New Phytologist

Research

Azolla ferns testify: seed plants and ferns share a common ancestor for leucoanthocyanidin reductase enzymes

Erbil Güngör[1], Paul Brouwer[1,2], Laura W. Dijkhuizen[1], Dally Chaerul Shaffar[1], Klaas G.J. Nierop[2], Ric C.H. de Vos[3], Javier Sastre Toraño[4], Ingrid M. van der Meer[3] and Henriette Schluepmann[1]

[1]Molecular Plant Physiology, Utrecht University, Padualaan 8, Utrecht 3584 CH, the Netherlands; [2]Earth Sciences, Utrecht University, Princetonlaan 8, Utrecht 3584 CB, the Netherlands; [3]Bioscience, Wageningen University & Research, Droevendaalsesteeg 1, Wageningen 6708 PB, the Netherlands; [4]Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht 3508 TB, the Netherlands

Author for correspondence:
Henriette Schluepmann
Email: h.schlupmann@uu.nl

Received: *5 June 2020*
Accepted: *12 August 2020*

New Phytologist (2020)
doi: 10.1111/nph.16896

Key words: *Azolla*, flavonoids, leucoanthocyanidin reductase, *Nostoc*, phenolics, PIP-family reductases, proanthocyanidins, RNA-sequencing.

Summary

- Questions about *in vivo* substrates for proanthocyanidin (PA) biosynthesis and condensation have not been resolved and wide gaps in the understanding of transport and biogenesis in 'tannosomes' persist. Here we examined the evolution of PA biosynthesis in ferns not previously reported, asking what PAs are synthesised and how.
- Chemical and gene-expression analyses were combined to characterise PA biosynthesis, leveraging genome annotation from the floating fern *Azolla filiculoides*. *In vitro* assay and phylogenomics of PIP-dehydrogenases served to infer the evolution of leucoanthocyanidin reductase (LAR).
- Sporophyte-synthesised (epi)catechin polymers, averaging only seven subunits, accumulated to 5.3% in *A. filiculoides*, and 8% in *A. pinnata* biomass dry weight. Consistently, a LAR active *in vitro* was highly expressed in *A. filiculoides*. LAR, and paralogous fern WLAR-enzymes with differing substrate binding sites, represent an evolutionary innovation of the common ancestor of fern and seed plants.
- The specific ecological niche of *Azolla* ferns, a floating plant–microbe mat massively fixing CO_2 and N_2, shaped their metabolism in which PA biosynthesis predominates and employs novel fern LAR enzymes. Characterisation of *in vivo* substrates of these LAR, will help to shed light on the recently assigned and surprising dual catalysis of LAR from seed plants.

Introduction

Azolla is a genus of floating ferns known for the Eocene Azolla event when floating fern mats covered parts of the North Pole Sea and adjacent fresh water bodies: the mats captured substantial amounts of CO_2 as they sunk to the seafloor, presumably contributing to climate cooling (Brinkhuis *et al.*, 2006; Neville *et al.*, 2019). The mats are deemed undesirable by ecologists as they cause anoxia in the water column and outcompete plants that cannot rely on atmospheric dinitrogen (N_2) in freshwater ecosystems (Wagner, 1997). There is renewed interest in *Azolla* fern mats for high-yield protein-rich biomass production in the complete absence of nitrogen fertiliser and, possibly, to build soils in subsiding delta regions (Brouwer *et al.*, 2018; Brouwer *et al.*, 2019). *Azolla* mats, moreover, represent a unique model in which ferns construct a floating ecological niche that includes several symbiotic bacteria (Dijkhuizen *et al.*, 2018). The most prominent symbiont is the obligate filamentous cyanobacterium *Nostoc azollae*, with an eroded genome that fixes sufficient N_2 for the symbiosis, even at the highest growth rates (Ran *et al.*, 2010; Brouwer *et al.*, 2017).

Niche construction was shown to influence the evolution of proanthocyanidin (PA; Fig. 1) production by plant–soil feedback (Arnoldi *et al.*, 2020). The PA content of *Azolla filiculoides* and *Azolla pinnata* was determined by colorimetric assay at 4.4% and a staggering 19.9% of the DW, respectively (Brouwer *et al.*, 2019). PA, therefore, could similarly define the niche of fern mats that contain little to no lignin (Nierop *et al.*, 2011). Fern PA biosynthesis pathways, however, have not been documented. This is unfortunate as PAs constitute a key step in evolution and have not been reported in the other, more ancient, plant lineages such as the bryophytes and the lycophytes (Markham, 1988; Alonso-Amelot *et al.*, 2004).

Symbiotic cyanobacteria in *Azolla* could, in principle, contribute to PA biosynthesis because cyanobacteria contain the flavonoid biosynthesis pathway (FBP) that, in angiosperms, provides the substrates for PA biosynthesis: cyanobacteria have recently been found to synthesise flavonols such as dihydroquercetin and quercetin (Żyszka-Haberecht *et al.*, 2018). By contrast, dihydroflavonol 4-reductase (DFR) may have evolved in the plant lineage (Campanella *et al.*, 2014). Enzymes shared between PA and anthocyanidin biosynthesis, including DFR and

