

# Evolution of chlorophyll degradation is associated with plant transition to land

Isabel Schumacher<sup>†</sup>, Damian Menghini<sup>†</sup>, Serguei Ovinnikov, Mareike Hauenstein, Niklaus Fankhauser, Cyril Zipfel , Stefan Hörtensteiner<sup>‡</sup> and Sylvain Aubry\* 

Department of Plant and Microbial Biology, University of Zürich, Zürich 8008, Switzerland

Received 2 November 2021; revised 9 December 2021; accepted 15 December 2021; published online 21 December 2021.

\*For correspondence (e-mail sylvain.aubry@botinst.uzh.ch).

<sup>†</sup>These authors are contributed equally to this work.

<sup>‡</sup>Deceased.

## SUMMARY

Colonization of land by green plants (Viridiplantae) some 500 million years ago was made possible by large metabolic and biochemical adaptations. Chlorophyll, the central pigment of photosynthesis, is highly photo-active. In order to mitigate deleterious effects of pigment accumulation, some plants have evolved a coordinated pathway to deal with chlorophyll degradation end-products, so-called phyllobilins. This pathway has been so far mostly unravelled in *Arabidopsis thaliana*. Here, large-scale comparative phylogenomic coupled to an innovative biochemical characterization strategy of phyllobilins allow a better understanding of how such a pathway appeared in Viridiplantae. Our analysis reveals a stepwise evolution of the canonical pheophorbide *a* monooxygenase/phyllobilin pathway. It appears to have evolved gradually, first in chlorophyte's chloroplasts, to ensure multicellularity by detoxifying chlorophyll catabolites, and in charophytes outside chloroplasts to allow adaptation of embryophytes to land. At least six out of the eight genes involved in the pathway were already present in the last common ancestor of green plants. This strongly suggests parallel evolution of distinct enzymes catalysing similar reactions in various lineages, particularly for the dephytylation step. Together, our study suggests that chlorophyll detoxification accompanied the transition from water to land, and was therefore of great importance for plant diversification.

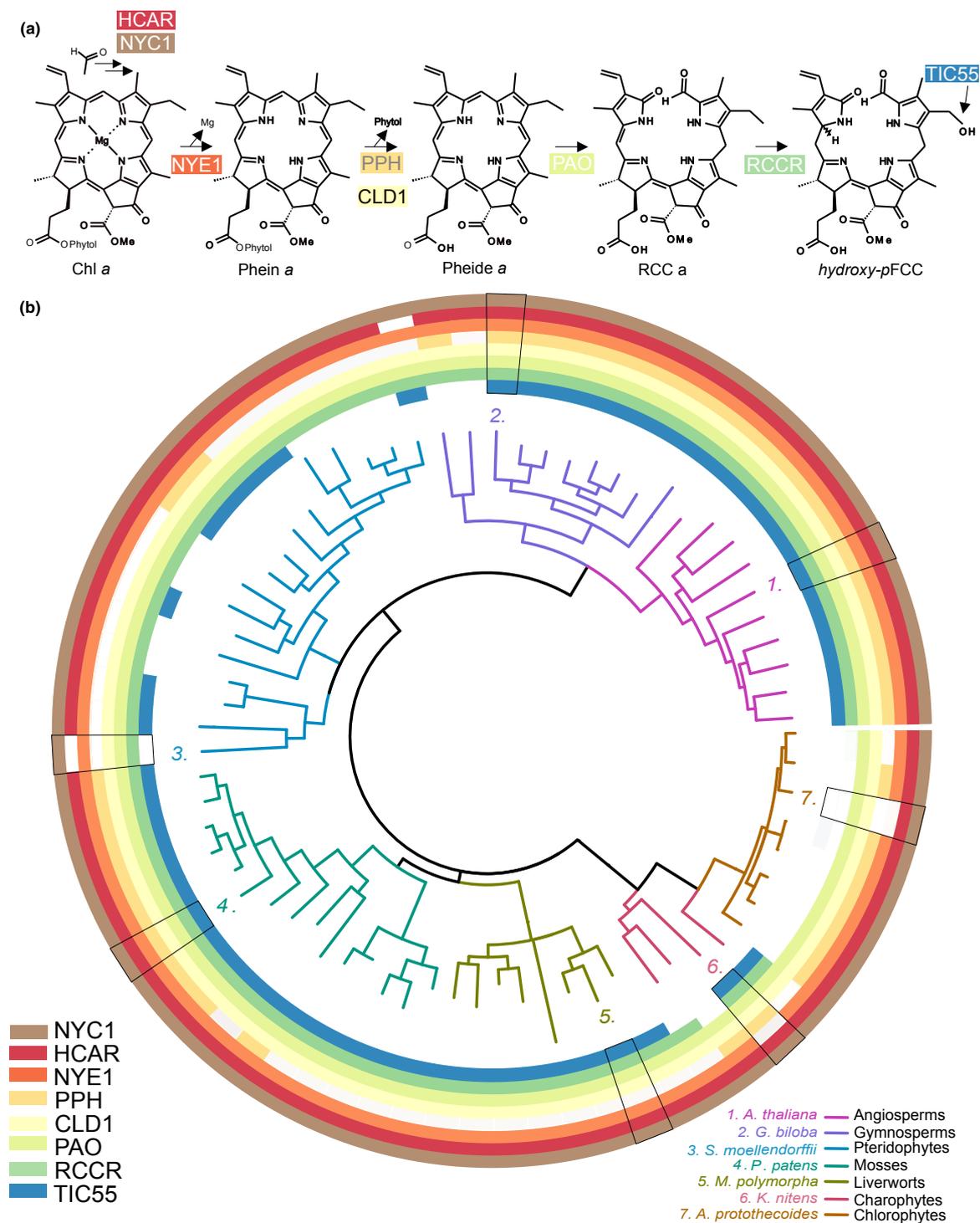
**Keywords:** chlorophyll degradation, phylogenomics, orthology inference, phyllobilins, water-to-land transition.

## INTRODUCTION

Photosynthetic organisms on Earth are largely dominated by green plants (Viridiplantae). They comprise two major clades: embryophytes (land plants) and algae (chlorophytes and streptophyte algae/charophytes). Embryophytes are thought to be derived from a single clade within streptophytes and diverged some 450–500 million years ago (Wickett et al., 2014; de Vries and Archibald, 2018). The photosynthetic apparatus is derived from ancestral endosymbiotic cyanobacteria, which use chlorophyll as one of the main pigments in their photosystems and light-harvesting complexes (Xiong and Bauer, 2002). While our understanding of the biochemistry and evolution of chlorophyll biosynthesis is fairly comprehensive (Tanaka and Tanaka, 2006; Chew and Bryant, 2007), the pigment's turnover and degradation have mostly been characterized in angiosperms with no actual understanding of when the pathway appeared and whether these processes are conserved in other plants (Kuai et al., 2017). Despite

chlorophyll degradation being one of the most visually striking biochemical reactions on Earth, with an estimated one billion tons turnover yearly (Hendry et al., 1987), its occurrence outside angiosperms and the associated evolutionary advantages to degrade chlorophyll using a tightly controlled enzymatic pathway are relatively unclear.

Most of the experimental work deciphering the pathway responsible for chlorophyll degradation has been performed in angiosperms, namely on senescing leaves of *Arabidopsis thaliana* (hereafter *Arabidopsis*; Hörtensteiner et al., 2019) and a few other ripening fruits of crop plants (Gómez et al., 2013). It is referred to as the 'pheophorbide *a* oxygenase (PAO)/phyllobilin pathway' (Kuai et al., 2017; Hörtensteiner et al., 2019; Figure 1a). In green leaves of *Arabidopsis*, chlorophyll *a* and *b* are embedded in the photosystem and light-harvesting antennas. The pigment is subject to an active turnover in the so-called 'chlorophyll cycle' (Tanaka and Tanaka, 2019). Upon leaf senescence, while thylakoids start to disassemble, the chlorophyll



**Figure 1.** Evolution of the canonical PAO/phyllobilin chlorophyll degradation pathway in Viridiplantae.

(a) The pathway of chlorophyll degradation in *Arabidopsis* leaves. Only early steps of the chloroplastic portion of the PAO/phyllobilin pathway that is further studied here are shown. TIC55 catalyses hydroxylation of pFCC in hydroxy pFCC. Further subsequent steps leading to the (hydroxy) pFCC to the chloroplast are detailed here: Kuai et al. (2017).

(b) Presence of gene homologues encoding chlorophyll degradation across 72 plant and algal species. Each concentric ring indicates an enzyme, and the presence (coloured) or absence (empty) of a homologue for that particular gene. Identity of species contributing to the phylogenetic tree with full species names given in Figure S1. NYC1, NONYELLOW COLORING1, CHLOROPHYLL *b* REDUCTASEs; HCAR, HYDROXY-CHL *a* REDUCTASE; NYE1, NON YELLOWING 1; PPH, PHEOPHYTINASE; PAO, PHEOPHORBIDE *a* MONO-OXYGENASE; RCCR, RCC REDUCTASE; TIC55, TRANSLOCON AT THE INNER CHLOROPLAST ENVELOPE 55; Phein *a*, pheophytin *a*; Pheide *a*, pheophorbide *a*; RCC, red chlorophyll catabolite; pFCC, primary FCC.

degradation pathway starts. It is generally considered to be divided into two successive phases: first, the release of the pigment from the thylakoid membrane up to the opening of the porphyrin ring; and second, the detoxification and transport of the linear tetrapyrrole until its final storage in the vacuole.

The first committed step to chlorophyll degradation is the conversion of chlorophyll *b* to *a* by a two-step reaction catalysed by NONYELLOW COLORING 1 (NYC1) and HYDROXYMETHYL CHLOROPHYLL *a* REDUCTASE (HCAR; Kusaba et al., 2007; Horie et al., 2009; Sato et al., 2009; Meguro et al., 2011). The pigment is then subsequently processed by the Mg-dechelatase NON YELLOWING (NYE1; Shimoda et al., 2016), the dephytylase PHEOPHYTYLASE (PPH; Schelbert et al., 2009), and the PHEOPHORBIDE *A* OXYGENASE (PAO) that catalyses the irreversible opening of the porphyrin ring (Pruzinska et al., 2003). After these coordinated reactions, which leave no detectable trace of intermediaries under normal physiological conditions, a first linear tetrapyrrole is produced: the red chlorophyll catabolite (RCC; Rodoni et al., 1997). In *Arabidopsis*, it is directly converted to a primary fluorescent catabolite (pFCC) by the RED CHLOROPHYLL CATABOLITE REDUCTASE (RCCR; Mühlecker et al., 1997; Hörtensteiner et al., 2000; Wuthrich et al., 2000). pFCCs are then modified and further exported from the chloroplast. While the first part of the pathway appears relatively straightforward, the fate of pFCC outside the chloroplast appears more diverse and involves various detoxifying enzymes. For example, in *Arabidopsis*, CYP89a9 and MES16 catalyse C1 deformylation and C8<sup>4</sup> demethylation, respectively (Christ et al., 2012, 2013, 2016). Alone in *Arabidopsis*, nine distinct phyllobilins – defined as any linear tetrapyrrole that is derived from chlorophyll – have been characterized to date (Kräutler and Hörtensteiner, 2014), suggesting many more plant phyllobilin-modifying enzymes are still to be identified (Kräutler, 2016).

