny, organization, and expression. *BMC Plant Biology* 9, 26.
1098 Barakate, A., Stephens, J., Goldie, A., Hunter, W.N., Marshall, D., Hancock, R.D., Lapierre,
1099 C., Morreel, K., Boerjan, W., & Halpin, C. (2011). Syringyl lignin is unaltered by severe
1100 sinapyl alcohol dehydrogenase suppression in tobacco. *The Plant Cell*, 23, 4492-4506.
1101 Barros, J., Serrani-Yarce, J.C., Chen, F., Baxter, D., Venables, B.J., & Dixon, R.A. (2016).
1102 Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. *Nature Plants*, 2,
1103 16050.
1104 Barros, J., & Dixon, R.A. (2020). Plant Phenylalanine/Tyrosine Ammonia-lyases. *Trends in*
1105 *Plant Science*, 25, 66-79.
1106 Bednarek, P., Schneider, B., Svatoš, A., Oldham, N.J., Hahlbrock, K. (2005). Structural
1107 complexity, differential response to infection, and tissue specficity of indolic and
1108 phenylpropanoid secondary metabiolisms in Arabidopsis roots. *Plant Physiology*, 138,
1109 1058–1070.
1110 Berens, M.L., Berry, H.M., Mine, A., Argueso, C.T., & Tsuda, K. (2017). Evolution of
1111 hormone signaling networks in plant defense. *Annual Review of Phytopathology*, 55,
1112 401-425.
1113 Berland, H., Albert, N.W., Stavland, A., Jordheim, M., McGhie, T.K., Zhou, Y., Zhang, H.,
1114 Deroles, S.C., Schwinn, K.E., Jordan, B.R., Davies, K.M., & Andersen, Ø.M. (2019).
1115 Auronidins are a previously unreported class of flavonoid pigments that challenges when
1116 anthocyanin biosynthesis evolved in plants. *Proceedings of the National Academy of*
1117 *Sciences USA*. 116, 20232-20239.
1118 Blanc, G., Agarkova, I., Grimwood, J., Kuo, A., Brueggeman, A., Dunigan, D. D., … Van
1119 Etten, J. L. (2012). The genome of the polar eukaryotic microalga *Coccomyxa*
1120 *subellipsoidea* reveals traits of cold adaptation. *Genome Biology*, 13, R39.
1121 Blázquez, M.A., Nelson, D.C., & Weijers, D. (2020). Evolution of plant hormone response
1122 pathways. *Annual Review of Plant Biology*, 71, 327-353.
1123 Booij-James, I.S., Dube, S.K., Jansen, M.A.K., Edelman, M., & Mattoo, A.K. (2000).
1124 Ultraviolet-B radiation impacts light-mediated turnover of the photosystem II reaction

center heterodimer in Arabidopsis mutants altered in phenolic metabolism. *Plant Physiology*, 124, 1275–1284

Bowman, J. L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., … Schmutz, J. (2017). Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell*, 171, 287-304.

Byeon, Y., Lee, H. J., Lee, H. Y., & Back, K. (2016). Cloning and functional characterization of the *Arabidopsis N*-acetylserotonin *O*-methyltransferase responsible for melatonin synthesis. *Journal of Pineal Research*, 60, 65-73.

Carella, P., Gogleva, A., Hoey, D.J., Bridgen, A.J., Stolze, S.C., Nakagami, H., & Schornack, S. (2019). Conserved biochemical defenses underpin host responses to oomycete infection in an early-divergent land plant lineage. *Current Biology*, 29, 2282–2294.e5.

Cheng, S., Xian, W., Fu, Y., Marin, B., Keller, J., Wu, T., … Melkonian, M. (2019) Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell*, 179, 1057-1067.e14.

Chezem, W.R., Memon, A., Li, F.-S., Weng, J.-K., Clay, N.K. (2017). SG2-type R2R3-MYB transcription factor MYB15 controls defense-induced lignification and basal immunity in Arabidopsis. *The Plant Cell*, 29, 1907-1926.

Clayton, W. A., Albert, N. W., Thrimawithana, A. H., McGhie, T. K., Deroles, S. C., Schwinn, K. E., … Davies, K. M. (2018). UVR8□mediated induction of flavonoid biosynthesis for UVB tolerance is conserved between the liverwort *Marchantia polymorpha* and flowering plants. *Plant Journal*, 96, 503–517.

Costa, M.A., Bedgar, D.L., Moinuddin, S.G.A., Kim, K.-W., Cardenas, C.L., Cochrane, F.C., … Lewis, N. G. (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.

Danielsson, M., Lundén, K., Elfstrand, M., Hu, J., Zhao, T., Arnerup, J., Ihrmark, K., Swedjemark, G., Borg-Karlson, A.K., & Stenlid, J. (2011). Chemical and transcriptional responses of Norway spruce genotypes with different susceptibility to *Heterobasidion* spp. infection. *BMC Plant Biology*, 11, 154.

D'Auria, J.C. (2006). Acyltransferases in plants: a good time to be BAHD. *Current Opinion in Plant Biology*, 9, 331-340.

De Clerck, O., Kao, S.-M., Bogaert, K. A., Blomme, J., Foflonker, F., Kwantes, M. … Bothwell, J. H. (2018) Insights into the evolution of multicellularity from the sea lettuce genome. *Current Biology*, 28, 2921-2933.e2925.

1159 Delwiche, C. F., Graham, L. E., Thomson, N. (1989). Lignin-like compounds and
1160 sporopollenin in *Coleochaete*, an algal model for land plant ancestry. *Science*, 245, 399–
1161 401.
1162 Devic, M., Guilleminot, J., Debeaujon, I., Bechtold, N., Bensaude, E., Koornneef, M.,
1163 Pelletier, G. & Delseny, M. (1999). The BANYULS gene encodes a DFR□like protein
1164 and is a marker of early seed coat development. *The Plant Journal*, 19, 387-398.
1165 de Azevedo Souza, C., Kim, S.S., Koch, S., Kienow, L., Schneider, K., McKim, S.M.,
1166 Haughn, G.W., Kombrink, E., & Douglas, C.J. (2009). A novel fatty acyl-CoA
1167 synthetase is required for pollen development and sporopollenin biosynthesis in
1168 *Arabidopsis*. *Plant Cell* 21, 507-525.
1169 de Vries, J., Curtis, B.A., Gould, S.B. & Archibald, J.M. (2018). Embryophyte stress
1170 signaling evolved in the algal progenitors of land plants. *Proceedings of the National*
1171 *Academy of Sciences USA*, 115, E3471–E3480.
1172 de Vries, J., de Vries, S., Slamovits, C.H., Rose, L.E. & Archibald, J.M. (2017) How
1173 embryophytic is the biosynthesis of phenylpropanoids and their derivatives in
1174 streptophyte algae? *Plant Cell Physiol* 58, 934–945.
1175 de Vries, J., & Archibald, J.M. (2018) Plant evolution: Landmarks on the path to terrestrial
1176 life. *New Phytologist* 217, 1428–1434.
1177 de Vries, J. de Vries, S., Curtis, B. A., Zhou, H., Penny, S., Feussner, K., Pinto, D. M.,
1178 Steinert, M., Cohen, A. M., von Schwartzenberg, K., & Archibald, J. M. (2020) Heat
1179 stress response in the closest algal relatives of land plants reveals conserved stress
1180 signalling circuits. *The Plant Journal*, 103, 1025-1048.
1181 de Vries, J., & Ischebeck, T. (2020) Ties between stress and lipid droplets pre-date seeds.
1182 *Trends Plant Sci*. 12, 1203-1214.
1183 de Vries, S., Herrfurth, C., Li, F.-W., Feussner, I., & de Vries, J. (2021) An ancient route
1184 towards salicylic acid and its implications for the perpetual *Trichormus–Azolla*
1185 symbiosis. *bioRxiv* preprint doi: https://doi.org/10.1101/2021.03.12.435107
1186 Dixon, R. A., & Paiva, N. L. (1995) Stress-induced phenylpropanoid metabolism. *Plant Cell*,
1187 7, 1085-1097.
1188 Dixon, R.A., Achnine, L., Kota, P., Liu, C.-J., Srinivasa Reddy, M.S., & Wang, L. (2002).
1189 The phenylpropanoid pathway and plant defence—a genomics perspective. *Molecular*
1190 *Plant Pathology* 3, 371-390.
1191 Do, C.-T., Pollet, B., Thévenin, J., Sibout, R., Denoue, D., Barrière, Y., Lapierre, C., &
1192 Jouanin, L. (2007) Both caffeoyl coenzyme A 3-*O*-methyltransferase 1 and caffeic

1193 acid *O*-methyltransferase 1 are involved in redundant functions for lignin, flavonoids and
1194 sinapoyl malate biosynthesis in Arabidopsis. *Planta*, 226, 1117–1129.
1195 Ehlting, J., Büttner, D., Wang, Q., Douglas, C.J., Somssich, I.E., & Kombrink, E. (1999).
1196 Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two
1197 evolutionarily divergent classes in angiosperms. *The Plant Journal*, 19, 9-20.
1198 Emiliani, G., Fondi, M., Fani, R. & Gribaldo, S. (2009). A horizontal gene transfer at the
1199 origin of phenylpropanoid metabolism: a key adaptation of plants to land. *Biology*
1200 *Direct*, 4, 7.
1201 Eudes, A., Pereira, J.H., Yogiswara, S., Wang, G., Benites, V.T., Baidoo, E.E.K., Lee, T.S.,
1202 Adams, P.D., Keasling, J.D., & Loqué, D. (2016) Exploiting the substrate promiscuity of
1203 hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl transferase to reduce lignin. *Plant*
1204 *Cell Physiol*, 57, 568-579.
1205 Fellenberg, C., Milkowski, C., Hause, B., Lange, P.□R., Böttcher, C., Schmidt, J., & Vogt,
1206 T. (2008). Tapetum□specific location of a cation□dependent O□methyltransferase
1207 in *Arabidopsis thaliana*. *The Plant Journal*, 56, 132– 145.
1208 Franks, N.P., Jenkins, A., Conti, E., Lieb, W.R., & Brick, P. (1998). Structural basis for the
1209 inhibition of firefly luciferase by a general anesthetic. *Biophysical Journal*, 75, 2205-
1210 2211.
1211 Ferrer, J.L., Zubieta, C., Dixon, R.A., & Noel, J.P. (2005). Crystal structures of alfalfa
1212 caffeoyl coenzyme A 3-O-methyltransferase. *Plant Physiology*, 137, 1009–1017.
1213 Fürst-Jansen, J.M.R., de Vries, S. & de Vries, J. (2020). Evo-physio: on stress responses and
1214 the earliest land plants. *Journal of Experimental Botany*, 11, 3254–3269.
1215 Goiris, K., Muylaert, K., Voorspoels, S., Noten, B., De Paepe, D., Baart, G. J. E., & De
1216 Cooman, L. (2014). Detection of flavonoids in microalgae from different evolutionary
1217 lineages. *Journal of Phycology*, 50, 483–492.
1218 Gouy, M., Guindon, S., & Gascuel, O. (2010) SeaView Version 4: A multiplatform graphical
1219 user interface for sequence alignment and phylogenetic tree building. *Molecular Biology*
1220 *and Evolution*, 27, 221–224.
1221 Güngör, E., Brouwer P., Dijkhuizen, L. W., Shaffar, D. C., Nierop, K. G. J., de Vos, R. C. H.,
1222 Toraño, J. S., van der Meer, I. N., & Schluepmann, H. (2021) *Azolla* ferns testify: seed
1223 plants and ferns share a common ancestor for leucoanthocyanidin reductase enzymes.
1224 *New Phytologist*, 229, 1118-1132.