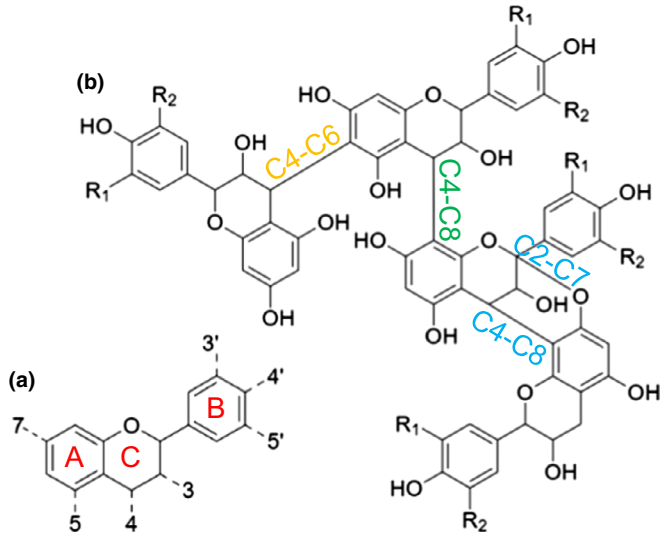


Fig. 1 General structure of flavonoids. Flavonoids are made of the A, B, and C rings and key features of their structures are hydroxyl groups at the numbered carbon atoms within their structure (a). Flavan-3-ols can polymerise into proanthocyanidins, with the C4–C8 (green) being the most widespread carbon–carbon bond (b).

anthocyanidin synthase (ANS), have already evolved in mosses and spike mosses, and have been identified already in ferns but not tested (Campanella *et al.*, 2014).

Evidence for a role for PA and biosynthesis intermediates in the *Azolla* symbiosis interaction is sketchy. The flavanone naringenin, the precursor of most flavonoid classes including PA, was isolated from the symbiosis and shown to stimulate growth of *Nostoc punctiforme*, a free-living facultative plant symbiont related to *N. azollae*; its rhamnoglycoside, naringin, is known to induce inhibition of hormogonia (Cohen *et al.*, 2002; Żyszka *et al.*, 2017). In addition, simple trichomes of the leaf cavities from *Azolla* ferns, known to mediate the interactions between symbiotic bacteria and fern host, contained PA (Pereira & Carrapiço, 2007). To begin understanding the role of PA in *Azolla* symbioses, chemical characterisation and knowledge of the biosynthesis pathway are required.

PA can range from soluble oligomers to large, insoluble polymers with a degree of polymerisation (DP) of over 50 flavan-3-ol subunits; PAs of flavan-3-ol subunits contain both or either catechin and epicatechin (Khanbabaee & van Ree, 2001; Dixon *et al.*, 2004). The flavanol subunits with their characteristic aromatic A- and B-rings (Fig. 1), and defining hydroxyl groups, are produced by enzymes of the FBP (Yonekura-Sakakibara *et al.*, 2019), but enzymes synthesising the recently discovered cysteinyl-linked substrate have not been identified (Liu *et al.*, 2016). Moreover, enzymes catalysing subunit polymerisation into PA have not been identified; polymerisation involving flavonoid carbocations without enzymes occurs at neutral pH unlike the acidic pH typical of the vacuole (Liu *et al.*, 2016; Wang *et al.*, 2020). PAs, however, are generally trapped in vacuoles, thus preventing any interaction with cytosolic proteins (Zhao *et al.*, 2010; Brillouet *et al.*, 2013). Uncertainty about the *in vivo* substrates in PA biosynthesis persists.

Recent advances have revealed a dual role of leucoanthocyanidin reductase (LAR) from an array of seed plants beyond the synthesis of (+)-catechin from leucocyanidin: the enzymes also accept 4β-(*S*-cysteinyl)-epicatechin (Cys-EP) which they cleave to Cys and (−)-epicatechin, the polymer starter subunit (Yu *et al.*, 2019). Cys-EC has been detected in several plant extracts and both genetic and biochemical evidence supports a role for LAR in determining DP (Liu *et al.*, 2016, 2018). Characterising FBP and PA polymerisation in the more ancient lineage of ferns, specifically those accumulating large amounts of PA, may help to close gaps in our current understanding of PA biosynthesis. The improved transcriptome assemblies and orthogroup classifications available from the 1000 plants transcriptomes initiative (1KP), published genome assemblies annotations and transcript profiles of ferns including *A. filiculoides* render the task possible (Brouwer *et al.*, 2017; Li *et al.*, 2018; Wong *et al.*, 2019).

This work, therefore, integrates PA chemical analysis with genomics to reconstruct the biosynthetic pathway in a fern. Using *A. filiculoides* and *A. pinnata*, we first determined flavonoid intermediates in biomass, and subunits in purified PA, using liquid chromatography coupled to mass spectrometry (LC-MS), and thermally assisted hydrolysis and methylation combined with gas chromatography and mass spectrometry (THM-GC-MS). Secondly, we reconstructed a pathway for PA biosynthesis in *A. filiculoides* with the knowledge from the expression of isoenzymes in diel transcript expression profiles, leveraging information from the first *Azolla* genome annotation (Li *et al.*, 2018). We finally focused on the gene of secondary metabolism most prominently expressed in *A. filiculoides* investigating its phylogenetic relationship to known LAR enzymes and the activity of its recombinant protein.

Materials and Methods

Chemicals

Phenolic standards (chlorogenic acid, catechin, naringin, tannic acid and quercetin) and substrates for the enzyme assay (NADPH and dihydroquercetin) were purchased from Sigma-Aldrich. Trace element analysis grade ethyl acetate, dichloromethane and methanol were purchased from VWR International (Amsterdam, the Netherlands), anhydrous acetone from Merck and acetonitrile from Biosolve.

Plant materials and growth conditions

Azolla filiculoides originated from a ditch in Utrecht (the Netherlands; 52.075°N 5.149°E). *Azolla pinnata* was provided by the International Rice Research Institute and originated from Sigiriya in Sri Lanka (IRRI code PI 535 in Watanabe *et al.*, 1992). The plants were grown in 16 h light with an intensity of 100 μmol m^{-2} s^{-1}, at 21°C and 70% humidity. The growth medium contained 0.65 mM $NaH_2PO_4{\cdot}H_2O$, 1.02 mM K_2SO_4, 1 mM $CaCl_2{\cdot}2H_2O$, 1.65 mM $MgSO_4{\cdot}7H_2O$, 17.9 μM Fe-EDTA, 9.1 μM $MnCl_2{\cdot}4\ H_2O$, 1.6 μM $Na_2MoO_4{\cdot}2H_2O$,

18.4 μM H_3BO_3, 0.8 μM $ZnSO_4{\cdot}7H_2O$, 0.3 μM $CuSO_4{\cdot}5H_2O$, 0.2 μM $CoCl_2{\cdot}6H_2O$. The plants were kept in the linear growth phase by continuous harvesting (Supporting Information Fig. S1) after which the biomass was freeze dried immediately. Biomass of both species was collected over a period of 5 wk and pooled. Under these growth conditions, the sporophytes were entirely green.