From a physiological point of view, maintaining a complex and costly enzymatic pathway to degrade chlorophyll in senescent leaves may allow nitrogen remobilization from both the pigment itself and their chlorophyll-binding proteins (Hörtensteiner and Feller, 2002; Wang and Grimm, 2021). Indeed, about two-thirds of leaf nitrogen is localized in chloroplast. However, the nature of the exact link between the pigment degradation and other senescing processes remains elusive. Importantly, leaf senescence appears to have evolved relatively recently in plants (Thomas et al., 2009), and nitrogen remobilization is unlikely to be a conserved evolutionary trigger for degrading chlorophyll in all chlorophyll-containing species. Another possible reason to actively degrade chlorophyll is linked to its phototoxic properties. Some mutants impaired in the PAO/phyllobilin pathway accumulate toxic intermediates: accumulation of pheophorbide *a* or RCC strongly impact

tissue integrity (Pruzinska et al., 2003, 2005; Aubry et al., 2020), while the end-products, mostly phyllobilins and monopyrroles, are photoinactive (Kräutler, 2016).

Our understanding of the PAO/phyllobilin pathway is therefore largely biased and limited to our knowledge of the pathway in angiosperms. For example, the vast majority of phyllobilins detected in angiosperms appear to be structurally related (with opening of the porphyrin ring at the 'northern' C1 position); therefore, suggesting that the PAO/phyllobilin pathway of degradation is largely conserved (Hörtensteiner et al., 2019). Some recent data showing the presence of phyllobilins (*iso*-phyllobilanones) in ferns (*Pteridium aquilinum*) suggest that the PAO/phyllobilin pathway is present outside angiosperms (Ehart et al., 2018). Indeed, chlorophyll catabolites were identified in the green alga *Auxenochlorella protothecoides* (Gossauer, 1994). Chlorophyll degradation might even have evolved outside the green lineage, in marine protists like the dinoflagellates *Pyrocystis lunula* (Wu et al., 2003). In these two particular cases, the exact nature of the enzymes involved remains to be determined. In *P. lunula*, the opening of the ring that produces luciferin occurs at a different position of the porphyrin ring as compared with plant PAO products (Wu et al., 2003). This is likely due to an independent evolution of the chlorophyll degradation pathway in this clade. However, *A. protothecoides* accumulates RCC-like compounds when grown heterotrophically. It excretes RCCs into the surrounding media, which therefore strongly suggests a PAO-like activity in this alga (Hörtensteiner et al., 2000).

Taking advantage of recent advances in phylogenomics; especially, the increased number of high-quality genomes and transcriptomes available, we performed a comparative analysis of homologous genes involved in the chlorophyll degradation in the green lineage (Cannell et al., 2020; Harris et al., 2020). Orthologues are genes that show some homology, and diverged after a species split rather than after duplication (Fitch, 1970; Emms and Kelly, 2015). We merged and computed genomes and transcriptomes from a subset of 72 species representative of seven major lineages in plant evolution (Figure S1). We based our analysis of potential homologues of the known chlorophyll degradation *Arabidopsis* proteins. In addition, we performed targeted metabolomics to detect the end-products of this pathway (phyllobilins) in some representative species of these lineages.

We aimed here to determine when during plant evolution a dedicated pathway to degrade chlorophyll might have evolved. It remains unclear whether a unique strategy (i.e. the PAO/phyllobilin pathway) for chlorophyll catabolites detoxification has been recruited. Eventually, we seek to understand the physiological relevance of detoxifying chlorophyll, and identify families or lineages where potential further studies could shed more light on the chlorophyll degradation pathway's evolutionary trajectory.

## RESULTS

### Identification of chlorophyll catabolic enzymes in plants by orthologue's inference

Despite some pioneering work in barley and fescue, the vast majority of the biochemistry of the PAO/phyllobilin pathway has been characterized in Arabidopsis. Recent works highlight the potential of orthology inference, using tools like Orthofinder (Emms and Kelly, 2020), to better understand the evolution of particular traits like stomata development (Harris et al., 2020) or various specialized metabolic pathways (Cannell et al., 2020). Here, using a similar approach, we performed orthology inference on 72 land plants and algae that have their genomes sequenced, and represent each of the major lineages in plant evolution: angiosperms, gymnosperms, lycophytes, bryophytes (mosses and liverworts), charophytes and chlorophytes (Figure 1b). Out of these species, 87.3% of genes could be assigned to 24 479 orthogroups. Then, orthogroups containing potential orthologues for eight major enzymes involved in the PAO/phyllobilin pathway were identified. Focusing on each enzymatic step at a time, we then checked for the presence/absence of orthologues in Viridiplantae and inferred the pathway evolution in these species (Figure 1b).

### The 'chlorophyll cycle' enzymes are well conserved throughout the green lineage

Conversion of chlorophyll *a* to *b* and its reverse reaction is catalysed by a set of four enzymes working in the so-called 'chlorophyll cycle'. Chlorophyll *a* is converted into chlorophyll *b* by CHLOROPHYLL A OXYGENASE (CAO), that is often also considered the 'last' step of chlorophyll synthesis. Given that chlorophyll *b* is also found in cyanobacteria (Satoh et al., 2001), it was expected that all the genes responsible for its synthesis would be present in all eukaryotic photosynthetic organisms. The vast majority of genes involved in chlorophyll synthesis, including CAO, are indeed present in all 72 species analysed here (Figure S2; supporting data in Table S1).

The chlorophyll *b* to *a* conversion is performed in a two-step process: NYC1 reduces chlorophyll *b* to 7-hydroxymethyl chlorophyll, and HCAR subsequently converts this intermediate into chlorophyll *a* (Zhao et al., 2020). Here, we focused on the initial step of chlorophyll degradation, namely its conversion from chlorophyll *b* to *a* and catalyzed by two enzymes: NYC1 and HCAR (Shimoda et al., 2012). Indeed, NYE1 and PAO do not accept 'b' pigments as substrates (Shimoda et al., 2016). Our analysis identifies orthologues of these two enzymes throughout the entire Viridiplantae (Figure 1b and Figure S3). Interestingly, a paralogue (i.e. a gene that derived from the same ancestral gene by duplication) of NYC1, NYC-LIKE 1 (NOL), belongs to the same orthogroup but is

phylogenetically clearly distinct from NYC1: it is absent from chlorophytes, suggesting NOL might have originated from a relatively recent gene duplication event (Figure S3).

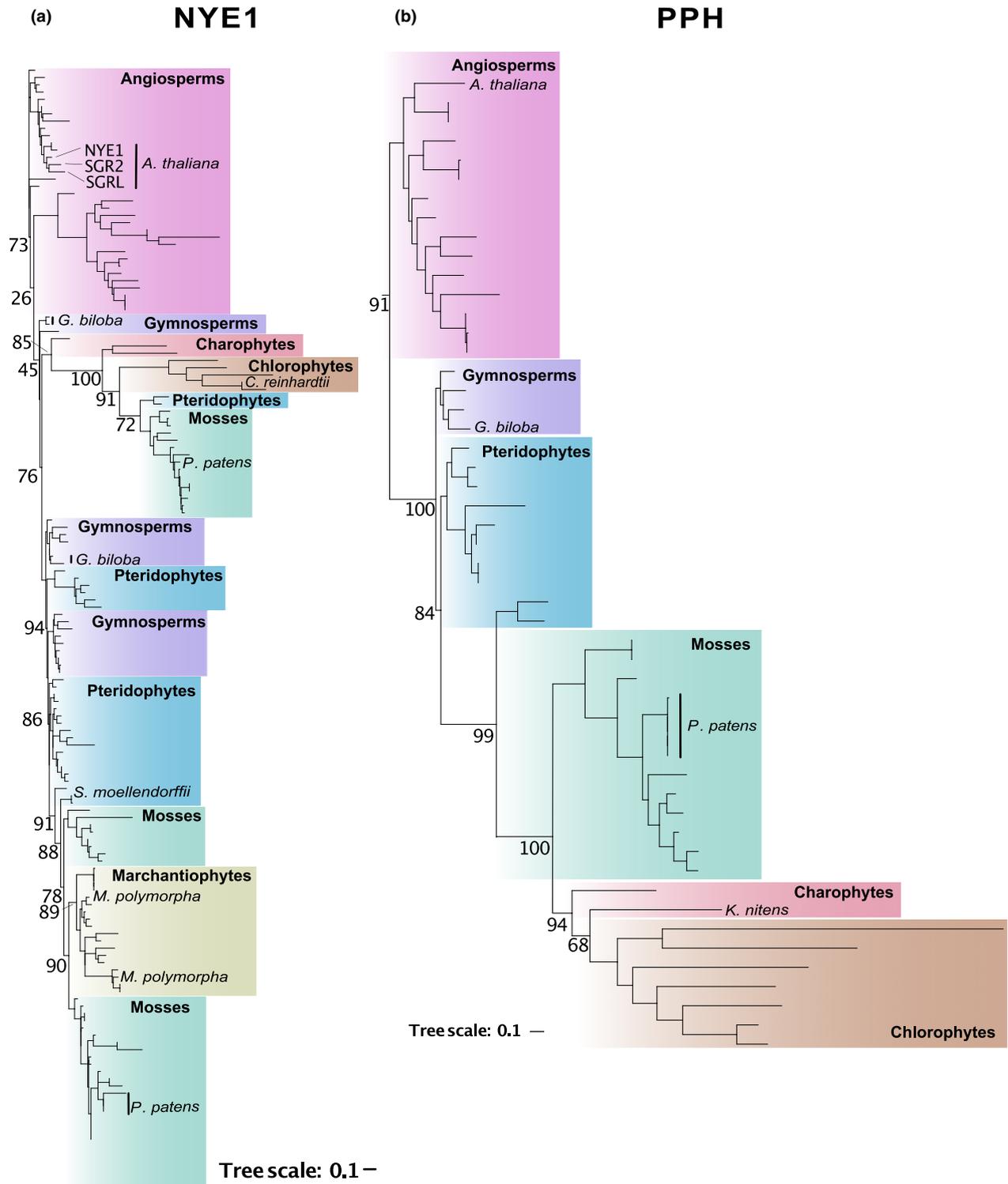
### Evolution of the Mg-dechelation

The *STAYGREEN* gene (*SGR*, also referred to as *NYE1*) was one of the first genetic loci indicative of a pathway responsible for chlorophyll degradation (Armstead et al., 2006, 2007; Aubry et al., 2008). Its role as a Mg-dechelation has only been shown recently (Shimoda et al., 2016). *NYE1* genes were identified in all Viridiplantae lineages (Figure 1b). Intriguingly, only two species did not show *NYE1* orthologues: the charophyte *Klebsormidium nitens* and the chlorophyte *A. protothecoides*. Given the presence of phyllobilins in *A. protothecoides* (Figure S4), and the fact that *NYE1* is present in all other green algae studied (both charophytes and chlorophytes), these absences are likely due to uneven genome quality/assembly and not to gene loss (Cannell et al., 2020). In Arabidopsis, three paralogues (*NYE1*, *STAY GREEN 2*; *SGR2* and *STAYGREEN LIKE*; *SGL*) are involved in the dechelation reaction and might differ in their substrate specificity (Sakuraba et al., 2014; Lin et al., 2016). Phylogenetic analysis suggests they all derived from a recent gene duplication event (Figure 2a). Most species analysed here show more than one copy of the *NYE1* gene (e.g. four copies in *Ginkgo biloba*; Figure 2a). While the enzymatic steps immediately preceding and succeeding the Mg-dechelation are apparently being encoded by single copy genes, the biochemical reason for having multiple copies of this particular gene remains unclear but might be related in some cases to the maintenance of photosystem homeostasis.