1225 Guo, D.-M., Ran, J.-H., & Wang, X.-Q. (2010). Evolution of the Cinnamyl/Sinapyl Alcohol
1226 Dehydrogenase (CAD/SAD) Gene Family: The Emergence of Real Lignin is Associated
1227 with the Origin of Bona Fide CAD. *Journal of Molecular Evolution*, 71, 202-218.
1228 Grienenberger, E., Kim, S. S., Lallemand, B., Geoffroy, P., Heintz, D., de Azevedo Souza,
1229 C., Heitz, T., Douglas, C. J., Legrand, M. (2010) Analysis of TETRAKETIDE α-
1230 PYRONE REDUCTASE function in Arabidopsis thaliana reveals a previously unknown,
1231 but conserved, biochemical pathway in sporopollenin monomer biosynthesis. *The Plant*
1232 *Cell* 22, 4067-4083.
1233 Gross, G. G., Stöckigt, J., Mansell, R. L., & Zenk, M. H. (1973). Three novel enzymes
1234 involved in the reduction of ferulic acid to coniferyl alcohol in higher plants: ferulate:
1235 CoA ligase, feruloyl-CoA reductase and coniferyl alcohol oxidoreductase. *FEBS Lett*. 31,
1236 283–286
1237 Ha, C.M., Escamilla-Trevino, L., Yarce, J.C.S., Kim, H., Ralph, J., Chen, F., & Dixon, R.A.
1238 (2016). An essential role of caffeoyl shikimate esterase in monolignol biosynthesis in
1239 *Medicago truncatula*. *The Plant Journal*, 86, 363-375.
1240 Hamberger, B., Ellis, M., Friedmann, M., de Azevedo Souza, C., Barbazuk, B., & Douglas,
1241 C.J. (2007). Genome-wide analyses of phenylpropanoid-related genes in *Populus*
1242 *trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the *Populus* lignin toolbox and
1243 conservation and diversification of angiosperm gene families. *Canadian Journal of*
1244 *Botany*, 85, 1182–1201.
1245 Hoffmann, L., Maury, S., Martz, F., Geoffroy, P., & Legrand, M. (2003). Purification,
1246 cloning, and properties of an acyltransferase controlling shikimate and quinate ester
1247 intermediates in phenylpropanoid metabolism. *The Journal of Biological Chemistry*,
1248 278, 95–103.
1249 Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., … Ohta, H.
1250 (2014). *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial
1251 adaptation. *Nature Communications*, 5, 3978.
1252 Hu, Y., Gai, Y., Yin, L., Wang, X., Feng, C., Feng, L., Li, D., Jiang, X.N., & Wang, D.C.
1253 (2010). Crystal structures of a *Populus tomentosa* 4-coumarate:CoA ligase shed light on
1254 its enzymatic mechanisms. *The Plant Cell*, 22, 3093-3104.
1255 Huerta-Cepas, J., Serra, F., & Bork, P. ETE 3: Reconstruction, analysis, and visualization of
1256 phylogenomic data. *Molecular Biology and Evolution*, 33, 1635–1638.
1257 Jahns, P., & Holzwarth, A.R. (2012). The role of the xanthophyll cycle and of lutein in
1258 photoprotection of photosystem II. *BBA – Bioenergetics*, 1817, 182–193.

1259 Jiao, C., Sørensen, I., Sun, X., Sun, H., Behar, H., Alseekh, S., … Rose, J. K. C. (2020). The
1260 *Penium margaritaceum* genome: hallmarks of the origins of land plants. *Cell*, 181,
1261 P1097-1111.E12
1262 Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., … Hunter, S. (2014).
1263 InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30, 1236-
1264 1240.
1265 Ju, C., Van de Poel, B., Cooper, E. D., Thierer, J. H., Gibbons, T. R., Delwiche, C. F., &
1266 Chang, C. (2015). Conservation of ethylene as a plant hormone over 450 million years of
1267 evolution. *Nature Plants*, 1, 14004.
1268 Kaur, H., Heinzel, N., Schöttner, M., Baldwin, I.T., & Gális, I. (2010). R2R3-NaMYB8
1269 regulates the accumulation of phenylpropanoid-polyamine conjugates, which are
1270 essential for local and systemic defense against insect herbivores in *Nicotiana attenuata*.
1271 *Plant Physiology*, 152, 1731-1747.
1272 Kim, S.-J., Kim, M.-R., Bedgar, D.L., Moinuddin, S.G.A., Cardenas, C.L., Davin, L.B.,
1273 Kang, C., & Lewis, N.G. (2004). Functional reclassification of the putative cinnamyl
1274 alcohol dehydrogenase multigene family in Arabidopsis. *Proceedings of the National*
1275 *Academy of Sciences USA*, 101, 1455-1460.
1276 Kim, R. J., Kim, H.J., Shim, D., & Suh, M.C. (2016). Molecular and biochemical
1277 characterizations of the monoacylglycerol lipase gene family of *Arabidopsis thaliana*.
1278 *The Plant Journal*, 85, 758-771.
1279 König, S., Feussner, K., Kaever, A., Landesfeind, M., Thurow, C., Karlovsky, P., Gatz, C.,
1280 Polle, A., & Feussner, I. (2014). Soluble phenylpropanoids are involved in the defense
1281 response of Arabidopsis against *Verticillium longisporum*. *New Phytologist*, 202, 823-
1282 837.
1283 Kriegshauser, L., Knosp, S., Grienenberger, E., Tatsumi, K., Gütle, D. D., Sørensen, I., …
1284 Renault, H. (2021) Function of the HYDROXYCINNAMOYL-CoA:SHIKIMATE
1285 HYDROXYCINNAMOYL TRANSFERASE is evolutionarily conserved in
1286 embryophytes. *The Plant Cell, in press, preprint on bioRxiv*, 2020.09.16.300285.
1287 Labar, G., Bauvois, C., Borel, F., Ferrer, J. L., Wouters, J., & Lambert, D.M. (2010), Crystal
1288 Structure of the Human Monoacylglycerol Lipase, a Key Actor in Endocannabinoid
1289 Signaling. *ChemBioChem*, 11, 218-227.
1290 Labeeuw, L., Martone, P.T., Boucher, Y., & Case, R.J. (2015) Ancient origin of the
1291 biosynthesis of lignin precursors. *Biology Direct* 10, 23.

1292 Lacombe, E., Hawkins, S., Doorsselaere, J.V., Piquemal, J., Goffner, D., Poeydomenge, O.,
1293 Boudet, A-M., & Grima-Pettenati, J. G. (1997). Cinnamoyl CoA reductase, the first
1294 committed enzyme of the lignin branch biosynthetic pathway: cloning, expression and
1295 phylogenetic relationships. *The Plant Journal* 11, 429–441.
1296 Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., … Rensing, S. A.
1297 (2018) The *Physcomitrella patens* chromosome-scale assembly reveals moss genome
1298 structure and evolution. *The Plant Journal*, 93, 515–533.
1299 Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., … Huala, E.
1300 (2011). The Arabidopsis Information Resource (TAIR): improved gene annotation and
1301 new tools. *Nucleic Acids Research*, 40, D1202-D1210.
1302 Le S.Q, & Gascuel O. (2008). An Improved General Amino Acid Replacement Matrix. *Mol.*
1303 *Biol. Evol.* 25, 1307-1320.
1304 Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., Deyholos, M. K., Gitzendanner, M. A.,
1305 Graham, S. W., … Wong, G. K.-S. (2019) One thousand plant transcriptomes and the
1306 phylogenomics of green plants. *Nature*, 574, 679–685.
1307 Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new
1308 developments. *Nucleic Acids Research*, 47, W256-W259.
1309 Li, W., Zhang, F., Wu, R., Jia, R., Li, G., Guo, Y., Liu, C., & Wang, W. (2017). A novel N-
1310 Methyltransferase in Arabidopsis appears to feed a conserved pathway for nicotinate
1311 detoxification among land plants and is associated with lignin biosynthesis. *Plant*
1312 *Physiology*, 174, 1492-1504.
1313 Li, F.-W., Nishiyama, T., Waller, M., Frangedakis, E., Keller, J., Li, Z. … Szövényi, P.
1314 (2020) Anthoceros genomes illuminate the origin of land plants and the unique biology
1315 of hornworts. *Nature Plants*, 6, 259-272.
1316 Li, F.-W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P.-M., … Pryer,
1317 K. M. (2018) Fern genomes elucidate land plant evolution and cyanobacterial symbioses.
1318 *Nature Plants*, 4, 460-472.
1319 Louie, G. V., Bowman, M. E., Tu, Y., Mouradov, A., Spangenberg, G., & Noel,
1320 J.P. (2010). Structure□function analyses of a caffeic acid *O*□methyltransferase from
1321 perennial ryegrass reveal the molecular basis for substrate preference. *The Plant Cell*, 22,
1322 4114– 4127.
1323 Maeda, H. A., & Fernie, A. R. (2021) Evolutionary history of plant metabolism. *Annual*
1324 *Review of Plant Biology*, 72, 10.1146/annurev-arplant-080620-031054

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

Matsuno, M., Compagnon, V., Schoch, G.A., Schmitt, M., Debayl, D., Bassard, J.-E. … Werck-Reichhart, D. (2009). Evolution of a novel phenolic pathway for pollen development. *Science* 325, 1688–1692.

Martone, P. T., Estevez, J. M., Lu, F., Ruel, K., Denny, M. W., Somerville, C., & Ralph, J. (2009). Discovery of lignin in seaweed reveals convergent evolution of cell-wall architecture. *Current Biology*, 19, 169-175.

Martz, F., Maury, S., Pincon, G., & Legrand, M. (1998). cDNA cloning, substrate specificity and expression study of tobacco caffeoyl- CoA 3-*O*-methyltransferase, a lignin biosynthetic enzyme. *Plant Molecular Biology*, 36, 427–437.

Maury, S., Geoffroy, P., & Legrand, M. (1999). Tobacco O-methyltransferases involved in phenylpropanoid metabolism: the different caffeoyl- coenzyme A/5-hydroxyferuloyl-coenzyme A 3/5-O-methyltransferase and caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase classes have distinct substrate specificities and expression patterns. *Plant Physiology*, 121, 215–224.

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., … Grossman, A. R. (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318, 245-250.

Miller, M., Owens, S.J., & Rørslett, B. (2011). Plants and colour: flowers and pollination. *Optics & Laser Technology*, 43, 282-294.

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

Moreau, H., Verhelst, B., Couloux, A., Derelle, E., Rombauts, S., Grimsley, N., … Vandepoele, K. (2012). Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biology*, 13, R74.

Nakatsu, T., Ichiyama, S., Hiratake, J., Saldanha, A., Kobashi, N., Sakata, K., & Kato, H. (2006). Structural basis for the spectral difference in luciferase bioluminescence. *Nature*, 440, 372-376.

Nelson, D., & Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *The Plant Journal*, 66, 194-211.