Purification of proanthocyanidins

Freeze-dried biomass was first extracted in a Soxhlet extractor with dichloromethane : methanol, 9 : 1 for lipid removal. The residual pellet was air dried and extracted three times with aqueous acetone (70% v/v) shaking at 200 rpm for 24 h and pooled, then the mixture was centrifuged at 5000 ***g*** and the supernatant was collected from three replicate extractions. The acetone was evaporated in a rotary evaporator. The remaining aqueous extract was extracted with ethyl acetate yielding the ethyl acetate fraction (EAF). The remaining aqueous extract was then mixed with an equal amount of methanol and loaded onto a column with Sephadex LH-20 resin and the flow-through and subsequent aqueous methanol washes were collected as the flow-through (FT) fraction. The column was washed with aqueous methanol (50% v/v) until the FT was colourless. Finally, elution was performed with aqueous acetone (70% v/v) yielding the SF fraction.

Determination of dry weight, C and N-content

Freeze-dried biomass was weighed as such, while the pellets and the phenolic fractions were first dried for 24 h at 105°C and cooled for 1 h in a desiccator for dry weight determination. Here, 1–2.5 mg of dried material was used for total carbon and nitrogen content determination using an elemental analyser (Fisons NA 1500 CNS).

Colorimetric assay to estimate total phenol content

Freeze-dried biomass, the dried pellets and phenolic fractions were extracted with aqueous acetone (70% v/v) for 24 h shaking at 200 rpm. Folin–Ciocalteu reagent was added to the extract to determine the phenol content, as described by Waterman & Mole (1994). Tannic acid was used as a standard.

Colorimetric assay to estimate proanthocyanidin content

Proanthocyanidin content was estimated using the acid butanol assay: 1.2 ml of butanol-HCl solution (5% v/v) and 20 μl of 2% $NH_4Fe(SO_4)_2$ in 2 N HCl was added to the freeze-dried biomass, the dried pellets and phenolic fractions followed by vigorous mixing and incubation for 50 min at 95°C. The absorbance was measured at 550 nm. Proanthocyanidins from the black spruce *Picea mariana* were used as a standard (Nierop *et al.*, 2005) because this PA is also very rich in (epi)catechin units and, therefore, resembles the reactivity of *Azolla* PA towards the assay.

Estimate of average degree of polymerisation

A vanillin assay as described by Butler *et al.* (1982) was performed on purified proanthocyanidins: the dried sample was dissolved in either methanol or glacial acetic acid, both containing 4% concentrated HCl and 0.5% vanillin. Samples were incubated for 20 min at 30°C, after which the absorbance was measured at 500 nm using a spectrophotometer. The average DP was estimated using the ratio of absorbance in methanol over glacial acetic acid.

Extraction of phenolic compounds in methanol and their analysis using liquid chromatography/mass spectrometry (LC-MS)

Methanolic extracts for LC-MS analysis were obtained by extracting freeze-dried biomass with aqueous methanol (75% v/v) acidified with formic acid (0.1% v/v) as described by De Vos *et al.*, 2007. Methanolic extracts and fractions obtained from proanthocyanidin purification were analysed by LC-MS as described by van der Hooft *et al.* (2012). Briefly, separation of extracted compounds using a Waters Acquity (U)HPLC was over a Luna C18 column (150 mm × 2.0 mm; 3 μm particles; Phenomenex) with a water–acetonitrile gradient in 0.1% v/v formic acid. Detection was by both a Waters Photodiode Array (200–600nm) and a LTQ Orbitrap FT-MS hybrid mass spectrometer (Thermo Fisher Scientific, Nieuwegein, the Netherlands) with negative electrospray ionisation, a mass resolution of 70 000 FWHM and a mass range of *m/z* 100–1300.

Thermally assisted hydrolysis and methylation combined with gas chromatography/mass spectrometry (THM-GC-MS) analysis of phenolic compounds

The method was as described by Nierop *et al.* (2005). Briefly, freeze-dried biomass was pressed onto Curie-Point wires together with a drop of aqueous tetramethylammonium hydroxide (25% v/v) and heated for 5 s at 600°C in a Curie-Point Pyrolyzer (Horizon Instruments, Heathfield, UK). Separation of the products was performed on-line using gas chromatography (Carlo Erba GC8060) on a fused silica column coated with CP-Sil 5 (Varian; 25 m × 0.32 mm; 0.40 μm). The carrier gas was helium. Oven temperature was programmed as 1 min at 40°C, heating to 180°C at a rate of 7°C min^{-1} and a final heating to 320°C at a rate of 20°C min^{-1} which then was held for 15 min. Detection was using the Fisons MD800 mass spectrometer (70 eV electron ionisation, mass range *m/z* 45-650, cycle time 0.7 s). The reference standards are listed in Fig. S2.