### Multiple enzymes might have been recruited for dephytylation reactions during evolution

The subsequent step to the Mg-dechelation in Arabidopsis is the removal of the phytol chain from pheophytin, which is catalysed by PPH (Schelbert et al., 2009). Removal of the phytol chain is a key step that allows the release of the pigment from the apoproteins. Impairments in PPH induce a stay-green phenotype in several angiosperms like Arabidopsis (Schelbert et al., 2009), *Lolium perenne* (Zhang et al., 2016) and pear (*Pyrus communis*; Cheng and Guan, 2014). Despite being functionally conserved in all angiosperms studied, our analysis shows that PPH is not uniformly present among all Viridiplantae. There is no PPH homolog in pteridophytes, lycophytes, hornworts and liverworts (Figure 1b). Interestingly, some streptophyte algae and chlorophytes apparently do contain PPH homologues (Figure 2b).

Because the lycophyte *Selaginella moellendorffii* and the liverwort *Marchantia polymorpha* are both able to produce phyllobilins (Table 1; Figure S5), it is therefore possible that these species express a distinct dechelation enzyme



**Figure 2.** Evolution of NYE1 and PPH in the Viridiplantae. (a) Maximum likelihood phylogenetic tree of NYE1. In *Arabidopsis*, NYE1 paralogues SGR2 and SGRL appear to be the result of relatively recent duplications events. Interestingly, the chlorophyte NYE1 orthologue in *Chlamydomonas reinhardtii*, shown to be involved in chlorophyll dephytylation, appears to be clustering within a branch containing mosses, gymnosperms, charophytes and pteridophytes. A distinct branch from liverworts till gymnosperms appears to be related to NYE1 but without known function. (b) Maximum likelihood phylogenetic tree of PPH. PPH appears to be derived from a common ancestor of land plants, and the gene was apparently lost in liverworts as well as some genera in pteridophytes and mosses. Only bootstrap support values for key branches are shown beside each corresponding branch.

**Table 1** Detection of homologues of the PAO pathway and phyllobilins across the green clade suggest some possible alternative pathways to chlorophyll degradation

	Model species	No. of orthologues	Phyllobilins detected?
Chlorophytes	<i>Auxenochlorella protothecoides</i>	4/8	RCC <sup>a,b</sup>
Charophytes	<i>Klebsormidium nitens</i>	6/8	–
Liverworts	<i>Marchantia polymorpha</i>	7/8	NCCs <sup>a,c</sup>
Mosses	<i>Physcomitrella patens</i>	8/8	–
Lycophytes	<i>Sellaginella moellendorffii</i>	5/8	NCCs <sup>a,d</sup>
Gymnosperms	<i>Ginkgo biloba</i>	8/8	NCCs <sup>a,d</sup>
Angiosperms	<i>Arabidopsis thaliana</i>	8/8	(D)NCCs <sup>e</sup>

RCC, red chlorophyll catabolite; NCC, colourless chlorophyll catabolite.

<sup>a</sup>This study.

<sup>b</sup>ApRCC1 and ApRCC2 (Figure S3).

<sup>c</sup>Previously undescribed NCC with multiple hydroxylation (Figure 3).

<sup>d</sup>NCC similar to DNCC\_618 (Kuai et al., 2017).

<sup>e</sup>Christ et al. (2016).

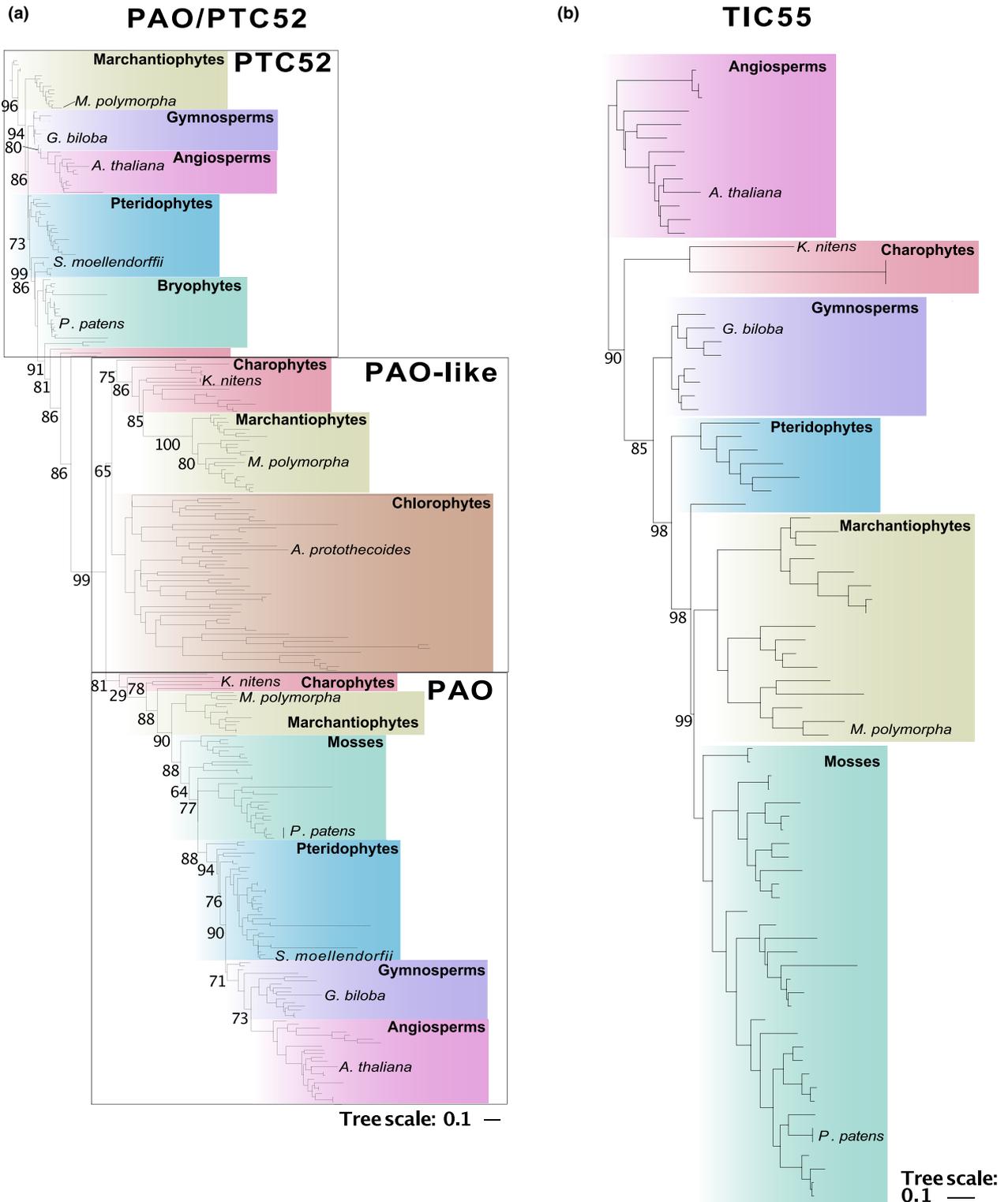
catalysing the conversion of pheophytin *a* into pheophorbide *a*. The hydrolysis of the phytol chain can be catalysed by esterases and lipases, which are relatively large families of enzymes in plant genomes (Schelbert et al., 2009). Forty-two chloroplast-localized  $\alpha/\beta$  hydrolases were screened during PPH identification in *Arabidopsis* (Schelbert et al., 2009). At least two other enzymes had previously been described to dephytylate chlorophyll: chlorophyllase (CLH; Schenk et al., 2007; Hu et al., 2015) and CHLOROPHYLL DEPHYTYLASE 1 (CLD1; Lin et al., 2016). CLH has been first thought to dephytylate chlorophyll, but its subcellular localization is still a matter of debate and may depend on leaf age (Schenk et al., 2007; Tian et al., 2021). CLH has also been shown to be involved in phyllobilin release upon defense activation against chewing herbivores (Hu et al., 2015). In this model, CLH accesses its substrate only after destruction of the chloroplastic membranes. CLH is therefore an unlikely candidate for dephytylation in the PPH-free lineages we identified here. Another dephytylase, CLD1, that is localized to the chloroplast, has been shown to be involved in chlorophyll steady-state turnover and possible recycling of the pigment in *Arabidopsis* (Lin et al., 2016). CLD1 is an interesting candidate for performing dephytylation in lycophytes, pteridophytes, liverworts and hornworts (Figure S6); although, this requires functional confirmation. Having at least three enzymes (CLH, PPH, CLD1) being able to catalyse the chlorophyll dephytylation suggests that there is a wider range of possible hydrolases that may catalyse this reaction outside angiosperms. More work is required to identify the enzyme(s) responsible for dephytylation in these lineages.

### Evolution of the irreversible opening of the porphyrin ring

One of the major steps of the chlorophyll degradation pathway is the opening of the porphyrin ring catalysed by

PAO (Pruzinska et al., 2003). PAO converts pheophorbide *a* into a RCC. PAO is a Rieske-type mononuclear non-heme iron oxygenase that has five close relatives in *Arabidopsis*: PAO, CAO, CHOLINE MONO-OXYGENASE (CMO), PROTOCHLOROPHYLLIDE-DEPENDENT TRANSLOCON COMPONENT 52 (PTC52) and TRANSLOCON AT THE INNER CHLOROPLAST ENVELOPE 55 (TIC55; Gray et al., 2004). Interestingly, PTC52, which has been suggested to have a flavone hydroxylase activity in basil (*Ocimum basilicum*; Berim et al., 2014), belongs to the same orthogroup as PAO but not TIC55 (Figure 3a). PAO orthologues are present across the green lineage (Figure 1b). Potential PAO orthologues in chlorophytes group in a distinct 'PAO-like' clade that also encompasses some charophytes, mosses and liverworts sequences (Figure 3a). Alignment of the Rieske-type and mononuclear iron domain of the *A. protothecoides* PAO show only about 40% identity to other PAO orthologues (Figure S7). Despite some early indications of a monooxygenase involvement in this reaction by application of specific inhibitors (Hortensteiner et al., 2000), the actual ability of this gene to catalyse pheophorbide *a* production remains to be experimentally confirmed.

TIC55 localizes in the chloroplast in *Arabidopsis* and catalyses the C3<sup>2</sup> hydroxylation of pFCC (Hauenstein et al., 2016). Multiple hydroxylated phyllobilins have been detected (Christ and Hörteneiner, 2014; Krätler, 2016), but their physiological relevance during leaf senescence remains unclear. TIC55 is absent from chlorophytes and is only present in some charophytes. It also appears missing in most pteridophytes (Figures 1b and 3b). Surprisingly, TIC55 predicted orthologues from charophytes do not support the 'expected' phylogeny, casting doubt of whether the proteins identified are actually 'real' orthologues of this gene. Alignments of conserved motifs of *K. nitens* TIC55 orthologues do not allow for clear discrimination of its



**Figure 3.** Evolution of PAO and TIC55 and PPH in the Viridiplantae.

(a) Maximum likelihood phylogenetic tree of PAO. Three genes appear to be clustered in one orthogroup that might have been present in the most common ancestor of plants: PTC52, PAO and PAO-like. While a *bona fide* PAO appears to be present in charophytes, the actual enzymatic activities of PAO-like subfamily members remain unknown. Noteworthy, *Auxenochlorella protothecoides* orthologue does belong to the PAO-like subfamily.