1359 Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K.
1360 K., … Rensing, S. A. (2018). The *Chara* genome: secondary complexity and
1361 implications for plant terrestrialization. *Cell*, 174, 448–464.
1362 Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., …
1363 Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution.
1364 *Nature*, 497, 579-584.
1365 Oliva, J., Rommel, S., Fossdal, C.G., Hietala, A.M., Nemesio-Gorriz, M., Solheim, H., &
1366 Elfstrand, M. (2015). Transcriptional responses of Norway spruce (*Picea abies*) inner
1367 sapwood against *Heterobasidion parviporum*. *Tree Physiology*, 35, 1007–1015.
1368 Omura, T. (1999). Forty years of cytochrome P450. *Biochemical and Biophysical Research*
1369 *Communications*, 266, 690-698.
1370 Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., … Buell, C. R. (2007).
1371 The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic*
1372 *Acids Research*, 35, D883–D887.
1373 Overdijk, E.J.R., de Keijzer, J., de Groot, D., Schoina, C., Bouwmeester, K., Ketelaar, T., &
1374 Govers, F. (2016). Interaction between the moss *Physcomitrella patens* and
1375 *Phytophthora*: a novel pathosystem for live□cell imaging of subcellular defence. *Journal*
1376 *of Microscopy*, 263, 171–180.
1377 Palenik, B., Grimwood, J., Aerts, A., Rouzé, P., Salamov, A., Putman, N., … Grigoriev, I. V.
1378 (2007). The tiny eukaryote Ostreococcus provides genomic insights into the paradox of
1379 plankton speciation. *Proceedings of the National Academy USA.* 104, 7705–7710.
1380 Pan, H., Zhou, R., Louie, G.V., Mühlemann, J. K., Bomati, E. K., Bowman, M. E., … Wang,
1381 X. (2014). Structural Studies of Cinnamoyl-CoA Reductase and Cinnamyl-Alcohol
1382 Dehydrogenase, key enzymes of monolignol biosynthesis. *The Plant Cell* 26, 3709.
1383 Piatkowski, B. T., Imwattana, K., Tripp, E. A., Weston, D. J., Healey, A., Schmutz, J., &
1384 Shaw, A. J. (2020). Phylogenomics reveals convergent evolution of red-violet coloration
1385 in land plants and the origins of the anthocyanin biosynthetic pathway. *Molecular*
1386 *Phylogenetics and Evolution*, 151, 106904.
1387 Ponce De León, I., Schmelz, E. A., Gaggero, C., Castro, A., Álvarez, A., & Montesano, M.
1388 (2012). *Physcomitrella patens* activates reinforcement of the cell wall, programmed cell
1389 death and accumulation of evolutionary conserved defence signals, such as salicylic acid
1390 and 12-oxo-phytodienoic acid, but not jasmonic acid, upon *Botrytis cinerea* infection.
1391 *Molecular Plant Pathology*, 13, 960-74.

1392 Prochnik, S.E., Umen, J., Nedelcu, A. M., Hallmann, A., Miller, S. M., Nishii, I., … Rokhsar,
1393 D. S. (2010). Genomic analysis of organismal complexity in the multicellular green alga
1394 *Volvox carteri*. *Science*, 329, 223–226
1395 Puttick, M.N., Morris, J.L., Williams, T.A., Cox, C.J., Edwards, D., Kenrick, P., Pressel, S.,
1396 Wellman, C.H., Schneider, H., Pisani, D., & Donoghue, P.C.J. (2018). The
1397 interrelationships of land plants and the nature of the ancestral embryophyte. *Current*
1398 *Biology*, 28, 733-745.
1399 Ralph, J., Lundquist, K., Brunow, G., Lu, F., Kim, H., Schatz, P. F., Marita, J. M., Hatfield,
1400 R. D., Ralph, S. A., Christensen, J. H., & Boerjan, W. (2004). Lignins: natural polymers
1401 from oxidative coupling of 4-hydroxyphenylpropanoids. *Phytochemistry Reviews*, 3, 29–
1402 60.
1403 Rupasinghe S., Baudry J. & Schuler M.A. (2003) Common active site architecture and
1404 binding strategy of four phenylpropanoid P450s from *Arabidopsis thaliana* as revealed
1405 by molecular modeling. *Protein Engineering Design and Selection*, 16, 721–731.
1406 Renault, H., Alber, A., Horst, N.A., Basilio Lopes, A., Fich, E. A., Kriegshauser, L., …
1407 Werck-Reichhart, D. (2017a). A phenol-enriched cuticle is ancestral to lignin evolution
1408 in land plants. *Nature Communications*, 8, 14713.
1409 Renault, H., De Marothy, M., Jonasson, G., Lara, P., Nelson, D. R., Nilsson, I., André, F.,
1410 von Heijne, G., & Werck-Reichhart, D. (2017b). Gene duplication leads to altered
1411 membrane topology of a cytochrome P450 enzyme in seed plants. *Molecular Biology*
1412 *and Evolution*, 34, 2041-2056.
1413 Renault, H., Werck-Reichhart, D., & Weng, J-K. (2019). Harnessing lignin evolution for
1414 biotechnological applications. *Current Opinion in Biotechnology*, 56, 105-111.
1415 Rensing, S.A. (2014). Gene duplication as a driver of plant morphogenetic evolution. *Current*
1416 *Opinion in Plant Biology*, 17, 43-48.
1417 Rensing, S.A. (2018) Great moments in evolution: the conquest of land by plants. *Current*
1418 *Opinion in Plant Biology*, 42, 49–54.
1419 Rippin, M., Becker, B. and Holzinger, A. (2017) Enhanced desiccation tolerance in mature
1420 cultures of the streptophytic green alga *Zygnema circumcarinatum* revealed by
1421 transcriptomics. *Plant & Cell Physiology*, 58, 2067–2084.
1422 Rippin, M., Pichrtová, M., Arc, E., Kranner, I., Becker, B. and Holzinger, A. (2019)
1423 Metatranscriptomic and metabolite profiling reveals vertical heterogeneity within a
1424 *Zygnema* green algal mat from Svalbard (High Arctic). *Environmental Microbiology*, 21,
1425 4283-4299.

1426 Ro, D. K., Mah, N., Ellis, B. E., & Douglas, C. J. (2001). Functional characterization and
1427 subcellular localization of poplar (*Populus trichocarpa* x *Populus deltoides*) cinnamate
1428 4-hydroxylase. *Plant Physiology*, 126, 317-329.
1429 Russell, D.W., Conn, E.E. (1967). The cinnamic acid 4-hydroxylase from pea seedlings.
1430 *Archives of Biochemistry and Biophysics* 122, 256-258.
1431 Saballos, A., Ejeta, G., Sanchez, E., Kang, C., & Vermerris, W. (2009) A genomewide
1432 analysis of the cinnamyl alcohol dehydrogenase family in sorghum [*Sorghum bicolor*
1433 (L.) Moench] identifies SbCAD2 as the brown midrib6 gene. *Genetics*, 181, 783–795.
1434 Saito, N., & Harborne, J.B. (1992). Correlations between anthocyanin type, pollinator and
1435 flower colour in the labiatae. *Phytochemistry*, 31, 3009-3015.
1436 Scheres, B., & van der Putten, W.H. (2017) The plant perceptron connects environment to
1437 development. *Nature* 543, 337-345.
1438 Schoenbohm, C., Martens, S., Eder, C., Forkmann, G., & Weisshaar, B. (2000) Identification
1439 of the *Arabidopsis thaliana* flavonoid 3'-hydroxylase gene and functional expression of
1440 the encoded P450 enzyme. *Biological Chemistry*, 381, 749-753.
1441 Sheahan, J.J. (1996) Sinapate esters provide greater UV-B attenuation than flavonoids in
1442 *Arabidopsis thaliana* (Brassicaceae). *American Journal of Botany*, 83, 679-686.
1443 Sheehan, H., Hermann, K., & Kuhlemeier, C. (2012). Color and Scent: How Single Genes
1444 Influence Pollinator Attraction. *Cold Spring Harbor Symposia on Quantitative Biology*,
1445 77, 117-133.
1446 Shockey, J.M., Fulda, M.S., & Browse, J. (2003). Arabidopsis contains a large superfamily of
1447 Acyl-Activating Enzymes. Phylogenetic and biochemical analysis reveals a new class of
1448 Acyl-Coenzyme A Synthetases. *Plant Physiology*, 132, 1065-1076.
1449 Sierro, N. Battey, J., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., Goepfert, S., Peitsch, M.
1450 C., & Ivanov, N. V. (2014). The tobacco genome sequence and its comparison with those
1451 of tomato and potato. *Nature Communications*, 5, 3833.
1452 Slotte, T., Hazzouri, K. M., Ågren, J. A., Koenig, D., Maumus, F., Guo, Y.-L., … Wrigth, S.
1453 I. (2013). The *Capsella rubella* genome and the genomic consequences of rapid mating
1454 system evolution. *Nature Genetics*, 45, 831–835
1455 Sørensen, I., Pettolino, F.A., Bacic, A., Ralph, J., Lu, F., O'Neill, M.A., Fei, Z., Rose, J.K.C.,
1456 Domozych, D.S., Willats, W.G.T. (2011). The charophycean green algae provide
1457 insights into the early origins of plant cell walls. *The Plant Journal*, 68, 201–211.
1458 Suzuki, S., & Umezawa, T. (2007). Biosynthesis of lignans and norlignans. *Journal of Wood*
1459 *Science* 53, 273-284.

Sytar, O., Zivcak, M., Bruckova, K., Brestic, M., Hemmerich, I., Rauh, C., Simko, I. 2018. Shift in accumulation of flavonoids and phenolic acids in lettuce attributable to changes in ultraviolet radiation and temperature. *Scientia Horticulturae*, 239, 193-204.

Szövényi, P., Frangedakis, E., Ricca, M., Quandt, D., Wicke, S., Langdale, J.A. (2015). Establishment of *Anthoceros agrestis* as a model species for studying the biology of hornworts. *BMC Plant Biology*, 15, 1-7.

Szövényi, P., Gunadi, A., Li, F.-W. (2021). Charting the genomic landscape of seed-free plants. *Nature Plants*, https://doi.org/10.1038/s41477-021-00888-z

Tan, D.-X., Hardeland, R., Manchester, L.C., Korkmaz, A., Ma, S., Rosales-Corral, S., Reiter, R.J. (2012). Functional roles of melatonin in plants, and perspectives in nutritional and agricultural science. *Journal of Experimental Botany*, 63, 577-597.

Tanaka, A., Shigemitsu, T., Yokota, Y., Shika, N. (1997). A new *Arabidopsis* mutant induced synthesis with spotted pigmentation. *Genes & Genetic Systems*, 72, 141–148.

The International *Brachypodium* Initiative. (2010). Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature*, 463, 763–768.

Urban, P., Werck-Reichhart, D., Teutsch, H.G., Durst, F., Regnier, S., Kazmeier, M., & Pompon, D. (1994). Characterization of recombinant plant cinnamate 4-hydroxylase produced in yeast. Kinetic and spectral properties of the major plant P450 of the phenylpropanoid pathway. *European Journal of Biochemistry*, 222, 843–850.

Vanholme, R., Storme, V., Vanholme, B., Sundin, S., Christensen, J.H., Goemine, G., Halpin, C., Rohde, A., Morreel, K., Boerjan, W. (2012) A systems biology view of responses to lignin biosynthesis perturbations in Arabidopsis. *The Plant Cell*, 24, 3506-3529.

Vanholme, R., Cesarino, I., Rataj, K., Xiao, Y., Sundin, L., Goeminne, G., … Boerjan, W. (2013). Caffeoyl shikimate esterase (CSE) is an enzyme in the ligninbiosynthetic pathway in *Arabidopsis*. *Science*, 341, 1103–1106.

Vanholme, R., De Meester, B., Ralph, J., & Boerjan, W. (2019). Lignin biosynthesis and its integration into metabolism. *Current Opinion Biotechnology*, 56, 230-239.