Identification of candidate genes of phenylpropanoid pathway (PBP) and FBP

The work was initiated with transcript assembly accessions from Brouwer *et al.*, 2017 and then revised using the first assembly annotation of the *A. filiculoides* Galgenwaard genome (Li *et al.*, 2018). Proteins predicted were accessed at ftp://ftp.fernbase.org/

Azolla_filiculoides/Azolla_asm_v1.1/. FASTA files of both the high- and low-confidence gene models were used for local BLAST (https://www.ncbi.nlm.nih.gov/books/NBK279690/) with sequences of enzymes in the literature experimentally validated. In addition, predictions were used from the KEGG database and MERCATOR tool (https://plabipd.de/portal/mercator-sequence-an notation; Lohse *et al.*, 2014; Kanehisa *et al.*, 2016). For each enzyme of the FBP, proteins identified were then aligned using MAFFT software (https://mafft.cbrc.jp/alignment/software/, Katoh & Standley, 2013) to experimentally validated enzymes, GBLOCKS was used to remove poorly aligned regions (Talavera & Castresana, 2007) and phylogenetic trees were constructed using PHYML (Guindon *et al.*, 2010; Katoh & Standley, 2013). Identified candidate genes were named according to the most closely related known genes of the PBP and FBP.

Phylogenetic analyses of the LAR-like genes

Using the *Vitis vinifera* LAR accession number (GSVIVT01011958001), a precomputed 'LAR-orthogroup' was extracted from the 1KP database v.2 consisting of 6500 protein sequences (http://jlmwiki.plantbio.uga.edu/onekp/v2/, Wong *et al.*, 2019), then re-sampled to include all sequences from key model plants representing land-plant lineages, *Azolla* and *Salvinia* as well as a few other ferns and known PIP enzymes, then *Azolla* LAR and WLAR sequences were added. The final, resulting 808 protein sequences were aligned using MAFFT-linsi (Katoh & Standley, 2013). The alignment was filtered limiting column and sequence gaps using trimAll (Capella-Gutiérrez *et al.*, 2009), leaving 584 unique sequences with 305 parsimony informative sites. From the trimmed alignment, a maximum likelihood phylogenetic tree was computed using IQ-TREE (Nguyen *et al.*, 2015; Trifinopoulos *et al.*, 2016) using extended model fitting (best-fit model LG + R7; Kalyaanamoorthy *et al.*, 2017) and both ultrafast bootstraps and SH-aLRT support (Hoang *et al.*, 2018). Trees were visualised in iTOL (Letunic & Bork, 2019). All details concerning the phylogenetic analyses are available on the GitHub page: github.com/lauralwd/LAR_phylogeny_gungor-et-al-2020.

Analysis of the RNA-sequencing data to reveal transcript accumulation in ferns and Arabidopsis

Azolla RNA-sequencing raw data were for *A. filiculoides* Galgenwaard grown without or with nitrogen (2 mM NH_4NO_3) in the medium and sampled in three biological replicates at 2, 8 and 14 h into the 16 h light period, then at 4 h in the 8 h darkness period of the diel cycle (Brouwer *et al.*, 2017). Arabidopsis RNA-sequencing raw data were from 3-wk-old, unstressed *Arabidopsis thaliana* plants collected at four time points spanning the diel cycle (Coolen *et al.*, 2016). Quality, trimmed, paired-end reads were aligned to the gene models of each plant and only paired matches aligning to only one gene model were counted (CLC Genomics Workbench), as described by Brouwer *et al.* (2017). We realise that for isoenzymes with high similarities some reads were lost as multimappers, leading to a possible underestimation of the absolute read counts for some of these isoenzymes, but this handicap does not change interpretation of differential expression when comparing differing time points. Read counts were normalised to total mapped reads per sample as reads per million (RPM).

Cloning, expression and purification of recombinant *Azfi*LAR is detailed in Methods S1.

Assay of the leucoanthocyanidin reductase

The enzyme reaction contained 5 mM DTT, 0.5 mM dihydroquercetin, 0.5 mM NADPH, 5 μM *Vv*DFR and 5 μM *Azfi*LAR in a total volume of 100 μl with the buffer at pH 6.5, as described by Maugé *et al.*, 2010. The reaction was performed for 2 h at 30°C and stopped by extraction with ethyl acetate and drying overnight. The pellet was stored at −20°C and dissolved in aqueous methanol (50% v/v) before LC-MS analysis, essentially as described by Maugé *et al.* (2010), except that separation was on a C18 column (Varian; 150 × 3.0 mm; 3 μm particles) with an ultra-high performance (UHP) LC instrument coupled to a 6560 Ion Mobility Q-TOF mass spectrometer (Agilent Technologies, Amstelveen, the Netherlands).

Accession numbers

Novel DNA sequences were deposited at EBI's ENA Browser under accession number PRJEB39515.

Results

Ferns from either sections of the *Azolla* genus share a similar phenolic profile

To draw a more general conclusion about PA production in the *Azolla* genus, we first investigated which phenolic compounds could be detected in two *Azolla* species: *A. filiculoides* (section *Azolla*) thriving in temperate regions and *A. pinnata* (section *Rhizosperma*) thriving in (sub)tropical regions (Heckman, 1997; Metzgar *et al.*, 2007).

LC-MS profiles of aqueous methanol extracts from both species were dominated by phenylpropanoid-type caffeoylquinic acids (e.g. peaks 2, 12, 14 and 16; Fig. 2; Table 1). The profiles further contained small amounts of the flavonol glycosides quercetin-3-*O*-rutinoside, quercetin-3-*O*-glucoside and kaempferol-3-*O*-rutinoside (peaks 8–10; Fig. 2; Table 1); compared with *A. filiculoides* these were more abundant in *A. pinnata*, much like in *A. imbricata* that is also a species from the *Rhizosperma* section (Qian *et al.*, 2020). *Azolla pinnata* furthermore contained more (epi)catechin trimer (peak 7; Fig. 2; Table 1). By contrast, THM-GC-MS of the whole biomass yielded a large peak of 1,3,5-trimethoxybenzene characteristic of flavonoids and PA in both *A. filiculoides* and *A. pinnata* (peak 8; Fig. S3; Table S1), and confirming the prevalence of PA reported earlier in *Azolla* ferns (Nierop *et al.*, 2011; Brouwer *et al.*, 2019). In addition, THM-GC-MS profiles of whole biomass contained the THM products typical of caffeoylquinic acids in both species (peaks 5 and 12; Fig. S3; Table S1).