(b) Maximum likelihood phylogenetic tree of TIC55. TIC55 does appear as an innovation of the charophytes, and remains monophyletic with few gene duplications (except in mosses). TIC55 being a Rieske-type monooxygenase, it might have derived from a pre-existing enzyme, while its relationship to the PAO or CAO ancestors is unclear. Only bootstrap support values for key branches are shown beside each corresponding branch.

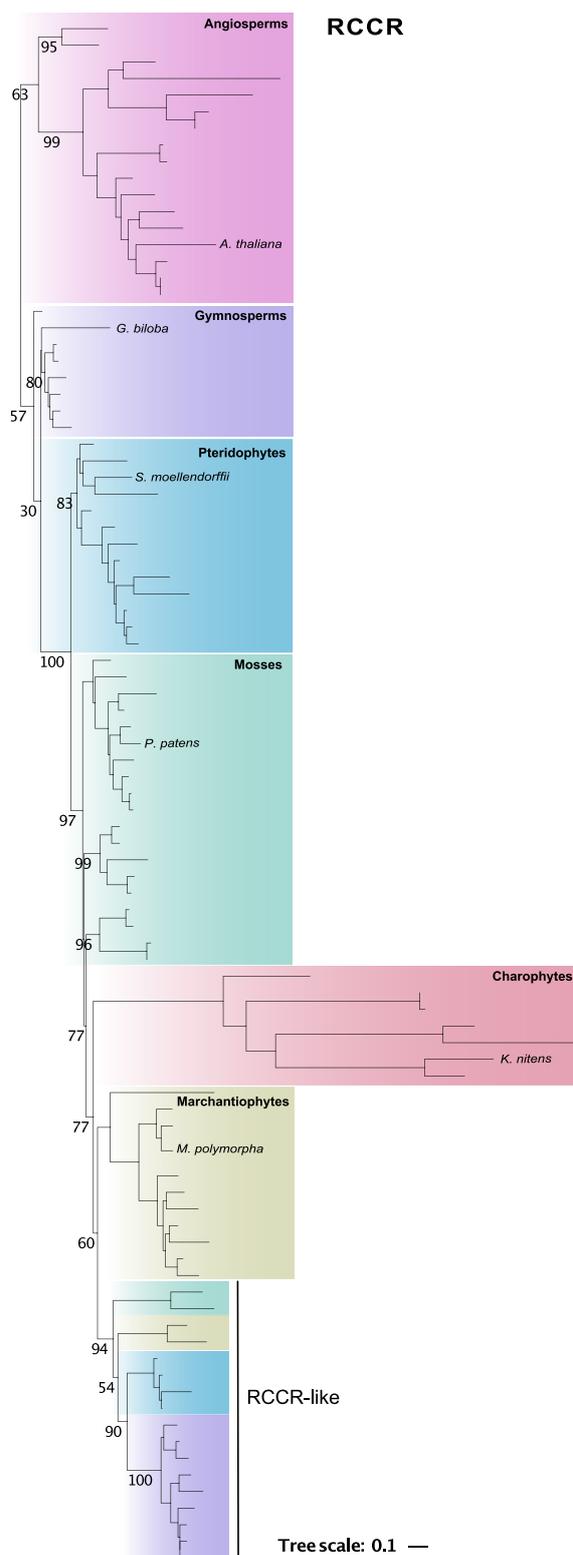
function (Figure S7), and further biochemical characterization would be necessary. While hydroxylation of phyllobilin modifies the polarity of phyllobilins to be exported, the presence of TIC55 is not mandatory for a functional degradation of phyllobilins and therefore not a very significant marker of the pathway's evolution.

#### Evolution of enzymatic steps leading to the phyllobilin's export from the chloroplast

During leaf senescence, it is essential that photoreactive chlorophyll intermediates do not accumulate, in order to avoid toxicity and tissue damages. Therefore, photoreactive RCCs produced by PAO need to be detoxified quickly. RCCR reduces the C15/16 double bond of RCC to produce less photoreactive pFCC (Pruzinska et al., 2005). The conversion from a closed ring into a linear tetrapyrrole is possibly the trigger allowing export of the catabolites from the chloroplast to the cytosol (by an yet unknown transporter; Kuai et al., 2017). RCCR orthologues are conserved in streptophytes (Figures 1b and 4), while they appear absent from chlorophytes. Some inconsistencies in the phylogeny for RCCR orthologues of charophytes and liverworts question the actual validity of orthologue prediction in these clades. Noteworthy, RCCR enzymatic activity was reported in liverworts, mosses and gymnosperms (Pruzinska et al., 2007). In a similar line, the green algae *A. protothecoides* and *Parachlorella kessleri* excrete large amounts of RCC in their growing medium when grown under low nitrogen and heterotrophic conditions (their chemical structure is described in Figure S4; Gossauer, 1994; Hortensteiner et al., 2000). While *A. protothecoides* and *P. kessleri* have evolved a dedicated detoxification of RCC, the extent to which a similar strategy is employed by other chlorophytes remains unclear.

#### Detection of phyllobilins in key species is consistent with orthologue's inference

We then focused on key species for each of the seven lineages for experimentally confirming the presence of the pathway's end-product, and logically infer the conservation of the PAO/phyllobilins pathways across the green lineage (Table 1). The presence/absence of phyllobilin was tested first using dark-induced senescence in detached leaves (or chlorophyll-containing tissues) for the embryophytes *Arabidopsis*, *S. moellendorffii*, *G. biloba*, *Physcomitrella patens* and *M. polymorpha* (see Experimental Procedures). Degradation of chlorophyll was induced in the two algae *A. protothecoides* and *K. nitens* after switching into heterotrophic conditions (low nitrogen and darkness). Phyllobilins could be detected in five of the seven clades investigated (Table 1). Despite several attempts and a diverse range of media and conditions tested [variations in nitrogen content before and after the dark-incubation, change in carbon sources (glucose/sucrose) and duration



**Figure 4.** Phylogenetic tree of RCCR. RCCR appears as an innovation from the charophytes and is present in all other clades. Interestingly, there is a distinct sub-family of RCCR-like genes in gymnosperms, mosses and pteridophytes of unknown function. Only bootstrap support values for key branches are shown beside each corresponding branch.

of the growth and dark-incubation, bubbling versus mixed cultures], no phyllobilins could be identified from either *K. nitens* or *P. patens*. However, shift to heterotrophic growth of *A. protothecoides* induced excretion of two types of RCC in the culture media (Figure S5). *Ap*-RCC1 and *Ap*-RCC2 structures only differ by a single carboxymethyl group bound to the isocyclic ring of chlorophyll in C8<sup>2</sup> position (Figure S5). In *Arabidopsis*, this reaction is catalysed by MES16 (Christ et al., 2012), suggesting that MES16-like activities are present in chlorophytes. Unfortunately, MES16 belongs to an orthogroup gathering many  $\alpha/\beta$ -hydrolases, making it very hard to properly identify orthologues. *Marchantia polymorpha* incubated in the dark accumulates multiple phyllobilins (Figure S5). At least four distinct compounds could be identified, including *Mp*-NCC-4, for which MS–MS fragmentation pattern was shown to be identical to NCC-662 (Kuai et al., 2017; Hörtensteiner et al., 2019). It seems that multiple successive rounds of hydroxylation are added to the core *Mp*-NCC4 in *M. polymorpha*, but their exact position on the porphyrin ring remains to be determined (Figure S5). The presence of NCCs in *M. polymorpha* confirms an active PAO/phyllobilin pathway in liverworts, as predicted by our phylogenomic approach (Figure 1b). But these results also highlight the complexity and versatility of cytosolic side-chain modification.

### Large-scale detection of phyllobilins show the extent of structural diversity

In order to better characterize the diversity of chlorophyll catabolites in phyllobilin-producing species, we took advantage of our laboratory localization, at the heart of the Botanical Garden of the University of Zürich. During three

consecutive autumns, we harvested naturally senescing leaves and extracted phyllobilins out of 203 species and analysed them by LC–MS/MS. Phyllobilins were detected in 183 species (90%; Table S2).

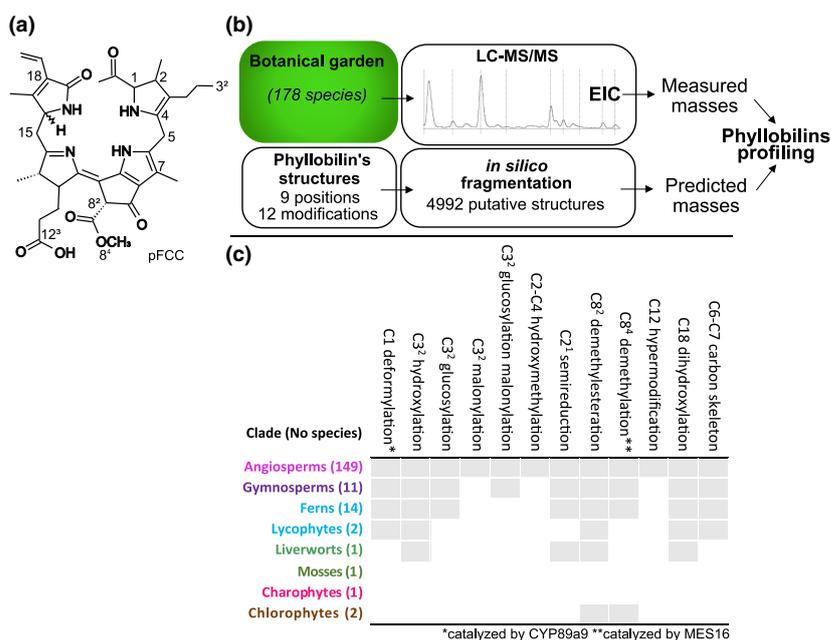
To increase our chance to identify phyllobilins, we screened the measured masses against an *in silico* produced list of ‘diagnostic ions’ with their predicted masses (Figure 5b). We based ourselves on one of the simplest structures of naturally occurring phyllobilin, a C1 deformed pyroRCC (m/z 556.2686; Figure 5). We then predicted all possible combinations of masses that could be originating from the 12 modifications over 8 positions (Figure 5). These modifications are namely C1 deformylation, C3 hydroxylation, malonylation/glucosylation or combination of both, C2–C4 hydroxymethylation, C2 semi-reduction, C8 demethylation, C8 demethylesteration, C12 hypermodification, C18 dihydroxylation, rearrangement of the carbon backbone where the C5/C6 is shifted to a C5/C7 bond (Hörtensteiner et al., 2019). A total of 4992 hypothetical structures could be listed, in which all 26 known phyllobilins could be retrieved from, as well as 15 profiles that are potentially yet undescribed phyllobilins (Figure 5c; Table S2). Interestingly, major structural modifications like the conversion in *iso*-phyllobilanones (m/z 619.2762 [M + H]<sup>+</sup> and m/z 635.2712 [M + H]<sup>+</sup>), recently reported in *P. aquilinum* (Erhart et al., 2018), could also be detected in most ferns and gymnosperms studied here (Table S2). Taken together, the diversity of cytosolic phyllobilins observed here could appear surprising considering the high degree of conservation of the (preceding) chloroplastic steps, and might be indicative of a more general strategy of cells to deal with potentially toxic catabolites.