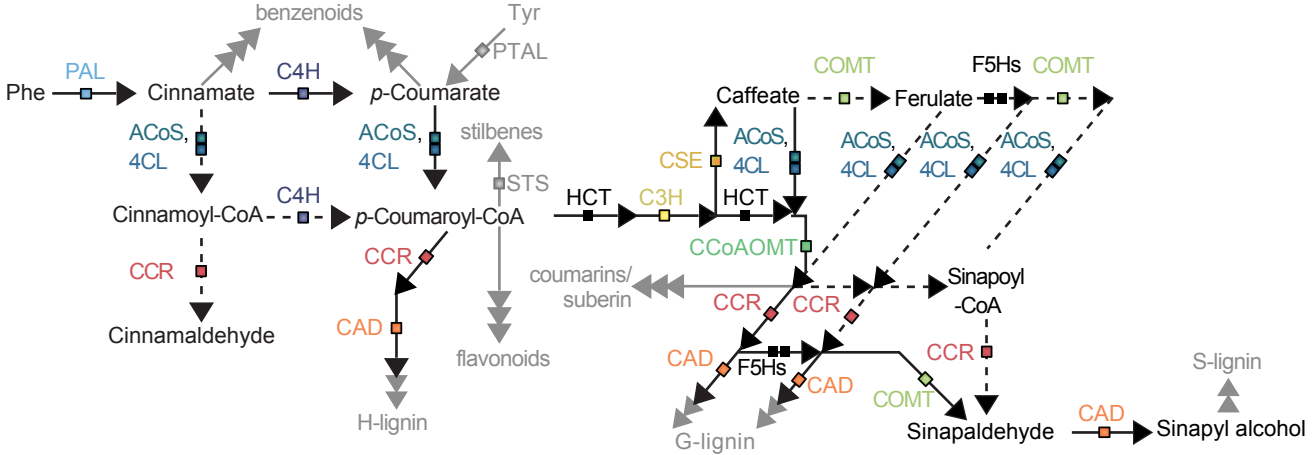
Vogt, T. (2010). Phenylpropanoid biosynthesis. *Molecular Plant*, 3, 2–20.

Wan, T. et al. (2018). A genome for gnetophytes and early evolution of seed plants. *Nature Plants*, 4, 82–89.

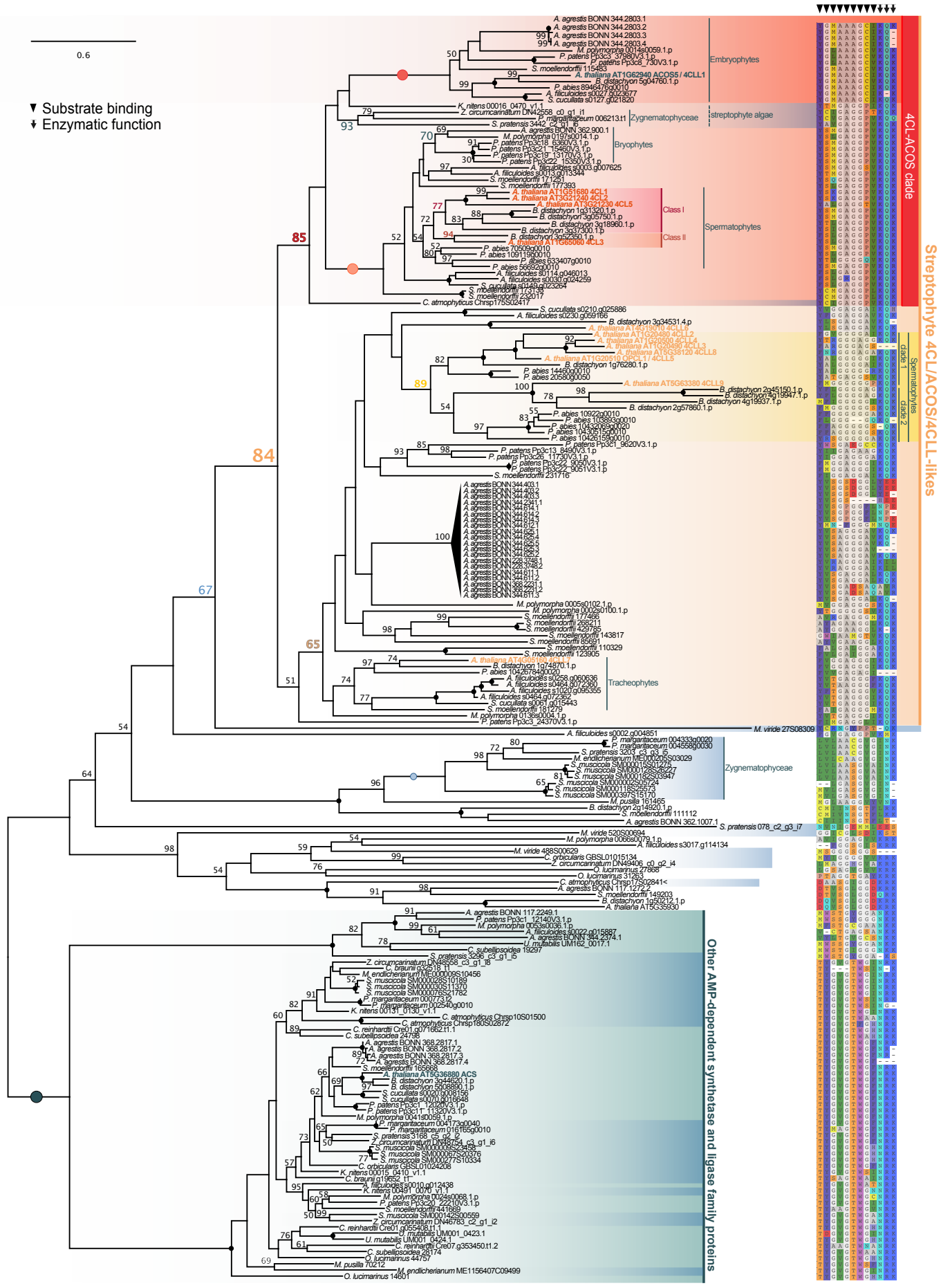
Wang, S., Li, L., Li, H., Sahu, S. K., Wang, H., Xu, Y., … Liu, X. (2020) Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nature Plants*, 6, 95-106.

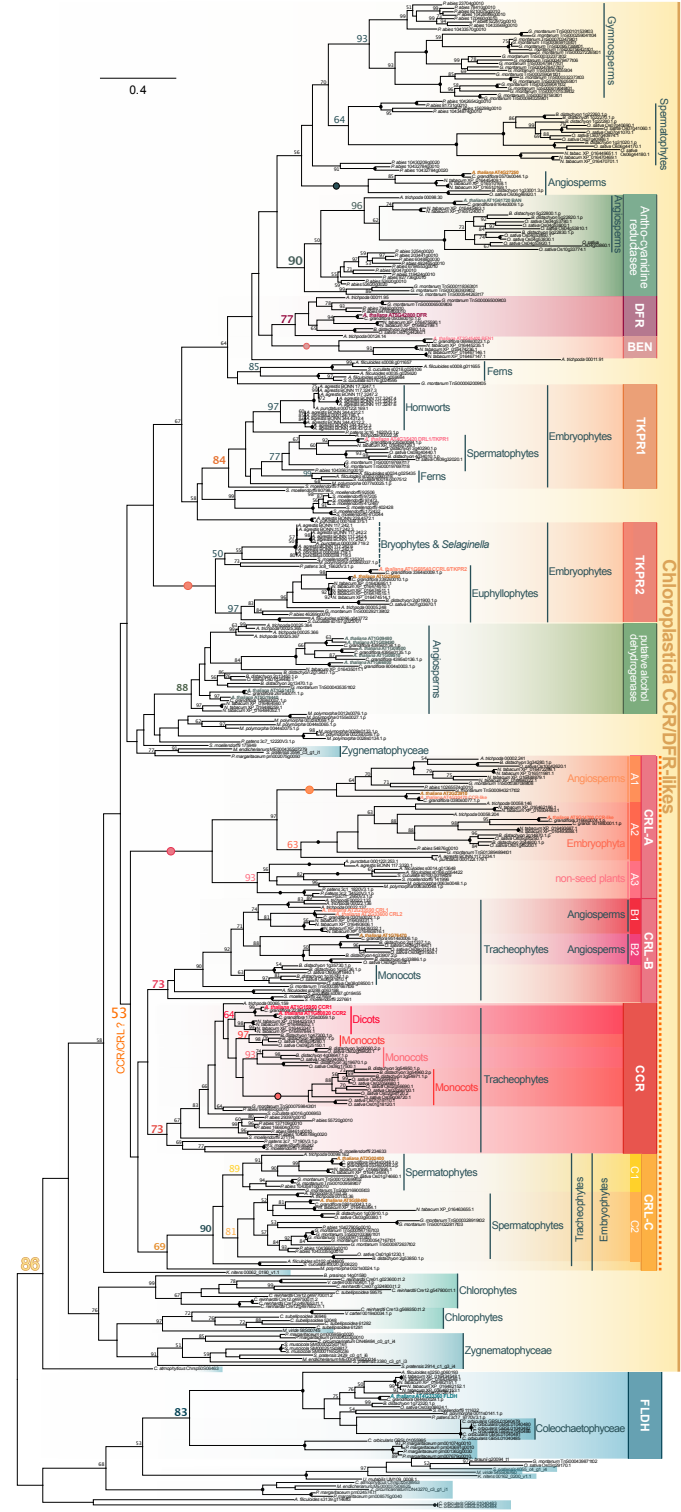
1494 Weng, J.-K., Li, X., Stout, J., & Chapple, C. (2008) Independent origins of syringyl lignin in
1495 vascular plants. *Proceedings of the National Academy of Sciences USA*, 105, 7887–7892.
1496 Weng, J.-K., & Chapple, C. (2010) The origin and evolution of lignin biosynthesis. *New*
1497 *Phytologist*, 187, 273-285.
1498 Weng, J.-K., Akiyama, T., Ralph, J., & Chapple, C. (2011) Independent recruitment of an O-
1499 methyltransferase for syringyl lignin biosynthesis in *Selaginella moellendorffii*. *The*
1500 *Plant Cell*, 23, 2708–2724.
1501 Weng, J. K. (2013) The evolutionary paths towards complexity: a metabolic perspective. *New*
1502 *Phytologist*, 201, 1141-1149.
1503 Whelan, S., & Goldman, N. (2001). A general empirical model of protein evolution derived
1504 from multiple protein families using a maximum-likelihood approach. *Molecular Biology*
1505 *and Evolution*, 18, 691-699.
1506 Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., … Leebens-
1507 Mack, J. (2014) Phylotranscriptomic analysis of the origin and early diversification of
1508 land plants. *Proceedings of the National Academy of Sciences USA*, 111, E4859–E4868.
1509 Wodniok, S., Brinkmann, H., Glöckner, G., Heidel, A. J., Philippe, H., Melkonian, M., &
1510 Becker, B. (2011) Origin of land plants: do conjugating green algae hold the key? *BMC*
1511 *Evol. Biol.* 11, 104.
1512 Wohl, J., & Petersen, M. (2020). Functional expression and characterization of cinnamic acid
1513 4□hydroxylase from the hornwort *Anthoceros agrestis* in *Physcomitrella patens*. *Plant*
1514 *Cell Reports*, 39, 597-607.
1515 Wolf, L., Rizzini, L., Stracke, R., Ulm, R., & Rensing, S. A. (2010) The molecular and
1516 physiological repsonses of *Physcomitrella patens* to ultraviolet-B radiation. *Plant*
1517 *Physiol.* 153, 1123-1134.
1518 Worden, A. Z., Lee, J.-H., Mock, T., Rouzé, P., Simmons, M. P., Aerts, A. L., … Grigoriev,
1519 I. V. (2009). Green evolution and dynamic adaptations revealed by genomes of the
1520 marine picoeukaryotes *Micromonas*. *Science*, 324, 268–272.
1521 Xu, Z., Zhang, D., Hu, J., Zhou, X., Ye, X., Reichel, K. L., … Yuan, J. S. (2009).
1522 Comparative genome analysis of lignin biosynthesis gene families across the plant
1523 kingdom. *BMC Bioinformatics* 10, S3.
1524 Xue, J.-S., Zhang, B., Zhan, H., Lv, Y-L., Jia, X.-L., Wang, T., …, Yang, Z.-N. (2020).
1525 Phenylpropanoid derivatives are essential components of sporopollenin in vascular
1526 plants. *Molecular Plant* 13, 1644-1653.