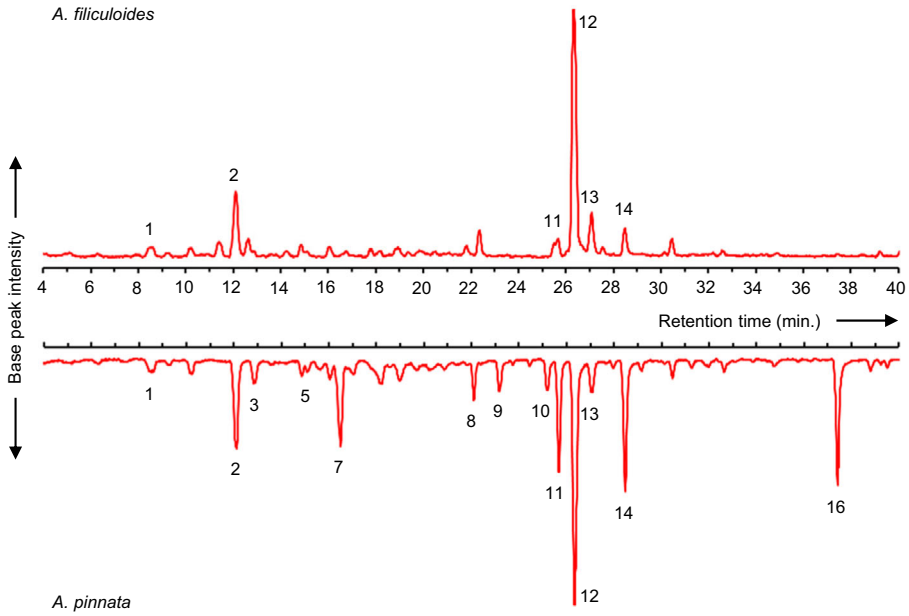


Fig. 2 Representative total ion chromatograms obtained during liquid chromatography-mass spectrometry (LC-MS) analyses of methanolic extracts from *Azolla filiculoides* and *Azolla pinnata* biomass. Base peak intensity (ion counts per second; *y*-axis) of compounds eluting from the column at indicated time after injection (min; *x*-axis). Peak numbers correspond to the compounds listed in Table 1. Upper, extract from *A. filiculoides*; lower, extract from *A. pinnata*.

Table 1 Overview of the (poly)phenolic compounds, their retention time (RT) and their *m/z* as ([M-H]-), that could be assigned to main peaks detected in the LC-MS chromatograms of the *Azolla filiculoides* and *Azolla pinnata* extracts analysed in this study (Figs 2, 4).

Peak number	RT (min)	[M-H]−	Compound
1	8.6	353.0874	1-Caffeoylquinic acid
2	12.1	353.0875	5-Caffeoylquinic acid
3	12.9	353.0875	3-Caffeoylquinic acid
4	13.6	577.1346	(Epi)catechin dimer
5	15.1	353.0875	4-Caffeoylquinic acid
6	15.3	289.0715	Epicatechin monomer
7	16.6	865.1876	(Epi)catechin trimer
8	22.1	609.1455	Quercetin-3-*O*-rutinoside
9	23.2	463.0877	Quercetin-3-*O*-glucoside
10	25.2	593.1504	Kaempferol-3-*O*-rutinoside
11	25.7	515.1192	Dicaffeoylquinic acid I
12	26.3	515.1192	Dicaffeoylquinic acid II
13	27.1	515.1192	Dicaffeoylquinic acid III
14	28.5	515.1192	Dicaffeoylquinic acid IV
15	32.6	515.1192	Dicaffeoylquinic acid V
16	37.4	677.1503	Tricaffeoylquinic acid I
17	38.8	677.1503	Tricaffeoylquinic acid II
18	39.5	677.1503	Tricaffeoylquinic acid III

Phenolic profiles are similar amongst species of the *Azolla* genus, we conclude. In *A. pinnata*, a member of the *Rhizosperma* section, flavonol glycosides, tricaffeyolquinic acids and (epi)catechin make up a higher proportion of the LC-MS profile than in *A. filiculoides*. PAs were not in the range *m/z* 100–1300 detected by the LC-MS profiles of whole-fern extracts. The strong flavonoid signals observed by THM-GC-MS in both species, likely therefore originated from PA with a molecular weight > 1300. To research this hypothesis, we set out to purify PAs from the polyphenol extracts and study these in more detail.

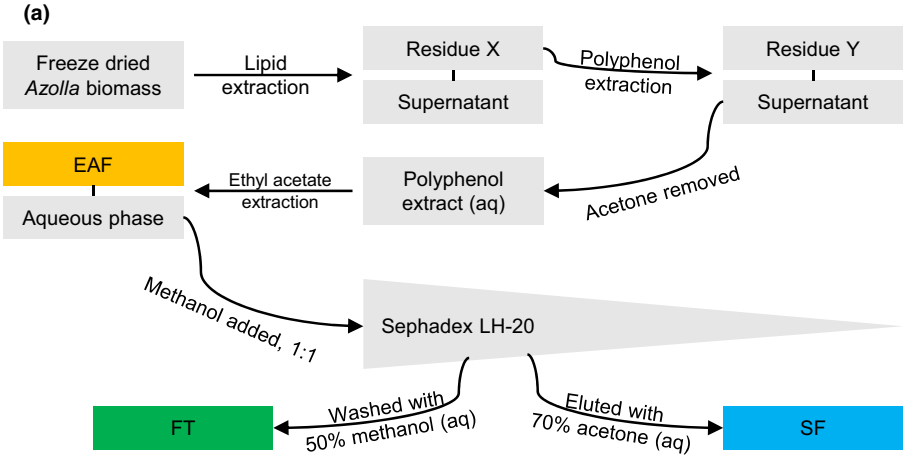
Azolla PA consists mainly of (epi)catechin subunits with an average degree of polymerisation of about seven