**Figure 5.** Diversity of cytosolic phyllobilins.

(a) Mapping of the various positions of modifications based on primary FCC. Modifications give rise to fluorescent phyllobilins (DFCCs and FCCS) that are isomerized further in the cytosol.

(b) Strategy used for identification of *de novo* phyllobilins.

(c) Table summarizing all side-chain modifications of phyllobilin described in this study.



## DISCUSSION

Comparative genomics allowed us to identify innovations and losses during the evolution of the chlorophyll degradation in Viridiplantae. This approach was complemented by detection of the pathway's end-products, phyllobilins. Our analysis provides a comprehensive view on possible ways by which plants have dealt with phototoxic chlorophyll catabolites during evolution.

### Chloroplastic degradation of chlorophyll may have facilitated transition to multicellularity

Some unicellular chlorophytes like *A. protothecoides* (Figure S4) and *P. kessleri* (Gossauer, 1994) degrade chlorophyll in their chloroplast, most likely using a PAO-like protein (Figures 1b and 2a), process phyllobilins to a certain extent (presence of modified RCCs in *A. protothecoides*; Figure S4), and eventually excrete phyllobilins into the media (Hortensteiner et al., 2000). It remains unclear whether the ability to produce RCC in *A. protothecoides* can be generalized to other chlorophytes, or is a *de novo* adaptation of this particular species, which allows the switch between autotrophic and heterotrophic growth without accumulation of phototoxic compounds (Gao et al., 2014). Our data suggest that, in chlorophytes, the enzymatic machinery is present but the actual degradation of chlorophyll (and associated gene expression) most likely depends on the lifestyle of each alga.

Interestingly, genes recruited in the PAO/phyllobilin pathway appear to have an ancestral and often distinct function and substrates. For example, an orthologue of *NYE1* in *Chlamydomonas reinhardtii* (belonging to the *NYE1*-like orthogroup; Figure 2a) has been proposed to facilitate photosystem II biosynthesis instead of acting as a magnesium dechelatease in plants (Chen et al., 2019). Homologues of *NYE1* could be identified in some glaucophytes and non-photosynthetic bacteria, but neither in cyanobacteria nor rhodophytes (Obata et al., 2019). This might be indicative of a recruitment of *NYE* in plants via horizontal gene transfer succeeding the cyanobacteria primary endosymbiosis. Meanwhile, the large variation of substrate specificity and dechelating efficiency of the homologues may question their physiological relevance. In other words, the extent to which the *in vitro* ability of a protein to dechelate magnesium is a good proxy for testing its actual orthology to the angiosperm's *NYE/SGR* remains unclear. A similar cyanobacterial origin for *HCAR* has also been suggested: it may have derived from a cyanobacterial ferredoxin-dependent divinyl reductase (F-DVR), responsible for the reduction of an 8-*vinyl* group in the chlorophyll biosynthesis (Ito and Tanaka, 2014). Therefore, the presence of *HCAR*, *NYC1* and *NOL* across all species from the green lineage may not be surprising.

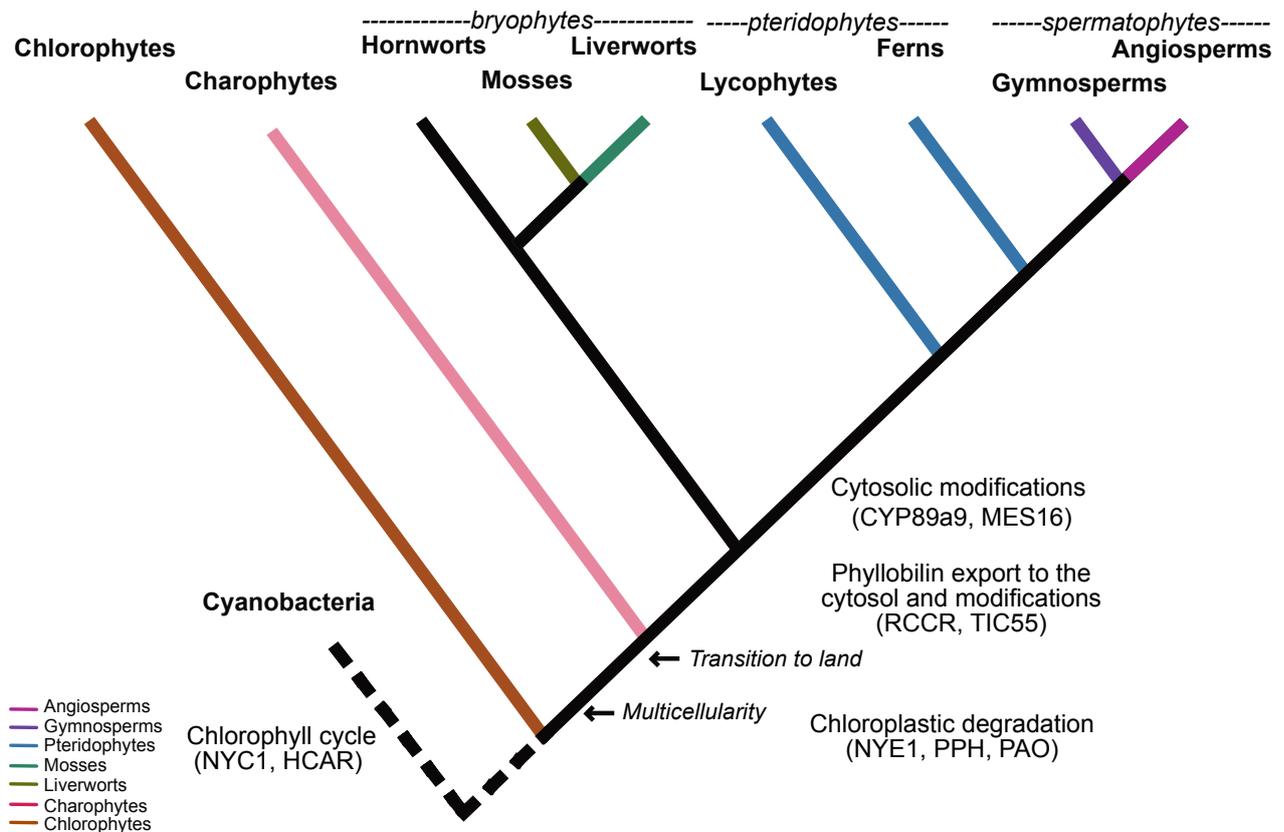
The PAO activity, as a mono-oxygenase, depends obviously on molecular oxygen presence; therefore, production of RCC must have logically succeeded the appearance of oxygenic phototrophs, in a very comparable way heme is degraded in bilins (Zhang et al., 2021). *PAO* and *TIC55* also appear to have homologues in several oxygenic cyanobacteria (Gray et al., 2004). As the ability to convert chlorophyll *a* to *b* by CAO has been suggested to appear several times in the prokaryotic prochlorophytes (Tomitani et al., 1999), it is therefore possible that *PAO* and *TIC55* did originate from duplication events from *CAO*. Chl *b* conversion by CAO is thought to have participated in aerobic photosynthesis evolution (Tomitani et al., 1999). But, again, ancestral monooxygenase homologues are very difficult to distinguish without proof of their biochemical activity.

In charophytes, despite no phyllobilins being detected in *K. nitens* (some pheophorbide *a* trace could be detected in *Chara vulgaris*; Table S2), all genes encoding proteins converting chlorophyll to RCCs are present (Figure 1b). This suggests that charophytes may degrade chlorophyll, although the growth conditions triggering chlorophyll degradation in these species remain elusive. While phototrophic lineages, including charophyte algae, successfully established on land multiple times independently (Lewis and McCourt, 2004), the land plant macroflora originates from a unique streptophyte ancestor (de Vries and Archibald, 2018). Comparative genomics support two genetic innovation bursts during emergence of multicellularity in streptophytes and terrestrialization (embryophytes; Bowles et al., 2020).

Taken together, our results suggest that the chloroplastic portion of the PAO/phyllobilin pathway (i.e. from chlorophyll up to RCC) was already present in the common ancestor of land plants (Figure 6). The recruitment of the PAO-RCCR couple allowed a swift conversion of the highly phototoxic pheophorbide *a* and its subsequent export (as a linear RCC or pFCC) from the chloroplasts (Hirashima et al., 2009; Aubry et al., 2020). This appeared as a key innovation crucial for allowing multicellularity of phototropic organisms by dealing with highly phototoxic chlorophyll catabolites embedded in thylakoid membranes.

### Phyllobilin production evolved coincidentally to water-to-land transition

Given the phototoxicity of chlorophyll catabolites intermediates, especially for multicellular organisms where excretion is not an option, it is interesting to observe that the enzymatic steps that take place within the chloroplasts appear under a much higher selective pressure compared with cytosolic processes. There are in fact only two catabolites (pFCC and hydroxy-pFCC) known to be exported from the chloroplasts, that are subsequently converted in more than 40 different phyllobilins in streptophytes (Table S2).



**Figure 6.** A model for the evolution of the PAO/phyllobilin pathway in plants. Cladogram of the major plant lineages and evolution of key traits that marked the transition from aquatic to a terrestrial life.

This diversity of cytosolic activities is most likely due to the high redundancy of detoxifying processes that appear to ensure an extensive range of enzymatic activities (reductases, hydrolases, etc.) and of relatively high promiscuity. Low substrate specificity allows the enzymes to react with virtually any toxic/xenobiotic compound that may be detrimental to the cellular activity. Interestingly, hydroxylations of specialized metabolites (diterpenoids) that also involve cytochrome P450s were shown to be involved in the defense of plant herbivores, while preserving plants from their own toxicity (Li et al., 2021). A similar function for the accumulation of phyllobilin remains to be investigated (Kräutler, 2016). It is therefore possible that these cytosolic steps mitigate the impact of light stress associated to terrestrial lifestyle, and were central to the transition to land.

#### Lower genomic resolution for the cytosolic phyllobilin-modifying steps

Despite at least nine different phyllobilins detected in senescing *Arabidopsis* leaves (Hörtensteiner et al., 2019), only three enzymes have been associated with phyllobilin side-chain modifications, namely TIC55, CYP89a9 and MES16 (Christ

et al., 2012, 2013; Hauenstein et al., 2016). CYP89a9 belongs to the large and highly interrelated multi-gene family of cytochrome P450 oxidase and MES16 to the  $\alpha/\beta$ -hydrolase family (Christ et al., 2012, 2013). It is therefore not surprising that they both clustered in two large orthogroups, which makes the identification of proper candidate genes difficult. For example, the cytochrome P450 family comprises 245 members in *Arabidopsis*, and has undergone rapid expansion and neofunctionalization (Nelson and Werck-Reichhart, 2011; Du et al., 2016). Therefore, screening for intra-specific (natural) variations between phyllobilins (e.g. deformed, demethylated or conjugated phyllobilins) that could be related to some specific loci by genome-wide association study might be a complementary strategy to identify more cytosolic enzymes involved in chlorophyll degradation.