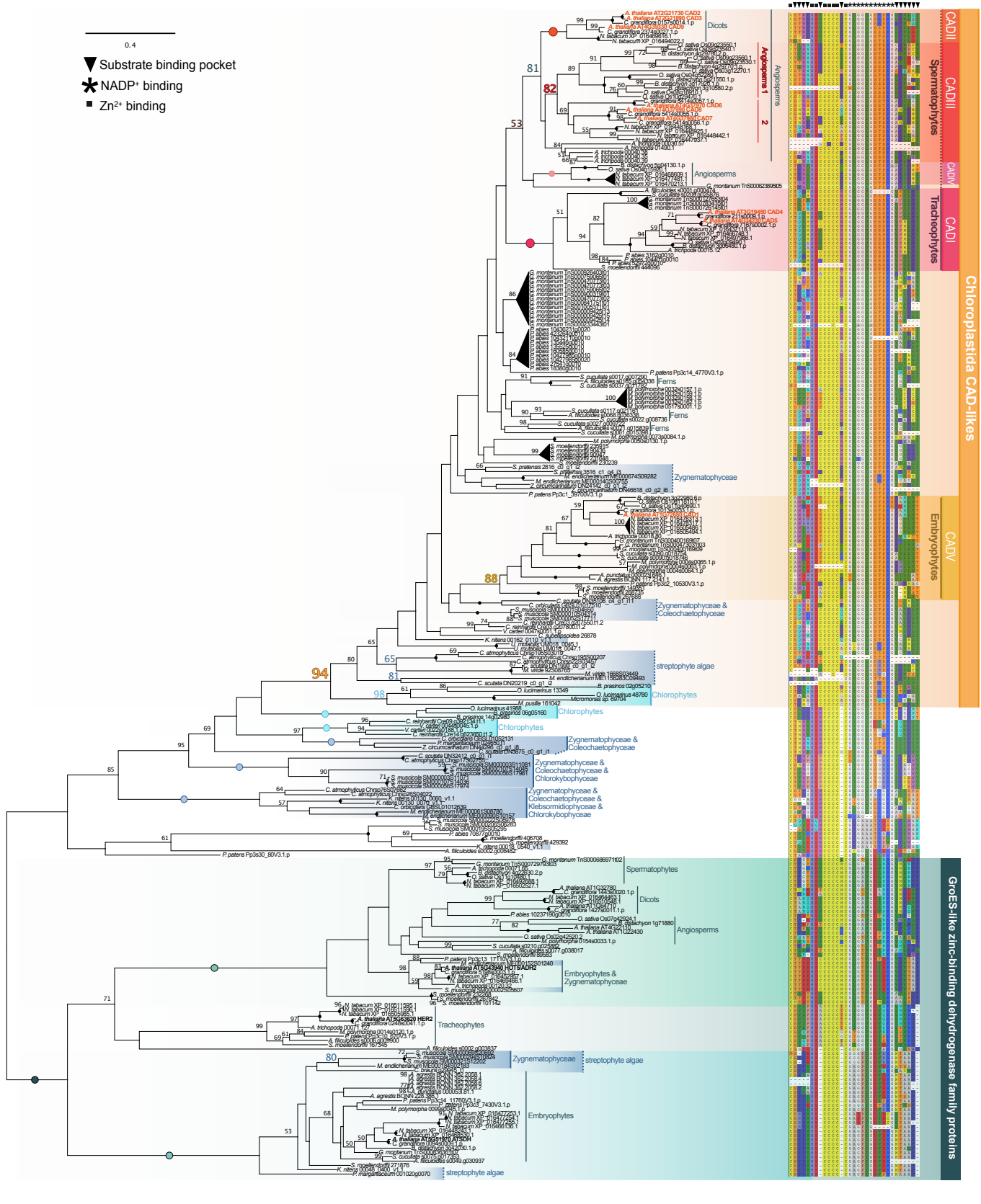
1527 Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite:
1528 protein structure and function prediction. *Nature Methods* 12, 7-8.
1529 Ye, Z.-H., Kneusel, R.E., Matern, U., & Varner, J. E. (1994). An alternative methylation
1530 pathway in lignin biosynthesis in *Zinnia*. *The Plant Cell*, 6, 1427–1439.
1531 Ye, Z.-H., & Varner, J.E. (1995). Differential expression of two *O*-methyltransferases in
1532 lignin biosynthesis in *Zinnia elegans*. *Plant Physiology*, 108, 459–467.
1533 Youn, B., Camacho, R., Moinuddin, S.G.A., Lee, C., Davin, L.B., Lewis, N.G., …, (2006)
1534 Crystal structures and catalytic mechanism of the *Arabidopsis* cinnamyl alcohol
1535 dehydrogenases AtCAD5 and AtCAD4. *Organic and Biomolecular Chemistry*, 4, 1687–
1536 1697.
1537 Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*,
1538 9, 40.
1539 Zhang, J., Fu, X.-X., Li, R.-Q., Zhao, X., Liu, Y., Li, M.-H., … Chen, Z.-D. (2020). The
1540 hornwort genome and early land plant evolution. *Nature Plants*, 6, 107-118.
1541 Zubieta, C., Kota, P., Ferrer, J.L., Dixon, R.A., & Noel, J.P. (2002). Structural basis for the
1542 modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-O-
1543 methyltransferase. *The Plant Cell*, 14, 1265-1277.


benzenoids
Tyr
PTAL
PAL
Phe
Cinnamate
C4H
p-Coumarate
ACoS,
4CL
ACoS,
4CL
stilbenes
STS
C4H
Cinnamoyl-CoA
p-Coumaroyl-CoA
CCR
CCR
Cinnamaldehyde
CAD
flavonoids
H-lignin
COMT
F5Hs
COMT
Caffeate
Ferulate
CSE
ACoS,
4CL
ACoS,
4CL
ACoS,
4CL
ACoS,
4CL
HCT
C3H
HCT
CCoAOMT
coumarins/
suberin
Sinapoyl
-CoA
CCR
CCR
CAD
F5Hs
CAD
CCR
COMT
S-lignin
G-lignin
Sinapaldehyde
CAD
Sinapyl alcohol

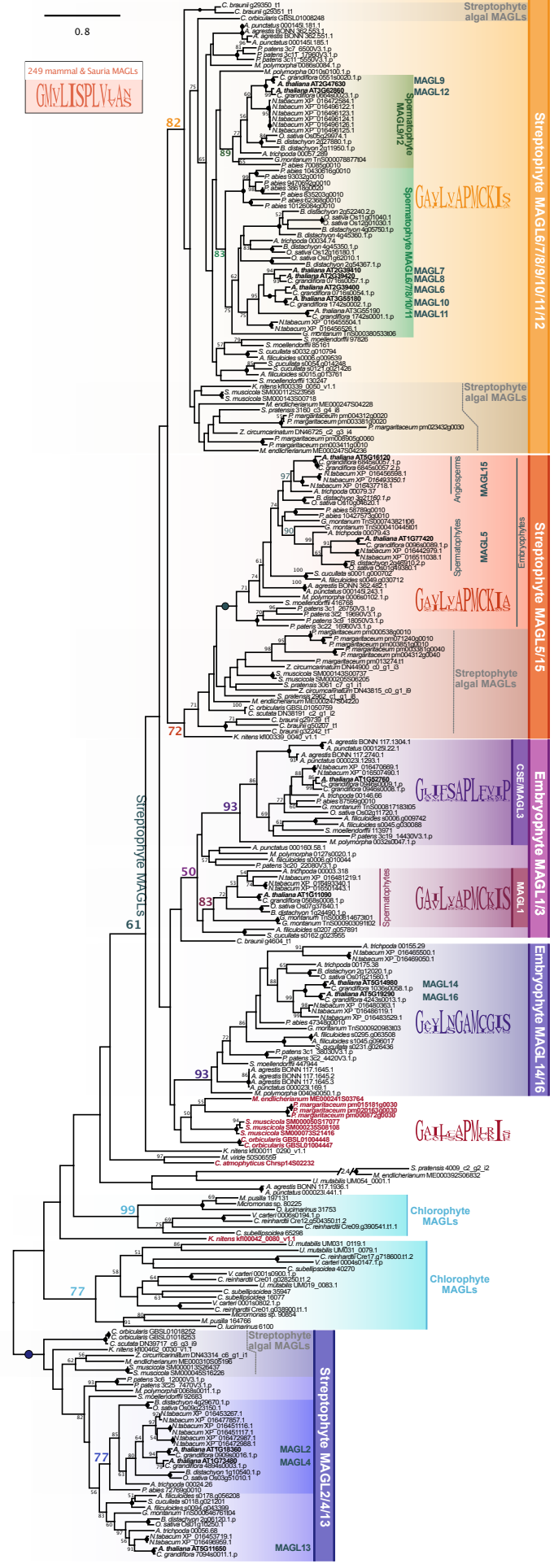

0.7
Angiosperms
Embryophytes
Embryophyte PALs
Fungal PALs
Bacterial PALs
& Klebsormidium
Eukaryotic and Prokaryotic HALs
96
82
54
68
85
K. nitens kfl00024_0250_v1.1
K. nitens kfl00104_0290_v1.1
C. braunii g66119_t1
C. braunii g34530_t1
C. braunii g57646_t1
A. thaliana AT2G37040 PAL1
A. thaliana AT3G53260 PAL2
A. thaliana AT3G10340 PAL4
A. thaliana AT5G04230 PAL3
M. polymorpha 0142s0036.1.p
B. subtilis WP_061186919.1
B. subtilis WP_051632929.1