To purify PA, *A. filiculoides* and *A. pinnata* whole biomass was extracted according to the scheme in Fig. 3(a); this resulted in three fractions: the EAF, the flow-through of Sephadex LH-20 including washes (FT), and the eluate from Sephadex LH-20 with 70% aqueous acetone (SF). Firstly, the dry weight and the nitrogen and carbon content was determined for each fraction. The FT contained most of the initial biomass, while the SF had the highest carbon content (Fig. 3b), close to that of pure PA from the black spruce *Picea mariana* (59–66%) (Nierop *et al.*, 2005).

Secondly, total phenols and PA content of each fraction were determined using colorimetric assays and expressed in equivalents of tannic acid (polygalloyl esters, hydrolyzable tannins) and black spruce tannin (poly-flavan-3-ols, proanthocyanidins), respectively. In line with the carbon content, the SF were highly enriched in phenols and especially in PA (Fig. 3b). The EAF were also highly enriched in phenols, with much lower levels of PA except for the EAF from *A. pinnata* that contained a significant amount of PA (Fig. 3b). The FT contained relatively low levels of phenols and PA.

Subsequently, the content of each fraction was profiled by both LC-MS and THM-GC-MS. LC-MS showed that the EAF and the FT contained mostly caffeoylquinic acids (Fig. 4; Table 1), also present in the original methanol extract (Fig. 2). THM-GC-MS of both fractions revealed THM products characteristic of caffeoylquinic acids (Fig. S4; Table S1). The caffeoylquinic acids (peaks 16, 17 and 18; Fig. 4; Table 1) were enriched in EAF and FT fractions compared with the original methanol extract (Fig. 3; Table 1). The (epi)catechin trimer (peak 7; Fig. 4; Table 1) in the EAF of *A. pinnata* corresponded to the PA content of this fraction and, therefore, suggested that a proportion of *A. pinnata* PA was extracted with ethyl acetate (Fig. 3b).

(a)



(b)

	Phenolics	PA	Nitrogen	Carbon
Whole *A. filiculoides* biomass	3.4 ± 0.4	5.3 ± 0.2	3.7 ± 0.1	42.8 ± 0.4
Whole *A. pinnata* biomass	5.0 ± 1.2	8.0 ± 0.1	2.9 ± 0.0	41.6 ± 0.7

Fraction	Portion				*A. filiculoides*
Residue X	86 ± 0	na	4.7 ± 0.0	3.7 ± 0.0	33.4 ± 0.4
Residue Y	62 ± 5	na	0.8 ± 0.2	3.3 ± 0.5	27.6 ± 1.7
EAF	1.3 ± 0.4	49 ± 3	10 ± 3	na	na
FT	13.0 ± 1.1	10 ± 7	1.6 ± 0.7	1.0 ± 0.0	19.3 ± 0.5
SF	4.0 ± 0.8	63 ± 11	101 ± 23	0.4 ± 0.1	55.7 ± 3.6
					A. pinnata
Residue X	87 ± 0.0	na	8.3 ± 0.8	2.9 ± 0.0	41.6 ± 0.4
Residue Y	65 ± 0.0	na	0.7 ± 0.3	2.6 ± 0.2	34.7 ± 0.3
EAF	2.7 ± 0.1	53 ± 2	38 ± 15	na	na
FT	9.3 ± 1.1	9.1 ± 5.3	2.3 ± 2.0	1.0 ± 0.1	23.0 ± 2.6
SF	5.8 ± 2.6	60 ± 3	97 ± 12	0.3 ± 0.1	60.5 ± 2.2

Fig. 3 Polyphenol extraction from whole *Azolla* biomass and key features of the fractions obtained from it. (a) Overview of the extraction process, fractions obtained included the lipid extracted residue X, the (poly)phenol extracted residue Y, the ethyl acetate fraction (EAF), the flow-through (FT) and the fraction bound to Sephadex LH-20 then eluted with 70% acetone (SF). (b) Chemical analysis of whole *Azolla filiculoides* and *Azolla pinnata* biomass and the fractions obtained from (a). Total phenolic and proanthocyanidin (PA) content are respectively in percentage tannic acid and pure black spruce PA equivalents. Portion stands for the percentage of the initial biomass in each fraction. Values represent means of three biological replicates with SD. na, not available.

The SF of both *Azolla* species were massively enriched in PA, yielding a large series of PA peaks that were eventually restricted by the LC-MS detection limit of *m/z* 1300 (Fig. 4; Table 1). The THM products derived from flavonoid A- and B-rings, especially the highly abundant 1,3,5-trimethoxybenzene and 1-methyl-2,4,6-trimethoxybenzene (peaks 8 and 10, respectively; Fig. S4; Table S1), were most likely to have been derived from the flavan-3-ol subunits of the PA (Fig. S4; Table S1). Considering in addition the carbon-to-nitrogen ratios and the PA contents determined by colorimetric assay in these fractions, we concluded that the SF of both species contained highly purified PA. The SF furthermore contained large proportions, that is 81% and 86%, of PA found in the initial biomass of *A. filiculoides* and *A. pinnata* respectively (Fig. 3b). Finally, the mass balance and total phenolic content of the fractions, indicated that PAs were the most abundant, soluble, phenolic compounds in these *Azolla* species.