#### Diversity of the cytosolic modifications across phyllobilin-producing species

Monitoring phyllobilins in 183 species via a 'Botanical Garden metabolomics' approach provided a unique view on the diversity of cytosolic modifications originating from one single pigment: chlorophyll. While the sampling is clearly biased towards angiosperms, it provides a unique perspective on

the structural diversity of phyllobilins. Surprisingly, about 10% of the samples did not contain any detectable phyllobilins. This could be due to their respective leaf phenologies: despite using an extensive harvesting period, not all species or organs do induce senescence and degrade chlorophyll at the same time. Particularly mosses and charophytes were hardly degrading chlorophyll despite numerous attempts to induce senescence artificially in these species. Moreover, some species like tobacco do not accumulate phyllobilins to detectable amounts, while degrading chlorophyll. This might be due to a particularly quick transfer to acidic vacuoles or possibly to large amounts of secondary metabolites that might limit the extraction of phyllobilin. While our innovative strategy to predict *in silico* and screen any possible sorts of phyllobilin improved the pool of molecules to be screened for, it is still possible that some catabolites were simply missed during the analysis.

Since their first chemical characterization in 1996 in barley, a constantly increasing number of phyllobilins have been described (Christ et al., 2016; Kräutler, 2016). Two major families of phyllobilins can be distinguished: type I that are non-fluorescent (NCCs), 1-formyl, 19-oxobilins; and type II, dioxobilin-type non-fluorescent (DNCCs; Kräutler, 2014, 2016). Within these two categories, a wide range of side-chain modifications have been reported. Some modifications, like the C<sub>3</sub><sup>2</sup> hydroxylation (Figure 5c) are largely shared, while many other phyllobilins are restricted to only a few species. For example, hypermodified FCCs with functionalization in C<sup>12</sup> position have been detected in banana (*Musa acuminata*) and *Spathiphyllum wallisii* (Moser et al., 2012; Vergeiner et al., 2015). In *Ulmus glabra*, a bicycloglysidic NCC connecting C<sub>12</sub><sup>3</sup> and C<sub>3</sub><sup>2</sup> was reported (Scherl et al., 2016). Some other modifications appear restricted to some clades, like the major rearrangement of the carbon backbone that has been initially observed in *P. aquilinum* but found here to be happening in ferns and gymnosperms (Erhart et al., 2018; Table S2).

Finally, using this large dataset, no particular phylogenetic pattern could be observed that would indicate a common ancestor to any of the various modifications downstream of TIC55 (Table S2). Phyllobilins accumulating in senescent leaves, while having sometimes similar structures, may in many instances result from distinct (promiscuous) enzymatic activities. In terms of physiological relevance, this indicates that a relatively diverse and promiscuous set of cytosolic detoxication reactions have been recruited to change the polarity of phyllobilin towards vacuolar export. But, in contrast to the highly conserved chloroplastic portion of the pathway, these reactions are likely shared with other detoxication processes and do not necessarily evolve in a coordinated manner. More extensive analyses in other clades will be indicative of the actual extent of complexity of the cytosolic detoxifying processes during land plant diversification.

## CONCLUSIONS

Combining homology inference and biochemical characterization over a wide range of species covering the green lineage has proven to be a promising approach to help understanding evolution of biological pathways. Completion of more and better-quality genomes (particularly for large and repetitive genomes like in gymnosperms) will surely improve the quality of identification of orthologues at the species level (Cannell et al., 2020). Noteworthy, most gymnosperms that do not degrade chlorophyll in a coordinated way (with the notable exception of *Larix decidua*) still have the entire set of PAO/chlorophyll genes in their genome. This emphasizes the crucial role of transcriptional regulation in the onset of chlorophyll degradation. More generally, it also raises interesting questions about the possible reasons and dynamics for such genes to be kept or lost during evolution.

We could observe here a high degree of conservation of the genes involved in the early chloroplastic step of the PAO/phyllobilin pathway. Taking into consideration the few indications of potential prokaryotic homologues (at least for *HCAR*, *NYE1*, *PAO*; Gray et al., 2004; Ito and Tanaka, 2014; Obata et al., 2019), this raises interesting questions on their actual function in prochlorophytes, and the actual process and timing by which a tightly coordinated pathway has evolved in plants. When considering chlorophyll degradation, it is also important to consider the various cellular and tissue-specific gene expression patterns that relate to changing contexts in which chlorophyll is degraded in the green lineages. Indeed, the extent of co-evolution of senescence and the hormonal cues that regulate the PAO/phyllobilin pathway remain to be determined (Thomas et al., 2009; Bowles et al., 2020; Lai et al., 2020). In *Arabidopsis*, all eight genes involved in the core PAO/phyllobilins pathway are under a complex transcriptional control that involves multiple families of transcription factors that are in turn under hormonal control (Kuai et al., 2017; Aubry et al., 2020). Building on this work that focuses on key species, and helped by previous attempts in *Arabidopsis* senescent leaves (Hickman et al., 2017), it would be important to identify the gene regulatory networks that control the PAO/phyllobilin pathway in all these different clades. Again here, focusing on model species with high-quality genomes will be necessary to identify potential *cis* elements and transcription factors involved. It remains unclear how this pathway could be maintained under similar regulation (or at least in response to similar environmental cues, in a similar way) during the complex evolution/expansion/alteration of various transcription factor families in the green clade (Lai et al., 2020). More work is needed to identify the extent of conservation of the gene regulatory networks and their connection to various environmental/hormonal cues (Fürst-Jansen et al., 2020) that regulate the PAO/phyllobilin

pathway, and that degrade chlorophyll in a safe and efficient manner. Chlorophyll degradation is an overlooked yet apparently crucial aspect of the transition from water to land, and therefore is of great importance for our understanding of plant diversification.

## EXPERIMENTAL PROCEDURES

### Evolutionary bioinformatics and phylogenetic analysis

This study used genomes and transcriptomes sequences from 72 species, all sources and accession numbers of which are summarized in Table S1. Most angiosperm genome sequences were downloaded from Phytozome (Goodstein et al., 2012), the 1KP Project (Matasci et al., 2014), Fernbase (Li et al., 2018), ConGenie (Sundell et al., 2015) and UniProt (Bateman et al., 2021), or from individual publications. Genes for the 72 selected species (Data S1) were sorted into orthogroups using OrthoFinder v2.5.2 software (Emms and Kelly, 2015, 2020). Orthogroups are: “collections of orthologous and paralogous genes descending from one gene in the common ancestor of all species included in the analysis” (Emms and Kelly, 2015). Among the 24 479 orthogroups created, we focused our analysis on the orthogroups that contain known *Arabidopsis* genes of interests.

After manual curation, gene trees were inferred for each gene following published methods (Harris et al., 2020). Briefly, homologous amino acid sequences were aligned using MAFFT v7.475 (Kato and Standley, 2013), and the resulting multiple alignment was trimmed using BGME v4.0 with a BLOSUM30 matrix (Crisuolo and Gribaldo, 2010). The clean alignment was then processed by IQ-Tree v1.6.12 using bootstrapped maximum likelihood phylogenies with an empirical profile mixture models (C10–C60; Nguyen et al., 2015). The trees were visualized and edited in iTOL v5 (Letunic and Bork, 2019).

### Plant material growth and chlorophyll degradation induction

*Arabidopsis* and *S. moellendorffii* were grown on soil in 16 h day / 8 h night conditions, 23°C, 100  $\mu\text{mol m}^{-2} \text{sec}^{-1}$ , 60% humidity. *Ginkgo biloba* leaves were harvested mid-August in the Botanical Garden of the University of Zürich. Detached leaves from these three species were dark-incubated for 7 days to induce phyllobilin production. *Physcomitrella patens* was grown on solid Knopp medium (Frank et al., 2005) under long-day conditions and later transferred in nitrogen-free medium for dark-incubation. *Marchantia polymorpha* was grown in the same conditions as *Arabidopsis* (see above) on ½ Gamborg medium plates supplemented with 0.5% sucrose. Detached thalli were dark-incubated for 10 days on wet filter paper before phyllobilins were extracted. *Klebsormidium nitens* was grown following a published protocol (Hori et al., 2014), and nitrogen and light depletion experiments were carried out (data not shown). *Auxenochlorella protothecoides* (previously called *Chlorella protothecoides*) was first grown autotrophically for 10 days, and then transferred to nitrogen-depleted medium for 5 days following a published protocol (Hortensteiner et al., 2000).

### Phyllobilins extraction and LC-MS/MS analysis

After dark-incubation and the first sign of senescence, the plant materials were immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until further use. Prior to extraction and following a published protocol (Christ et al., 2016), the frozen samples were ground

in liquid nitrogen using mortar and pestle and resuspended in five volumes (w/v) of pre-cooled extraction buffer (80% MeOH, 20% water and 0.1% formic acid; Christ et al., 2016). Samples were centrifuged and supernatants were separated twice, and transferred on the LC-MS system. The samples were then analysed using a LC-MS/MS Thermo Scientific Dionex Ultimate 3000 Rapid Separation LC system (Thermo Fisher Scientific, Reinach, Switzerland) equipped with a photodiode array detector coupled to a Bruker Compact ESI-QTOF (Bruker Daltonics). Phyllobilins were separated on a C18 column (ACQUITY UPLC BEH, 1.7  $\mu\text{m}$ , 2.1  $\times$  9  $\times$  150 mm; Waters, Milford, MA USA), using a similar protocol as published previously (Christ et al., 2016). Briefly, a gradient of acetonitrile with 0.1% (v/v) formic acid was performed with water with 0.1% (v/v) formic acid as follows (all v/v): 30% acetonitrile for 0.5 min, 30–70% in 7.5 min, 70–100% in 0.1 min and 100% for 4 min. The instrument was set to acquire over the m/z range 50–1300, with an acquisition rate of four spectra per sec. All data were recalibrated internally using pre-run injection of sodium formate [10 mM sodium hydroxide in 0.2% formic acid, 49.8% water, 50% isopropanol (v/v)]. Each sample was run in data-dependent MS/MS mode.

### Mass- and fragmentation prediction and detection of phyllobilin

Most of the phyllobilins known so far were identified by LC-MS/MS and based on a set of diagnostic ions that are most likely fragmenting from a porphyrin ring. In order to mine for new variations in phyllobilins, LC-MS/MS data were compared with a list of *in silico* generated masses that gather all possible combinations of 12 reported modifications on 8 positions over the porphyrin's ring (Figure 5). A total of 4992 hypothetical structures and their corresponding predicted masses were produced and compared with the measured masses from the MS/MS data for each species using a home-made R script (available at [https://github.com/phyllomenous/phyllobilin\\_permutations](https://github.com/phyllomenous/phyllobilin_permutations)).

### ACKNOWLEDGEMENTS

The authors would like to dedicate this article to their colleague, mentor and friend, the late Stefan Hörtensteiner, whose seminal work on chlorophyll degradation inspired the early steps of this work. The authors would like to thank Bastien Christ, Bernhard Kräutler and Philipp Schlueter for helpful comments on the manuscript. This work was supported by the Swiss National Science Foundation (#31003A\_172977) and the University of Zurich. Open Access Funding provided by Universität Zurich.

### AUTHOR CONTRIBUTIONS

IS, SH and SA designed the research; IS, MH, NF, SO, DM helped collecting data and analysing; all authors helped with the interpretation; IS, DM and SA wrote the paper.

### CONFLICT OF INTEREST

The authors declare no competing interests.

### SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Table S1.** List of species used in this study and the data sources

**Table S2.** List of phyllobilins from 178 species sorted by modifications observed and sorted by position on the porphyrin ring and

their corresponding phyllobilins. The phyllobilin nomenclature used here was initially described in Hörtensteiner et al. (2019).