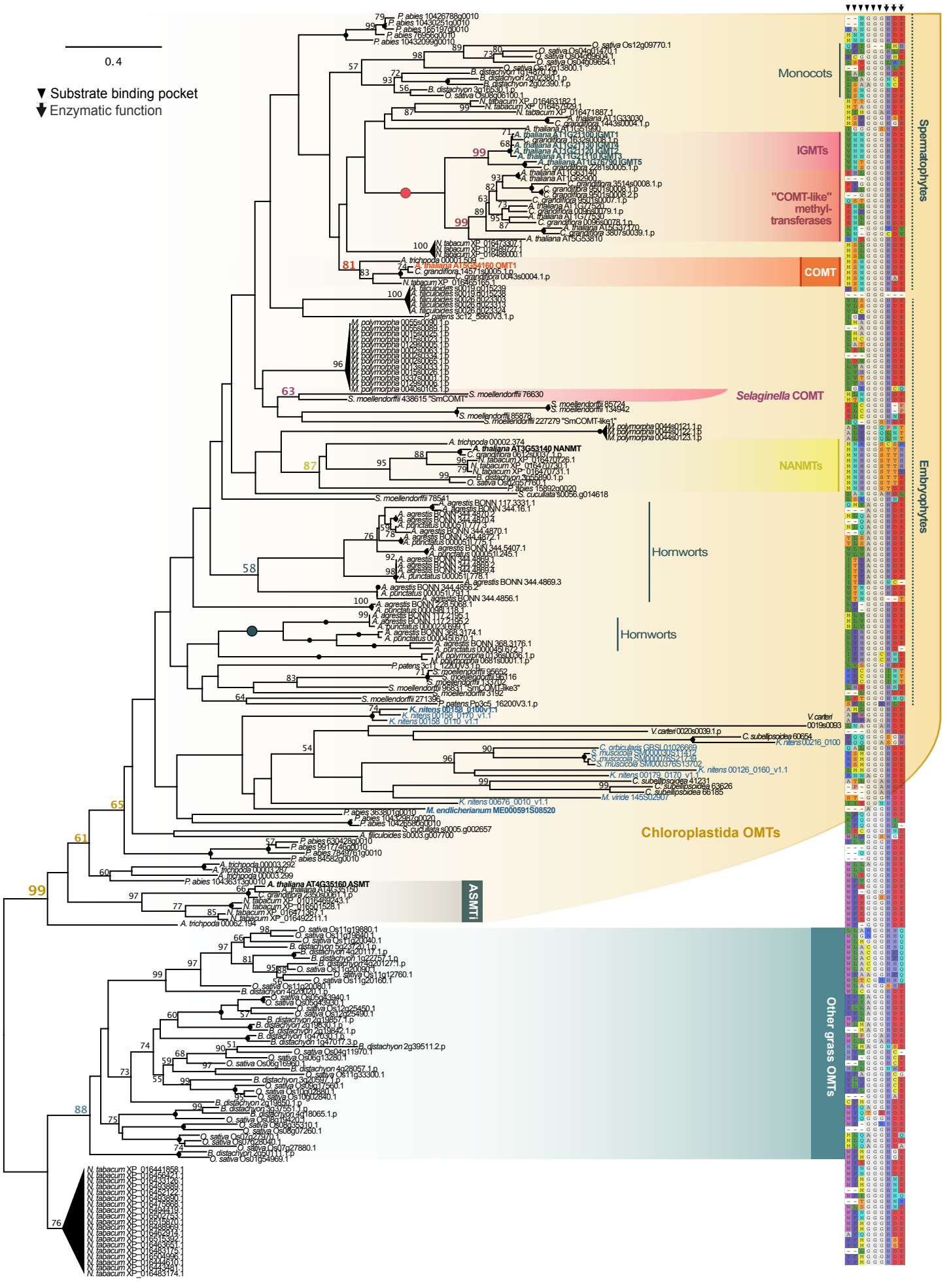

0.6
▼ Substrate binding
↓ Enzymatic function
A. agrestis BONN 344.2803.1
A. agrestis BONN 344.2803.2
A. agrestis BONN 344.2803.3
A. agrestis BONN 344.2803.4
M. polymorpha 0013s0059.1.p
P. patens Pp3c3_37960V3.1.p
P. patens Pp3c8_730V3.1.p
S. moellendorffii 115483
A. thaliana AT1G62940 ACOS5 / 4CLL1
B. distachyon 5g04760.1.p
P. abies 8946478c0010
A. filiculoides s0027 0023677
S. cucullata s0127.g021820
Embryophytes
K. nitens 00016_0470_v1.1
Z. circumcarinatum DN42558_c0_g1_i1
P. margaritaceum 006213.t1
Zygnematophyceae
streptophyte algae
S. pratensis 3442_c2_g1_i8
A. agrestis BONN 362.900.1
M. polymorpha 0197s0014.1.p
P. patens Pp3c18_8360V3.1.p
P. patens Pp3c21_15460V3.1.p
P. patens Pp3c19_13170V3.1.p
P. patens Pp3c22_15350V3.1.p
A. filiculoides s0003.g007625
A. filiculoides s0013.g013344
S. moellendorffii 171251
S. moellendorffii 177393
Bryophytes
A. thaliana AT1G51680 4CL1
A. thaliana AT3G21240 4CL2
A. thaliana AT3G21230 4CL5
B. distachyon 1g31320.1.p
B. distachyon 3g05750.1.p
B. distachyon 3g18960.1.p
Class I
A. thaliana AT1G65060 4CL3
B. distachyon 3g37300.1.p
B. distachyon 3g52550.1.p
Class II
P. abies 7059s0010
P. abies 10911960010
P. abies 633407g0010
P. abies 56692g0010
A. filiculoides s0114.g046013
A. filiculoides s0030.g024259
S. cucullata s0149.g023264
S. moellendorffii 171158
S. moellendorffii 232017
Spermatophytes
C. atmophyticus Chrsp175S02417
4CL-ACOS clade
S. cucullata s0210.g025886
A. filiculoides s0230.g069166
B. distachyon 3g34531.4.p
A. thaliana AT1G20480 4CLL2
A. thaliana AT1G20500 4CLL4
A. thaliana AT1G20490 4CLL3
A. thaliana AT5G38120 4CLL8
A. thaliana AT1G20510 OPCL1 / 4CLL5
B. distachyon 1g76280.1.p
P. abies 144600g0010
P. abies 2058g0050
A. thaliana AT5G63380 4CLL9
B. distachyon 2g45150.1.p
B. distachyon 4g18947.1.p
B. distachyon 4g19937.1.p
B. distachyon 2g57860.1.p
P. abies 10922g0010
P. abies 10389g00010
P. abies 10439g0010
P. abies 10426159g0010
P. patens Pp3c1_9620V3.1.p
P. patens Pp3c13_8490V3.1.p
P. patens Pp3c26_11730V3.1.p
P. patens Pp3c22_9650V3.1.p
P. patens Pp3c22_9051V3.1.p
S. moellendorffii 231716
clade 1
Spermatophytes
clade 2
Streptophyte 4CL/ACOS/4CLL-likes
A. agrestis BONN 344.03.1
A. agrestis BONN 344.4032
A. agrestis BONN 344.4033
A. agrestis BONN 344.2941.1
A. agrestis BONN 344.6141.1
A. agrestis BONN 344.6142
A. agrestis BONN 344.6143
A. agrestis BONN 344.6121
A. agrestis BONN 344.6251
A. agrestis BONN 344.6254
A. agrestis BONN 344.6255
A. agrestis BONN 344.6253
A. agrestis BONN 344.6252
A. agrestis BONN 228.3748.1
A. agrestis BONN 228.3748.2
A. agrestis BONN 344.6111
A. agrestis BONN 344.6112
A. agrestis BONN 368.2211.1
A. agrestis BONN 368.2211.2
A. agrestis BONN 344.611.3
M. polymorpha 0005s0102.1.p
M. polymorpha 0002s0100.1.p
S. moellendorffii 177466
S. moellendorffii 268211
S. moellendorffii 429785
S. moellendorffii 143817
S. moellendorffii 85691
S. moellendorffii 110329
S. moellendorffii 123905
A. thaliana AT4G05160 4CL7
B. distachyon 1g74870.1.p
P. abies 10426784g0020
A. filiculoides s0258.g060636
A. filiculoides s0464.g072360
A. filiculoides s1020.g095355
A. filiculoides s0464.g072362
S. cucullata s0061.g015443
S. moellendorffii 181279
M. polymorpha 0136s0004.1.p
P. patens Pp3c3_24370V3.1.p
Tracheophytes
M. viride 27S08309
A. filiculoides s0002.g004851
P. margaritaceum 004333g0020
P. margaritaceum 004558g0030
S. pratensis 3203_c3_g3_i5
M. endlicherianum ME000205S03029
S. muscicola SM000015S01275
S. muscicola SM000128S26227
S. muscicola SM000162S03947
S. muscicola SM000062S05724
S. muscicola SM000118S25573
S. muscicola SM000397S15170
Zygnematophyceae
M. pusilla 161465
B. distachyon 2g14920.1.p
S. moellendorffii 111112
A. agrestis BONN 362.1007.1
S. pratensis 078_c2_g3_i7
M. viride 520S00694
M. polymorpha 0066s0079.1.p
A. filiculoides s3017.g114134
M. viride 488S00629
C. orbicularis GBSL01015134
Z. circumcarinatum DN49406_c0_g2_i4
O. lucimarinus 31263
C. atmophyticus Chrsp17S02841<
A. agrestis BONN 117.1272.2
S. moellendorffii 149203
B. distachyon 1g52212.1.p
A. thaliana AT5G35930
A. agrestis BONN 117.2249.1
P. patens Pp3c1_12140V3.1.p
M. polymorpha 0053s0036.1.p
A. filiculoides s0022.g015897
A. agrestis BONN 344.2374.1
U. mutabilis UM162_0017.1
C. subellipsoidea 19297
S. pratensis 3256_c3_g1_i5
Z. circumcarinatum DN48558_c3_g1_i8
C. braunii g32518.11
M. endlicherianum ME000009S10456
S. muscicola SM000028S10189
S. muscicola SM000035S11370
S. muscicola SM000076S21782
P. margaritaceum 000773.t2
P. margaritaceum 002540g0010
K. nitens 00131_0130_v1.1
C. atmophyticus Chrsp10S01500
C. atmophyticus Chrsp180S02872
C. reinhardtii Cre01.g071662.t1.1
C. subellipsoidea 24798
A. agrestis BONN 368.2817.1
A. agrestis BONN 368.2817.2
A. agrestis BONN 368.2817.3
A. agrestis BONN 368.2817.4
S. moellendorffii 165668
A. thaliana AT5G36880 ACS
B. distachyon 3g44620.1.p
B. distachyon 5g08890.1.p
S. cucullata s0002.g008196
S. cucullata s0070.g016648
P. patens Pp3c1_12020V3.1.p
P. patens Pp3c1_11320V3.1.p
M. polymorpha 0041s0059.1.p
P. margaritaceum 004173g0040
P. margaritaceum 016165g0010
S. pratensis 3168_c5_g2_i2
Z. circumcarinatum DN48754_c3_g1_i6
S. muscicola SM000008S23458
S. muscicola SM000067S20376
S. muscicola SM000277S10334
C. orbicularis GBSL01024208
K. nitens 00015_0410_v1.1
C. braunii g19652.1t1
A. filiculoides s0010.g012438
K. nitens 00181_0006_v1.1
M. polymorpha 0024s0068.1.p
P. patens Pp3c20_22210V3.1.p
S. moellendorffii 441689
S. muscicola SM000142S00559
Z. circumcarinatum DN46783_c2_g1_i2
C. reinhardtii Cre01.g055408.t1.1
U. mutabilis UM001_0423.1
U. mutabilis UM001_0424.1
C. reinhardtii Cre07.g353450.t1.2
C. subellipsoidea 28174
O. lucimarinus 44757
M. pusilla 70212
M. endlicherianum ME1156407C09499
O. lucimarinus 14801
Other AMP-dependent synthetase and ligase family proteins