Most flavonoid B-rings detected in the SF of both species with THM-GC-MS had two hydroxyl groups (Fig. S4; Table S1). We therefore concluded that both *A. filiculoides* and *A. pinnata* PAs mainly consisted of catechin and epicatechin subunits. An additional vanillin assay on the SF of both species indicated an average DP of about 7 (data not shown). Although *A. filiculoides* and *A. pinnata* differed slightly in their phenol content, the characteristics of their PAs were the same. We therefore assumed that PA biosynthesis pathway was most likely to be shared amongst species of the *Azolla* genus and focused on the genome of *A. filiculoides* for subsequent identification of enzymes from the FBP.

Azolla filiculoides invests more in transcripts of the FBP than *Arabidopsis thaliana*

The first genome annotation of *A. filiculoides* (Li *et al.*, 2018) served to identify genes involved in PA production. Automatic functional annotation of each candidate gene was systematically verified or corrected with previously identified genes of the

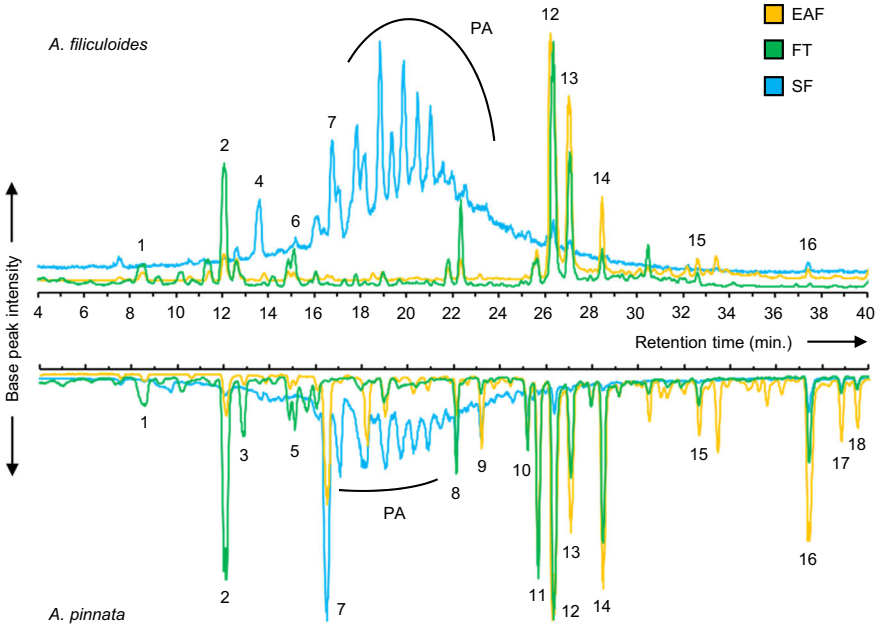


Fig. 4 Liquid chromatography-mass spectrometry (LC-MS) analyses of the three fractions obtained from polyphenol extractions for *Azolla filiculoides* and *Azolla pinnata*. Representative chromatograms of the ethyl acetate (EAF), the Sephadex LH-20 flow-through (FT) and the Sephadex LH-20 eluted (SF) fractions for *A. filiculoides* (upper) and *A. pinnata* (lower). Peak numbers correspond to the compounds listed in Table 1. The large series of peaks in the SF of both species could be assigned to (epi)catechin polymers, thus proanthocyanidin (PA).

phenylpropanoid and the FBP from *A. thaliana* and *Vitis vinifera* (FBP including LAR). Transcript profiles at four time points during the diel cycle of *A. filiculoides* sporophytes (Brouwer *et al.*, 2017) were then compared with those from 3-wk-old and non-stressed control *A. thaliana* plants (Coolen *et al.*, 2016).

Firstly, seven PAL-like, two C4H-like and eight 4CL-like genes were identified (Table S2). All these genes had a higher read count, normalised per million reads and averaged over four time points in the diel cycle, in the ferns compared with *Arabidopsis* plants and, therefore, a higher transcript accumulation in *A. filiculoides* compared with *A. thaliana* (Fig. 5). Secondly, four CHS-like, one CHI-like, two F3H-like, eight F3'H-like, two DFR-like, one FNSII-like, one LAR-like and one FLS/ANS-like genes were identified (Table S2). The latter having a shared phylogeny with both FLS and ANS. Most of these genes had again a higher transcript accumulation in ferns compared with *A. thaliana*, exceptions were FNSII-like and FLS/ANS-like (Fig. 5).

A. filiculoides, we concluded, generally accumulates more transcripts of the phenylpropanoid pathway and the FBP compared with *A. thaliana*. Markedly, *A. filiculoides* LAR-like was the 44th most accumulating transcript of all genes annotated and one of the highest accumulating transcripts of genes encoding enzymes involved in secondary metabolism. Given the role of LAR enzymes in PA biosynthesis in seed plants, the observation was consistent with the large accumulation of PA in *A. filiculoides* measured in Fig. 3. We next wondered whether the expression of LAR-like correlated with that of other genes encoding enzymes for PA biosynthesis.

Candidate genes involved in PA production in *Azolla filiculoides* are co-regulated

As PAs are so abundant in *Azolla*, enzymes for their biosynthesis may dominate the FBP. We therefore analysed the transcript accumulation of PBP and FBP candidate genes at four different time points during the diel cycle in the RNA-sequencing data following Brouwer *et al.*, 2017 and using the genome annotation from Li *et al.*, 2018. Given that many isoenzymes were found for each of the conversions of both the PBP and FBP, and given that we did not know *a priori* to which final product they would contribute, we added up the read counts of the isoenzymes to estimate transcriptional investment over the diel cycle in each enzymatic step. The CHS-like, CHI-like, DFR-like and LAR-like genes shared similar transcript accumulation patterns in ferns under different conditions: transcripts peaked at 08:00, 2 h into the 16 h light period, but did not peak when nitrogen was supplied (Fig. 6a). Generally, transcripts of only one or two isoenzyme genes dominated the read counts for every enzymatic step (Table S2). The shared expression pattern observed when summing up read counts indicated co-regulation of enzymes for PA biosynthesis that dominated the transcriptional investment in secondary metabolism pathways in the ferns. R2R3MYB transcription factors from class VIII-E that generally control FBP and PBP in seed plants were present in *A. filiculoides* (Jiang & Rao, 2020), their read counts were low compared with those from class VIII-D (Fig. 6b,c). Class VIII-D R2R3MYB from the clade containing At*NOECK(MYB106)* were the most highly expressed (Fig. 6d, Azfi_s0001.g000839 and Azfi_s0016.g014344). Correlated changes in read counts and expressed potential R2R3MYB regulators strengthened the hypothesis that *A. filiculoides* LAR-like (*Azfi*LAR-like from this point forward) may be involved in PA production. We therefore investigated this gene in more detail.