**Figure S1.** Phylogenetic tree showing evolutionary relationship between all 72 species involved in the orthogroups analysis. The same species tree was used to draw the presence/absence map in Figure 1(b). Model species used for phyllobilin detection are shown in bold.

**Figure S2.** Presence of gene homologues encoding chlorophyll synthesis across 72 plant and algal species. Each concentric ring indicates an enzyme, and the presence (coloured) or absence (empty) of a homologue for that particular gene. Identity of species contributing to the phylogenetic tree with full species names given in Figure S1. The chlorophyll synthesis pathway appears mostly conserved throughout the entire Viridiplantae.

**Figure S3.** Maximum likelihood phylogenetic tree of (a) NYC1/NOL and (b) HCAR orthologues. Only bootstrap support values for key branches are shown beside each corresponding branch.

**Figure S4.** *Auxenochlorella protothecoides* excretes two types of RCC in the growing media under heterotrophic conditions. (a) Base peak chromatogram (BPC) of the extract and extracted ion chromatogram for ApRCC1 (light blue) and ApRCC2 (dark blue) (b) MS (top) and MS/MS (bottom) spectra and predicted structure models of ApRCC1, (c) MS (top) and MS/MS (bottom) spectra and predicted chemical structure of ApRCC2. RCC, red chlorophyll catabolites.

**Figure S5.** Phyllobilins detected in *Marchantia polymorpha*. *M. polymorpha* detached thalli were dark-incubated for 10 days in the dark to induce chlorophyll degradation. (a) Base peak chromatogram (in black) and extracted ion chromatogram of the four new phyllobilins detected by UPLC-MS/MS. (b) Putative structure models of all four phyllobilins detected and corresponding MS/MS fragmentation pattern. The red star in MpNCC1 indicates a hypothetical position for the hydroxyl group that cannot be resolved based on the data available.

**Figure S6.** Maximum likelihood phylogenetic tree of CLD1 orthologues. Only bootstrap support values for key branches are shown beside each corresponding branch.

**Figure S7.** Alignment of Rieske and mononuclear iron domain of proteins from the PAO family. Alignment was performed using MAFFT and motif highlighted in colour.

## OPEN RESEARCH BADGES



This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at [https://figshare.com/articles/dataset/Evolution\\_of\\_chlorophyll\\_degradation\\_is\\_associated\\_with\\_plant\\_transition\\_to\\_land/17128793](https://figshare.com/articles/dataset/Evolution_of_chlorophyll_degradation_is_associated_with_plant_transition_to_land/17128793) and [https://github.com/phyllobilins/phyllobilin\\_permutations](https://github.com/phyllobilins/phyllobilin_permutations).

## DATA AVAILABILITY STATEMENT

All relevant data can be found within the manuscript and its supporting materials.

## REFERENCES

- Armstead, I., Donnison, I., Aubry, S., Harper, J., Hörtensteiner, S., James, C. et al. (2006) From crop to model to crop: identifying the genetic basis of the staygreen mutation in the *Lolium/Festuca* forage and amenity grasses. *New Phytologist*, **172**, 592–597.
- Armstead, I., Donnison, I., Aubry, S., Harper, J., Hörtensteiner, S., James, C. et al. (2007) Cross-species identification of Mendel's I Locus. *Science*, **315**, 73–73.
- Aubry, S., Fankhauser, N., Ovinnikov, S., Pruzinská, A., Stirnemann, M., Zienkiewicz, K. et al. (2020) Pheophorbide a may regulate jasmonate signaling during dark-induced senescence. *Plant Physiology*, **182**, 776–791.
- Aubry, S., Mani, J. & Hörtensteiner, S. (2008) Stay-green protein, defective in Mendel's green cotyledon mutant, acts independent and upstream of pheophorbide a oxygenase in the chlorophyll catabolic pathway. *Plant Molecular Biology*, **67**, 243–256.
- Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S. et al. (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, **49**, D480–D489.
- Berim, A., Park, J.-J. & Gang, D.R. (2014) Unexpected roles for ancient proteins: flavone 8-hydroxylase in sweet basil trichomes is a Rieske-type, PAO-family oxygenase. *The Plant Journal*, **80**, 385–395.
- Bowles, A.M.C., Bechtold, U. & Paps, J. (2020) The origin of land plants is rooted in two bursts of genomic novelty. *Current Biology*, **30**, 530–536.e2.
- Cannell, N., Emms, D.M., Hetherington, A.J., MacKay, J., Kelly, S., Dolan, L. et al. (2020) Multiple metabolic innovations and losses are associated with major transitions in land plant evolution. *Current Biology*, **30**, 1783–1800.e11.
- Chen, Y., Shimoda, Y., Yokono, M., Ito, H. & Tanaka, A. (2019) Mg-dechelataase is involved in the formation of photosystem II but not in chlorophyll degradation in *Chlamydomonas reinhardtii*. *The Plant Journal*, **97**, 1022–1031.
- Cheng, Y. & Guan, J. (2014) Involvement of pheophytinase in ethylene-mediated chlorophyll degradation in the peel of harvested 'Yali' pear. *Journal of Plant Growth Regulation*, **33**, 364–372.
- Chew, A.G.M. & Bryant, D.A. (2007) Chlorophyll biosynthesis in bacteria: the origins of structural and functional diversity. *Annual Review of Microbiology*, **61**, 113–129.
- Christ, B., Hauenstein, M. & Hörtensteiner, S. (2016) A liquid chromatography-mass spectrometry platform for the analysis of phyllobilins, the major degradation products of chlorophyll in *Arabidopsis thaliana*. *The Plant Journal*, **88**, 505–518.
- Christ, B. & Hörtensteiner, S. (2014) Mechanism and significance of chlorophyll breakdown. *Journal of Plant Growth Regulation*, **33**, 4–20.
- Christ, B., Schelbert, S., Aubry, S., Süssenbacher, I., Müller, T., Kräutler, B. et al. (2012) MES16, a member of the methyltransferase protein family, specifically demethylates fluorescent chlorophyll catabolites during chlorophyll breakdown in *Arabidopsis*. *Plant Physiology*, **158**, 628–641.
- Christ, B., Süssenbacher, I., Moser, S., Bichsel, N., Egert, A., Müller, T. et al. (2013) Cytochrome P450 CYP89A9 is involved in the formation of major chlorophyll catabolites during leaf senescence in *Arabidopsis*. *The Plant Cell*, **25**, 1868–1880.
- Crisuolo, A. & Grimaldo, S. (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology*, **10**, 210.
- Du, H., Ran, F., Dong, H.-L., Wen, J., Li, J.-N. & Liang, Z. (2016) Genome-wide analysis, classification, evolution, and expression analysis of the cytochrome P450 93 family in land plants. *PLoS One*, **11**, e0165020.
- Emms, D.M. & Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, **16**, 157.
- Emms, D.M. & Kelly, S. (2020) Benchmarking orthogroup inference accuracy: revisiting orthobench. *Genome Biology and Evolution*, **12**, 2258–2266.
- Erhart, T., Vergeiner, S., Kreutz, C., Kräutler, B. & Müller, T. (2018) Chlorophyll breakdown in a fern-discovery of phyllobilin isomers with a rearranged carbon skeleton. *Angewandte Chemie International Edition*, **57**, 14937–14941.
- Fitch, W.M. (1970) Distinguishing homologous from analogous proteins. *Systematic Zoology*, **19**, 99.
- Frank, W., Ratnadewi, D. & Reski, R. (2005) *Physcomitrella patens* is highly tolerant against drought, salt and osmotic stress. *Planta*, **220**, 384–394.
- Fürst-Jansen, J.M.R., de Vries, S. & de Vries, J. (2020) Evo-physio: on stress responses and the earliest land plants. *Journal of Experimental Botany*, **71**, 3254–3269.
- Gao, C., Wang, Y., Shen, Y., Yan, D., He, X., Dai, J. et al. (2014) Oil accumulation mechanisms of the oleaginous microalga *Chlorella protothecoides* revealed through its genome, transcriptomes, and proteomes. *BMC Genomics*, **15**, 582.