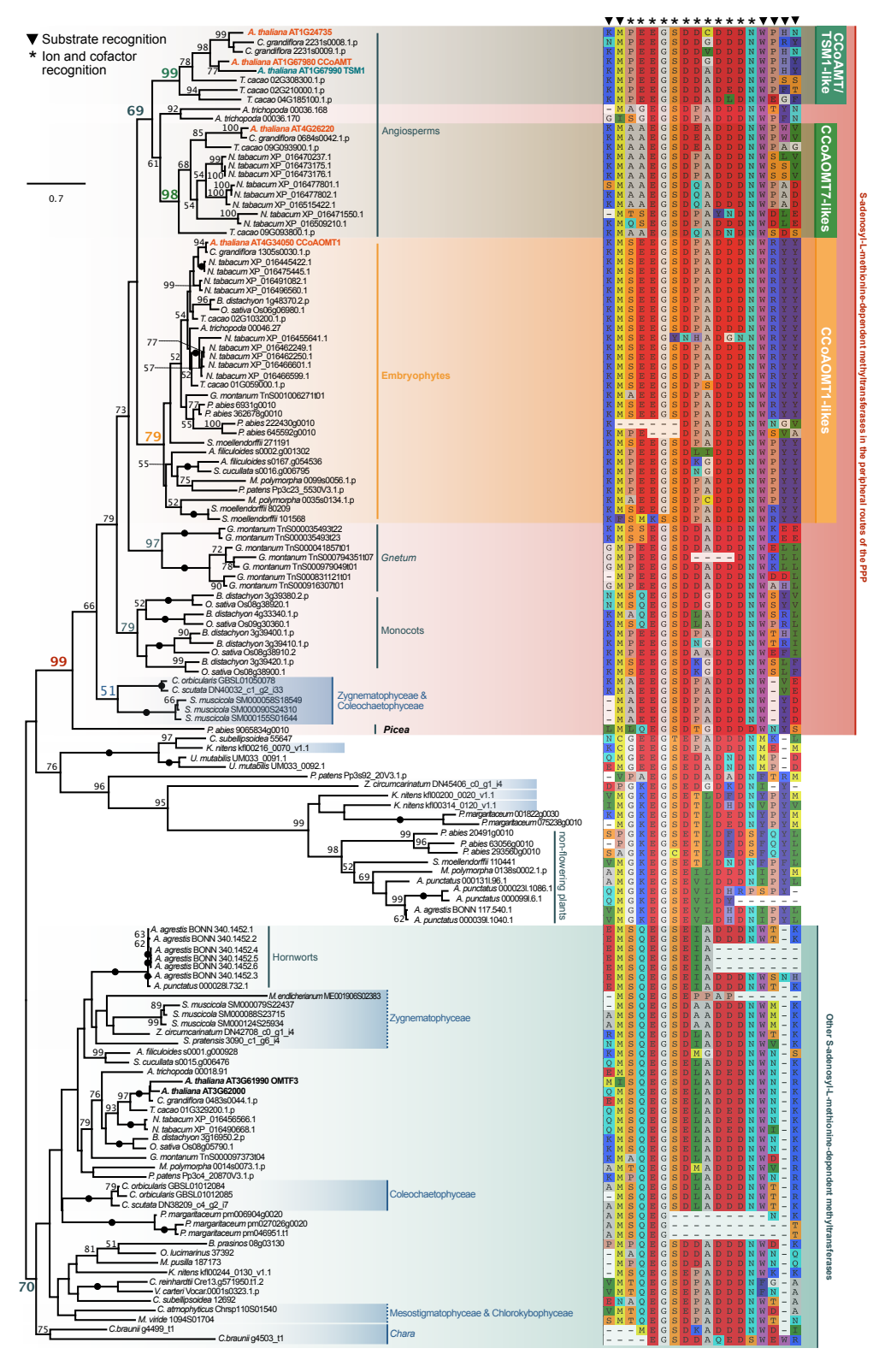

0.4
Chloroplastida CCR/DFR-likes
Gymnosperms
Spermatophytes
Angiosperms
Anthocyanidine reductase
DFR
BEN
Ferns
TKPR1
Hornworts
Embryophytes
Spermatophytes
Ferns
TKPR2
Bryophytes & Selaginella
Euphyllophytes
Embryophytes
putrescine/spermidine dehydrogenase
Angiosperms
Zygnematophyceae
CRL-A
A1
A2
A3
Angiosperms
Embryophyta
non-seed plants
CRL-B
B1
B2
Angiosperms
Angiosperms
Tracheophytes
Monocots
CCR
Dicots
Monocots
Monocots
Monocots
Tracheophytes
CRL-C
C1
C2
Spermatophytes
Spermatophytes
Tracheophytes
Embryophytes
Chlorophytes
Chlorophytes
Zygnematophyceae
FLDH
Coleochaetophyceae
CCR/CRL ?

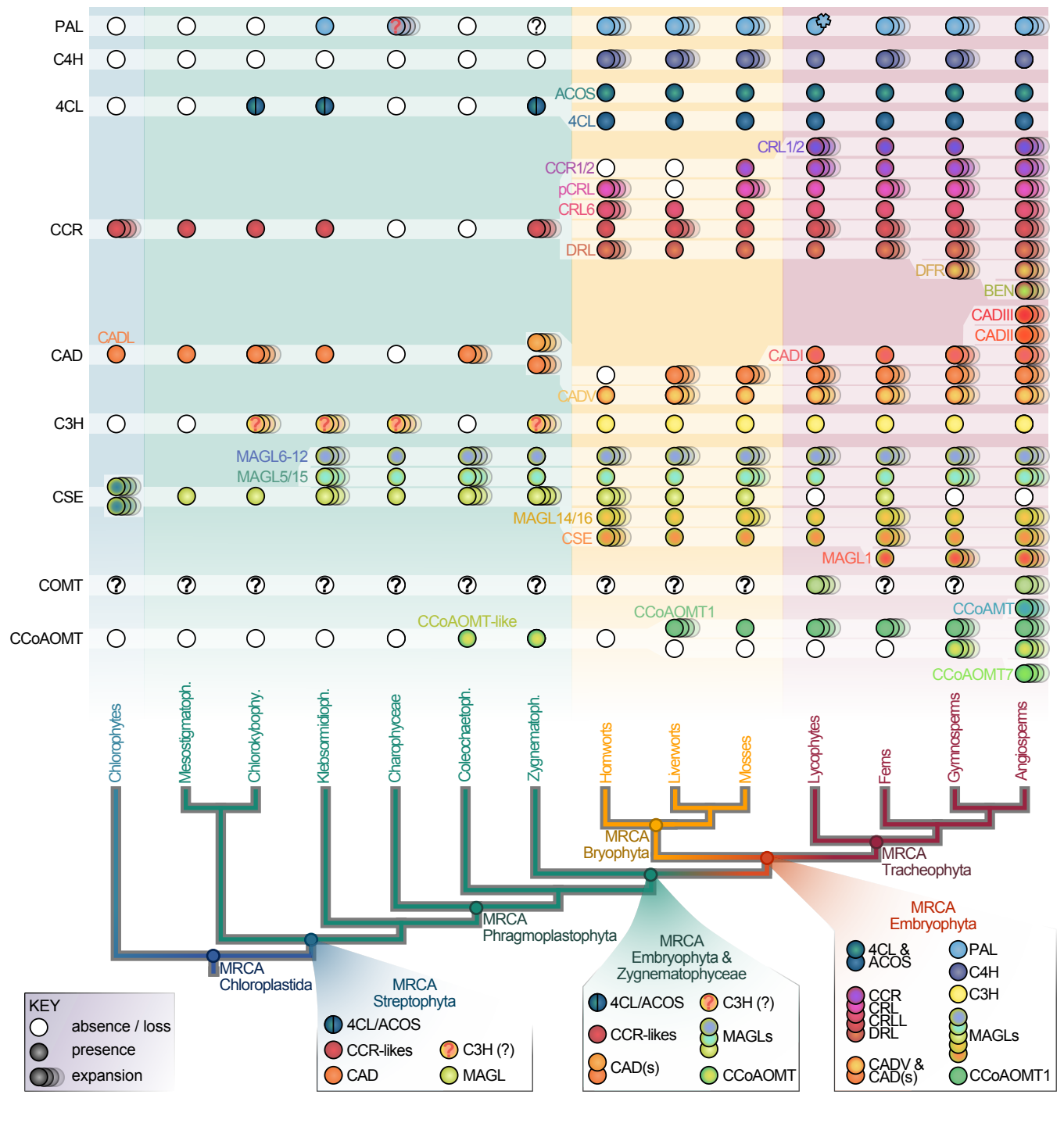

0.4
Substrate binding pocket
NADP+ binding
Zn2+ binding
Dicots
Angiosperms 1
Angiosperms 2
Angiosperms
Ferns
Zygnematophyceae
Zygnematophyceae & Coleochaetophyceae
streptophyte algae
Chlorophytes
Zygnematophyceae & Coleochaetophyceae & Chlorokybophyceae
Zygnematophyceae & Coleochaetophyceae & Klebsormidiophyceae & Chlorokybophyceae
Spermatophytes
Tracheophytes
Embryophytes & Zygnematophyceae
Embryophytes
CADII
CADIII
CADIV
CADI
CADV
Spermatophytes
Tracheophytes
Embryophytes
Chloroplastida CAD-likes
GroES-like zinc-binding dehydrogenase family proteins