- Gómez, M.D., Vera-Sirera, F. & Pérez-Amador, M.A. (2013) Molecular programme of senescence in dry and fleshy fruits. *Journal of Experimental Botany*, **65**, 4515–4526.
- Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J. et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, **40**, D1178–D1186.
- Gossauer, A. (1994) Catabolism of tetrapyrroles. *Chimia*, **48**, 352–361.
- Gray, J., Wardzala, E., Yang, M., Reinbothe, S., Haller, S. & Pauli, F. (2004) A small family of LLS1-related non-heme oxygenases in plants with an origin amongst oxygenic photosynthesizers. *Plant Molecular Biology*, **54**, 39–54.
- Harris, B.J., Harrison, C.J., Hetherington, A.M. & Williams, T.A. (2020) Phylogenomic evidence for the monophyly of bryophytes and the reductive evolution of stomata. *Current Biology*, **30**, 2001–2012.e2.
- Hauenstein, M., Christ, B., Das, A., Aubry, S. & Hörtensteiner, S. (2016) A role for TIC55 as a hydroxylase of phyllobilins, the products of chlorophyll breakdown during plant senescence. *The Plant Cell*, **28**, 2510–2527.
- Hendry, G.A.F., Houghton, J.D. & Brown, S.B. (1987) The degradation of chlorophyll - A biological enigma. *New Phytologist*, **107**, 255–302.
- Hickman, R., Van Verk, M.C., Van Dijken, A.J.H., Mendes, M.P., Vroegop-Vos, I.A., Caarls, L. et al. (2017) Architecture and dynamics of the jasmonic acid gene regulatory network. *The Plant Cell*, **29**, 2086–2105.
- Hirashima, M., Tanaka, R. & Tanaka, A. (2009) Light-independent cell death induced by accumulation of pheophorbide a in arabidopsis thaliana. *Plant and Cell Physiology*, **50**, 719–729.
- Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M. et al. (2014) *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nature Communications*, **5**, 3978.
- Horie, Y., Ito, H., Kusaba, M., Tanaka, R. & Tanaka, A. (2009) Participation of chlorophyll b reductase in the initial step of the degradation of light-harvesting chlorophyll a/b-protein complexes in Arabidopsis. *Journal of Biological Chemistry*, **284**, 17449–17456.
- Hortensteiner, S., Chinner, J., Matile, P., Thomas, H. & Donnison, I.S. (2000) Chlorophyll breakdown in *Chlorella protothecoides*: characterization of degreening and cloning of degreening-related genes. *Plant Molecular Biology*, **42**, 439–450.
- Hörtensteiner, S. & Feller, U. (2002) Nitrogen metabolism and remobilization during senescence. *Journal of Experimental Botany*, **53**, 927–937.
- Hörtensteiner, S., Hauenstein, M. & Kräutler, B. (2019) Chlorophyll breakdown—Regulation, biochemistry and phyllobilins as its products. In *Advances in Botanical Research*, **90**, 213–271.
- Hörtensteiner, S., Rodoni, S., Schellenberg, M., Vicentini, F., Nandi, O.I., Qui, Y.-L. et al. (2000) Evolution of Chlorophyll degradation: the significance of RCC reductase. *Plant Biology*, **2**, 63–67.
- Hu, X., Makita, S., Schelbert, S., Sano, S., Ochiai, M., Tsuchiya, T. et al. (2015) Reexamination of chlorophyllase function implies its involvement in defense against chewing herbivores. *Plant Physiology*, **167**, 660–670.
- Ito, H. & Tanaka, A. (2014) Evolution of a new chlorophyll metabolic pathway driven by the dynamic changes in enzyme promiscuous activity. *Plant and Cell Physiology*, **55**, 593–603.
- Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.
- Kräutler, B. (2014) Phyllobilins – the abundant bilin-type tetrapyrrolic catabolites of the green plant pigment chlorophyll. *Chemical Society Reviews*, **43**, 6227–6238.
- Kräutler, B. (2016) Breakdown of chlorophyll in higher plants—phyllobilins as abundant, yet hardly visible signs of ripening, senescence, and cell death. *Angewandte Chemie International Edition*, **55**, 4882–4907.
- Kräutler, B. & Hörtensteiner, S. (2014) Chlorophyll breakdown: chemistry, biochemistry and biology. In: Ferreira, G.C., Kadish, K.M., Smith, K.M. & Guillard, R. (Eds.) *Handbook of Porphyrin Science - Chlorophyll, Photosynthesis and Bio-inspired Energy*. Singapore: World Scientific Publishing, pp. 117–185.
- Kuai, B., Chen, J. & Hörtensteiner, S. (2017) The biochemistry and molecular biology of chlorophyll breakdown. *Journal of Experimental Botany*, **69**, 751–767.
- Kusaba, M., Ito, H., Morita, R., Iida, S., Sato, Y., Fujimoto, M. et al. (2007) Rice NON-YELLOW coloring1 is involved in light-harvesting complex II and Grana degradation during leaf senescence. *The Plant Cell*, **19**, 1362–1375.
- Lai, X., Chahtane, H., Martin-Arevalillo, R., Zubieta, C. & Parcy, F. (2020) Contrasted evolutionary trajectories of plant transcription factors. *Current Opinion in Plant Biology*, **54**, 101–107.
- Letunic, I. & Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*, **47**, W256–W259.
- Lewis, L.A. & McCourt, R.M. (2004) Green algae and the origin of land plants. *American Journal of Botany*, **91**, 1535–1556.
- Li, F.-W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P.-M. et al. (2018) Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants*, **4**, 460–472.
- Li, J., Halitschke, R., Li, D., Paetz, C., Su, H., Heiling, S. et al. (2021) Controlled hydroxylations of diterpenoids allow for plant chemical defense without autotoxicity. *Science*, **371**, 255–260.
- Lin, Y.-P., Wu, M.-C. & Charrg, Y. (2016) Identification of a chlorophyll dephytylase involved in chlorophyll turnover in Arabidopsis. *The Plant Cell*, **28**, 2974–2990.
- Matasci, N., Hung, L.-H., Yan, Z., Carpenter, E.J., Wickett, N.J., Mirarab, S. et al. (2014) Data access for the 1000 Plants (1KP) project. *GigaScience*, **3**, 17.
- Meguro, M., Ito, H., Takabayashi, A., Tanaka, R. & Tanaka, A. (2011) Identification of the 7-hydroxymethyl chlorophyll a reductase of the chlorophyll cycle in Arabidopsis. *The Plant Cell*, **23**, 3442–3453.
- Moser, S., Müller, T., Holzinger, A., Lütz, C. & Kräutler, B. (2012) Structures of chlorophyll catabolites in bananas (*Musa acuminata*) reveal a split path of chlorophyll breakdown in a ripening fruit. *Chemistry - A European Journal*, **18**, 10873–10885.
- Mühlecker, W., Ongania, K.-H., Kräutler, B., Matile, P. & Hörtensteiner, S. (1997) Tracking down chlorophyll breakdown in plants: elucidation of the constitution of a “Fluorescent” chlorophyll catabolite. *Angewandte Chemie International Edition in English*, **36**, 401–404.
- Nelson, D. & Werck-Reichhart, D. (2011) A P450-centric view of plant evolution. *The Plant Journal*, **66**, 194–211.
- Nguyen, L.-T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. (2015) iq-tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**, 268–274.
- Obata, D., Takabayashi, A., Tanaka, R., Tanaka, A. & Ito, H. (2019) Horizontal transfer of promiscuous activity from nonphotosynthetic bacteria contributed to evolution of chlorophyll degradation pathway. *Molecular Biology and Evolution*, **36**, 2830–2841.
- Pruzinska, A., Anders, I., Aubry, S., Schenk, N., Tapernoux-Luthi, E., Muller, T. et al. (2007) In vivo participation of red chlorophyll catabolite reductase in chlorophyll breakdown. *Plant Cell*, **19**, 369–387.
- Pruzinska, A., Tanner, G., Anders, I., Roca, M. & Hörtensteiner, S. (2003) Chlorophyll breakdown: pheophorbide a oxygenase is a Rieske-type iron-sulfur protein, encoded by the accelerated cell death 1 gene. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 15259–15264.
- Pruzinska, A., Tanner, G., Aubry, S., Anders, I., Moser, S., Muller, T. et al. (2005) Chlorophyll breakdown in senescent Arabidopsis leaves. Characterization of chlorophyll catabolites and of chlorophyll catabolic enzymes involved in the degreening reaction. *Plant Physiology*, **139**, 52–63.
- Rodoni, S., Mühlecker, W., Anderl, M., Kräutler, B., Moser, D., Thomas, H. et al. (1997) Chlorophyll breakdown in senescent chloroplasts (Cleavage of pheophorbide a in two enzymic steps). *Plant Physiology*, **115**, 669–676.
- Sakuraba, Y., Park, S.-Y., Kim, Y.-S., Wang, S.-H., Yoo, S.-C., Hörtensteiner, S. et al. (2014) Arabidopsis STAY-GREEN2 is a negative regulator of chlorophyll degradation during leaf senescence. *Molecular Plant*, **7**, 1288–1302.
- Sato, Y., Morita, R., Katsuma, S., Nishimura, M., Tanaka, A. & Kusaba, M. (2009) Two short-chain dehydrogenase/reductases, NON-YELLOW COLORING 1 and NYC1-LIKE, are required for chlorophyll b and light-harvesting complex II degradation during senescence in rice. *The Plant Journal*, **57**, 120–131.
- Satoh, S., Ikeuchi, M., Mimuro, M. & Tanaka, A. (2001) Chlorophyll b expressed in cyanobacteria functions as a light-harvesting antenna in photosystem I through flexibility of the proteins. *Journal of Biological Chemistry*, **276**, 4293–4297.
- Schelbert, S., Aubry, S., Burla, B., Agne, B., Kessler, F., Krupinska, K. et al. (2009) Pheophytin pheophorbide hydrolase (pheophytinase) is involved in chlorophyll breakdown during leaf senescence in Arabidopsis. *The Plant Cell*, **21**, 767–785.

- Schenk, N., Schelbert, S., Kanwischer, M., Goldschmidt, E.E., Dörmann, P. & Hörtensteiner, S. (2007) The chlorophyllases AtCLH1 and AtCLH2 are not essential for senescence-related chlorophyll breakdown in *Arabidopsis thaliana*. *FEBS Letters*, **581**, 5517–5525.
- Scherl, M., Müller, T., Kreutz, C.R., Huber, R.G., Zass, E., Liedl, K.R. et al. (2016) Chlorophyll catabolites in fall leaves of the wych elm tree present a novel glycosylation motif. *Chemistry - A European Journal*, **22**, 9498–9503.
- Shimoda, Y., Ito, H. & Tanaka, A. (2012) Conversion of chlorophyll b to chlorophyll a precedes magnesium dechelation for protection against necrosis in *Arabidopsis*. *The Plant Journal*, **72**, 501–511.
- Shimoda, Y., Ito, H. & Tanaka, A. (2016) *Arabidopsis* STAY-GREEN, Mendel's green cotyledon gene, encodes magnesium-dechelataase. *The Plant Cell*, **28**, 2147–2160.
- Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y., Sjödin, A. et al. (2015) The plant genome integrative explorer resource: PlantGen. *New Phytologist*, **208**, 1149–1156.
- Tanaka, A. & Tanaka, R. (2006) Chlorophyll metabolism. *Current Opinion in Plant Biology*, **9**, 248–255.
- Tanaka, A. & Tanaka, R. (2019) Chapter Six - The biochemistry, physiology, and evolution of the chlorophyll cycle. *Advances in Botanical Research*, **90**, 183–212. <https://doi.org/10.1016/bs.abr.2019.03.005>
- Thomas, H., Huang, L., Young, M. & Ougham, H. (2009) Evolution of plant senescence. *BMC Evolutionary Biology*, **9**, 163.
- Tian, Y., Zhong, R., Wei, J., Luo, H., Eyal, Y., Jin, H. et al. (2021) *Arabidopsis* CHLOROPHYLLASE 1 protects young leaves from long-term photodamage by facilitating FtsH-mediated D1 degradation in photosystem II repair. *Molecular Plant*, **14**, 1149–1167.
- Tomitani, A., Okada, K., Miyashita, H., Matthijs, H.C.P., Ohno, T. & Tanaka, A. (1999) Chlorophyll b and phycobilins in the common ancestor of cyanobacteria and chloroplasts. *Nature*, **400**, 159–162.
- Vergeiner, C., Ulrich, M., Li, C., Liu, X., Müller, T. & Kräutler, B. (2015) Stereo- and regioselective phyllobilane oxidation in leaf homogenates of the peace lily (*Spathiphyllum wallisii*): hypothetical endogenous path to yellow chlorophyll catabolites. *Chemistry - A European Journal*, **21**, 136–149.
- de Vries, J. & Archibald, J.M. (2018) Plant evolution: landmarks on the path to terrestrial life. *New Phytologist*, **217**, 1428–1434.
- Wang, P. & Grimm, B. (2021) Connecting chlorophyll metabolism with accumulation of the photosynthetic apparatus. *Trends in Plant Science*, **26**, 484–495.
- Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N. et al. (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, E4859–E4868.
- Wu, C., Akimoto, H. & Ohmiya, Y. (2003) Tracer studies on dinoflagellate luciferin with [<sup>15</sup>N]-glycine and [<sup>15</sup>N]-l-glutamic acid in the dinoflagellate *Pyrocystis lunula*. *Tetrahedron Letters*, **44**, 1263–1266.
- Wuthrich, K.L., Bovet, L., Hunziker, P.E., Donnison, I.S. & Hörtensteiner, S. (2000) Molecular cloning, functional expression and characterisation of RCC reductase involved in chlorophyll catabolism. *The Plant Journal*, **21**, 189–198.
- Xiong, J. & Bauer, C.E. (2002) Complex evolution of photosynthesis. *Annual Review of Plant Biology*, **53**, 503–521.
- Zhang, W., Willows, R.D., Deng, R., Li, Z., Li, M., Wang, Y. et al. (2021) Bilin-dependent regulation of chlorophyll biosynthesis by GUN4. *Proceedings of the National Academy of Sciences of the United States of America*, **118**, e2104443118.
- Zhang, J., Yu, G., Wen, W., Ma, X., Xu, B. & Huang, B. (2016) Functional characterization and hormonal regulation of the PHEOPHYTINASE gene LpPPH controlling leaf senescence in perennial ryegrass. *Journal of Experimental Botany*, **67**, 935–945.
- Zhao, X., Jia, T. & Hu, X. (2020) HCAR is a limitation factor for chlorophyll cycle and chlorophyll b degradation in chlorophyll-b-overproducing plants. *Biomolecules*, **10**, 1639.