0.5
A. agrestis BONN 228.160.1
A. punctatus 000126l.243.1
P. patens Pp3c22_19010V3.1.p
M. polymorpha 0037s0087.1.p
S. moellendorffii 271465
A. filiculoides s0355.g066730
S. cucullata s0001.g000031
P. abies 109548g0010
G. montanum TnS000810643t02
A. trichopoda 00040.62
A. trichopoda 00101.79
CYP98A with high 4-coumaroyl-anthranilate activity?
79
65
sinapoylspermidine
A. thaliana AT1G74540 CYP98A8
C. grandiflora 0402s0071.1.p
A. thaliana AT1G74550 CYP98A9
C. grandiflora 0402s0070.1.p
N. tabacum XP_016442382.1
N. tabacum XP_016506467.1
N. tabacum XP_016443317.1
N. tabacum XP_016443318.1
N. tabacum XP_016446948.1
N. tabacum XP_016446949.1
N. tabacum XP_016470640.1
N. tabacum XP_016479719.1
A. thaliana AT2G40890 C3H (CYP98A3)
C. grandiflora 1642s0055.1.p
99
58
N. tabacum XP_016476077.1
N. tabacum XP_016476078.1
N. tabacum XP_016482834.1
B. distachyon 2g21300.1.p
O. sativa Os05g41440.1
Embryophyte CYP98A
A. trichopoda 00181.7
90
A. thaliana AT3G26300 CYP71B34
51
A. thaliana AT3G26310 CYP71B35
C. grandiflora 8404s0001.1.p
C. grandiflora 2848s0009.1.p
C. grandiflora 0104s0022.1.p
C. grandiflora 1714s0003.1.p
B. distachyon 1g15190.1.p
O. sativa Os03g39760.1
82
A. filiculoides s0001.g000331
S. cucullata s0015.g006592
A. filiculoides s0422.g069198
Angiosperms
CYP71B
A. thaliana AT2G45550 CYP76C4
99
A. thaliana AT2G45560 CYP76C1
87
C. grandiflora 0698s0005.1.p
C. grandiflora 0698s0004.1.p
G. montanum TnS000818903t01
S. cucullata s0074.g017295
Spermato-phytes
CYP76C
A. thaliana AT5G07990 CYP75B1/TT7
C. grandiflora 1365s0004.1.p
84
B. distachyon 1g17180.1.p
89
B. distachyon 3g04750.1.p
O. sativa Os10g17260.1
92
B. distachyon 1g24840.1.p
B. distachyon 4g16560.1.p
O. sativa Os10g16974.1
G. montanum TnS000091781t07
G. montanum TnS000145195t03
G. montanum TnS000145195t04
Spermatophytes
CYP75B/TT7
B. distachyon 1g57000.1.p
52
B. distachyon 4g31367.1.p
68
B. distachyon 4g31380.1.p
O. sativa Os09g26960.1
86
O. sativa Os09g26980.1
O. sativa Os08g35510.1
72
O. sativa Os09g26940.1
88
O. sativa Os09g08990.1
97
O. sativa Os02g29960.1
B. distachyon 5g08374.1.p
B. distachyon 5g08570.1.p
Monocots
91
A. trichopoda 00066.179
C. grandiflora 4290s0011.1.p
A. agrestis BONN 117.3202.1
A. punctatus 000122l.139.1
A. punctatus 000124l.47.1
S. moellendorffii 183227
S. moellendorffii 271925
54
M. polymorpha 0001s0289.1.p
M. polymorpha 0094s0036.1.p
P. abies 53288g0010
78
P. abies 10434848g0010
S. moellendorffii 169986
S. moellendorffii 228641
M. polymorpha 0031s0060.1.p
M. polymorpha 0086s0077.1.p
54
P. patens Pp3c12_8440V3.1.p
P. patens Pp3c1_39440V3.1.p
non-flowering plants
A. trichopoda 00092.5
B. distachyon 3g14390.1.p
94
O. sativa Os08g03682.1
99
C. grandiflora 1968s0071.1.p
99
A. filiculoides s0004.g008667
S. cucullata s0019.g007724
S. cucullata s0115.g020980
Euphyllophytes
99
A. punctatus 000122l.186.1
97
M. polymorpha 0197s0009.1.p
95
58
M. polymorpha 0142s0024.1.p
72 Anthoceros sp. CYP450s
Bryophytes
A. agrestis BONN 369.483.1
A. punctatus 000148l.58.1
P. patens Pp3c8_21260V3.1
P. patens Pp3c24_17830V3.1
85
N. tabacum XP_016461598.1
97
96
P. abies 31070g0010
66
P. abies 541418g0010
P. abies 43714g0010
P. abies 137345g0010
A. filiculoides s0367.g067007
Embryophytes
S. pratensis 1558_c0_g1_i1
S. pratensis 3443_c0_g1_i6
87
S. pratensis 3243_c0_g2_i8
S. pratensis 3088_c0_g1_i2
Z. circumcarinatum DN49507_c5_g2_i1
Zygnematalean CYP450s
B. distachyon 2g53470.1.p
98
N. tabacum XP_016468779.1
99
N. tabacum XP_016509005.1
N. tabacum XP_016493546.1
99
N. tabacum XP_016505655.1
G. montanum TnS000953283t02
Spermatophytes
K. nitens 00035_0080_v1.1
78
K. nitens 00038_0230_v1.1
K. nitens 00588_0070_v1.1
Klebsormidium CYP450s
51
52
C. braunii g2943_t1
C. braunii g12412_t1
96
99
C. braunii g12414_t1
57
C. braunii g12415_t1
88
C. braunii g12428_t1
98
C. braunii g12416_t1
C. braunii g23074_t1
Chara CYP450s
K. nitens 00209_0130_v1.1
C. atmophyticus Chrsp31S08934
C. atmophyticus Chrsp61S07011
C. atmophyticus Chrsp180S02881
86
C. atmophyticus Chrsp180S02882
Chlorokybus CYP450s
Streptophyte CYP450s
C. orbicularis GBSL01045479
C. orbicularis GBSL01052004
87 algal seqs.


0.8
249 mammal & Sauria MAGLs
GMVLISPLVLAA
Streptophyte algal MAGLs
Streptophyte MAGL9/12
MAGL9
MAGL12
GAVLVAPMCKIS
Streptophyte MAGL6/7/8/9/10/11/12
MAGL7
MAGL8
MAGL6
MAGL10
MAGL11
Streptophyte algal MAGLs
Streptophyte MAGL5/4/5
Angiosperms
MAGL15
Spermatophytes
MAGL5
Embryophytes
GAVLVAPMCKIA
Streptophyte algal MAGLs
Embryophyte MAGL1/3
CSE/MAGL3
GLIESAPLEVLP
Spermatophytes
MAGL1
GAVLVAPMCKIS
Embryophyte MAGL14/16
MAGL14
MAGL16
GEVLNGAMCGIS
Chlorophyte MAGLs
Chlorophyte MAGLs
Streptophyte algal MAGLs
Streptophyte MAGL2/4/13
MAGL2
MAGL4
MAGL13




▼ Substrate recognition
* Ion and cofactor recognition
A. thaliana AT1G24735
C. grandiflora 2231s0008.1.p
C. grandiflora 2231s0009.1.p
A. thaliana AT1G67980 CCoAMT
A. thaliana AT1G67990 TSM1
T. cacao 02G308300.1.p
T. cacao 02G210000.1.p
T. cacao 04G185100.1.p
A. trichopoda 00036.168
A. trichopoda 00036.170
Angiosperms
A. thaliana AT4G26220
C. grandiflora 0684s0042.1.p
T. cacao 09G093900.1.p
N. tabacum XP_016470237.1
N. tabacum XP_016473175.1
N. tabacum XP_016473176.1
N. tabacum XP_016477801.1
N. tabacum XP_016477802.1
N. tabacum XP_016515422.1
N. tabacum XP_016471550.1
N. tabacum XP_016509210.1
T. cacao 09G093800.1.p
A. thaliana AT4G34050 CCoAOMT1
C. grandiflora 1305s0030.1.p
N. tabacum XP_016445422.1
N. tabacum XP_016475445.1
N. tabacum XP_016491082.1
N. tabacum XP_016496560.1
B. distachyon 1g48370.2.p
O. sativa Os06g08980.1
T. cacao 02G103200.1.p
A. trichopoda 00046.27
N. tabacum XP_016455641.1
N. tabacum XP_016462249.1
N. tabacum XP_016462250.1
N. tabacum XP_016466601.1
N. tabacum XP_016466599.1
T. cacao 01G059000.1.p
Embryophytes
G. montanum TnS001006271t01
P. abies 6931g0010
P. abies 3626789g0010
P. abies 222430g0010
P. abies 645592g0010
S. moellendorffii 271191
A. filiculoides s0002.g001302
A. filiculoides s0167.g054536
S. cucullata s0016.g006795
M. polymorpha 0099s0056.1.p
P. patens Pp3c23_5530V3.1.p
M. polymorpha 0035s0134.1.p
S. moellendorffii 80209
S. moellendorffii 101568
G. montanum TnS000035493t22
G. montanum TnS000035493t23
G. montanum TnS000041857t01
G. montanum TnS000794351t07
G. montanum TnS000979049t01
G. montanum TnS000831121t01
G. montanum TnS000916307t01
Gnetum
B. distachyon 3g39380.2.p
O. sativa Os08g38920.1
B. distachyon 4g33340.1.p
O. sativa Os09g30360.1
B. distachyon 3g39400.1.p
B. distachyon 3g39410.1.p
O. sativa Os08g38910.2
B. distachyon 3g39420.1.p
O. sativa Os08g38900.1
Monocots
C. orbicularis GBSL01050078
C. scutata DN4032_c1_g2_i33
S. muscicola SM000058S18549
S. muscicola SM000090S24310
S. muscicola SM000155S01644
Zygnematophyceae & Coleochaetophyceae
Picea
P. abies 9065834g0010
C. subellipsoidea 55647
K. nitens kfl00216_0070_v1.1
U. mutabilis UM033_0091.1
U. mutabilis UM033_0092.1
P. patens Pp3s92_20V3.1.p
Z. circumcarinatum DN45406_c0_g1_i4
K. nitens kfl00200_0020_v1.1
K. nitens kfl00314_0120_v1.1
P. margaritaceum 001822g0030
P. margaritaceum 075238g0010
P. abies 20491g0010
P. abies 630569g0010
P. abies 293560g0010
S. moellendorffii 110441
M. polymorpha 0138s0002.1.p
A. punctatus 000131l.96.1
A. punctatus 000023l.1086.1
A. punctatus 000099l.6.1
A. agrestis BONN 117.540.1
A. punctatus 000039l.1040.1
non-flowering plants
A. agrestis BONN 340.1452.1
A. agrestis BONN 340.1452.2
A. agrestis BONN 340.1452.4
A. agrestis BONN 340.1452.5
A. agrestis BONN 340.1452.6
A. agrestis BONN 340.1452.3
A. punctatus 000028l.732.1
Hornworts
M. endlicherianum ME001906S02383
S. muscicola SM000079S22437
S. muscicola SM000088S23715
S. muscicola SM000124S25934
Z. circumcarinatum DN42708_c0_g1_i4
S. pratensis 3090_c1_g6_i4
Zygnematophyceae
A. filiculoides s0001.g000928
S. cucullata s0015.g006476
A. trichopoda 00018.91
A. thaliana AT3G61990 OMTF3
A. thaliana AT3G62000
C. grandiflora 0483s0044.1.p
T. cacao 01G329200.1.p
N. tabacum XP_016456566.1
N. tabacum XP_016490668.1
B. distachyon 3g16950.2.p
O. sativa Os08g05790.1
G. montanum TnS000097373t04
M. polymorpha 0014s0073.1.p
P. patens Pp3c4_20870V3.1.p
C. orbicularis GBSL01012084
C. orbicularis GBSL01012085
C. scutata DN38209_c4_g2_i7
Coleochaetophyceae
P. margaritaceum pm006904g0020
P. margaritaceum pm027026g0020
P. margaritaceum pm046951.t1
B. prasinos 08g03130
O. lucimarinus 37392
M. pusilla 187173
K. nitens kfl00244_0130_v1.1
C. reinhardtii Cre13.g571950.t1.2
V. carteri Vocar.0001s0323.1.p
C. subellipsoidea 12692
C. atmophyticus Chrsp110S01540
Mesostigmatophyceae & Chlorokybophyceae
M. viride 1094S01704
C.braunii g4499_t1
C.braunii g4503_t1
Chara
0.7
CCoAMT/ TSM1-like
CCoAOMT7-likes
CCoAOMT1-likes
S-adenosyl-L-methionine-dependent methyltransferases in the peripheral routes of the PPP
Other S-adenosyl-L-methionine-dependent methyltransferases


PAL
C4H
4CL
CCR
CAD
C3H
CSE
COMT
CCoAOMT
ACOS
4CL
CRL1/2
CCR1/2
pCRL
CRL6
DRL
DFR
BEN
CADIII
CADII
CADI
CADV
CADL
MAGL6-12
MAGL5/15
MAGL14/16
CSE
MAGL1
CCoAOMT1
CCoAOMT-like
CCoAMT
CCoAOMT7
Chlorophytes
Mesostigmatoph.
Chlorokybophy.
Klebsormidioph.
Charophyceae
Coleochaetoph.
Zygnematoph.
Hornworts
Liverworts
Mosses
Lycophytes
Ferns
Gymnosperms
Angiosperms
MRCA Bryophyta
MRCA Tracheophyta
MRCA Phragmoplastophyta
MRCA Chloroplastida
MRCA Streptophyta
4CL/ACOS
CCR-likes
CAD
C3H (?)
MAGL
MRCA Embryophyta & Zygnematophyceae
4CL/ACOS
CCR-likes
CAD(s)
C3H (?)
MAGLs
CCoAOMT
MRCA Embryophyta
4CL & ACOS
CCR CRL CRLL DRL
CADV & CAD(s)
PAL
C4H
C3H
MAGLs
CCoAOMT1
KEY
absence / loss
presence
expansion