*Azfi*LAR-like phylogeny places it at the base of other characterised LARs and its recombinant proteins synthesises catechin *in vitro*

LAR belongs to the PIP protein family, including the enzymes pinoresinol-lariciresinol reductase (PLR), isoflavone reductase

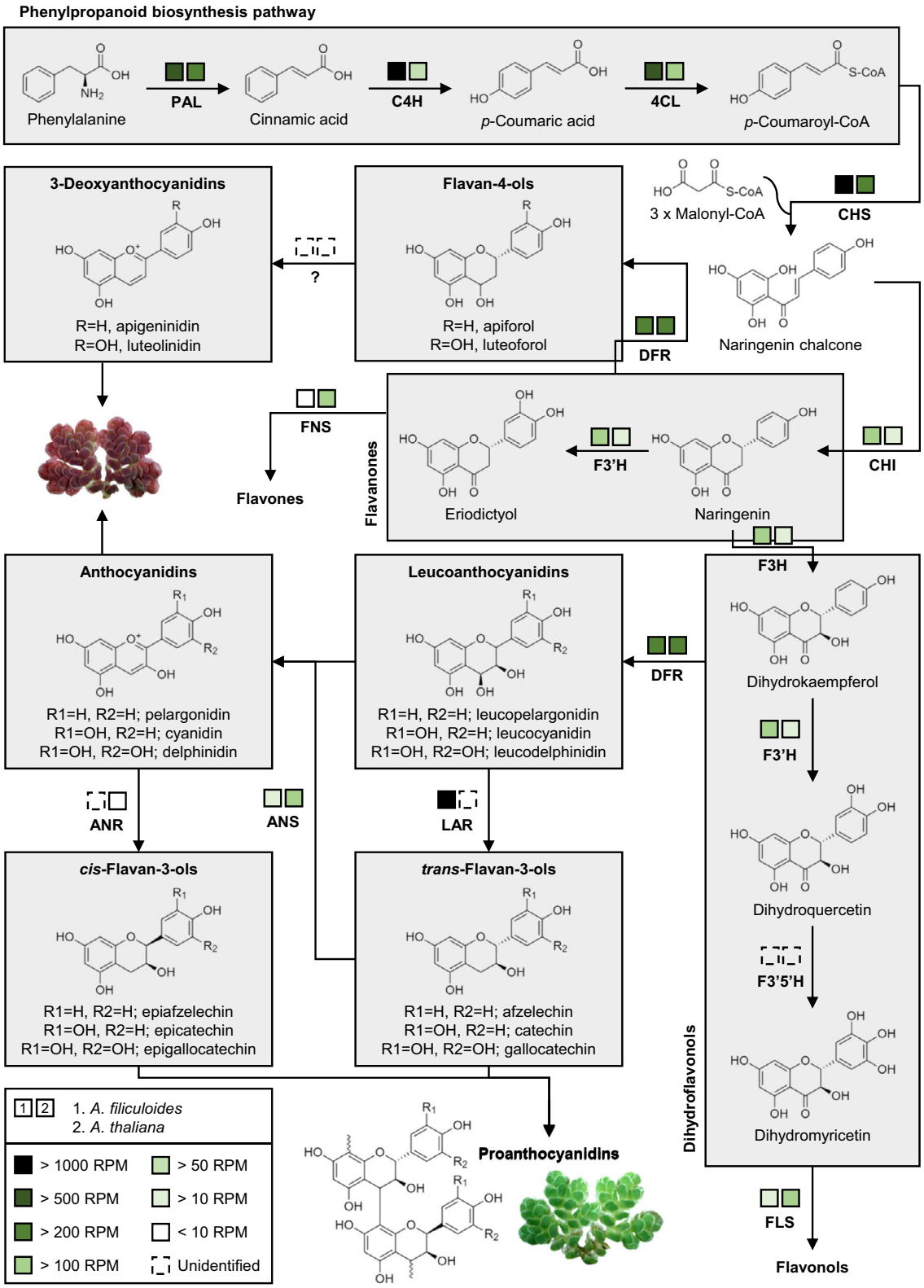


Fig. 5 Transcript accumulation of enzymes involved in the phenylpropanoid and the flavonoid biosynthesis pathways in *Azolla filiculoides* sporophytes and *Arabidopsis thaliana* 3-wk-old plants. The average transcript accumulation, normalised as reads per million (RPM), over four time points during the day was summed up for each isoenzyme of a catalytic step to reflect the transcriptional investment of each plant in each conversion. 4CL, 4-coumarate-CoA ligase; ANR, anthocyanidin reductase; ANS, anthocyanidin synthase; C4H, cinnamic acid 4-hydroxylase; CHI, chalcone isomerase; CHS, chalcone synthase; DFR, dihydrokaempferol 4-reductase; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; F3′5′H, flavonoid 3′5′-hydroxylase; FLS, flavonol synthase; FNS, flavone synthase; IFS, isoflavone synthase; LAR, leucoanthocyanidin reductase; PAL, phenylalanine ammonia-lyase.

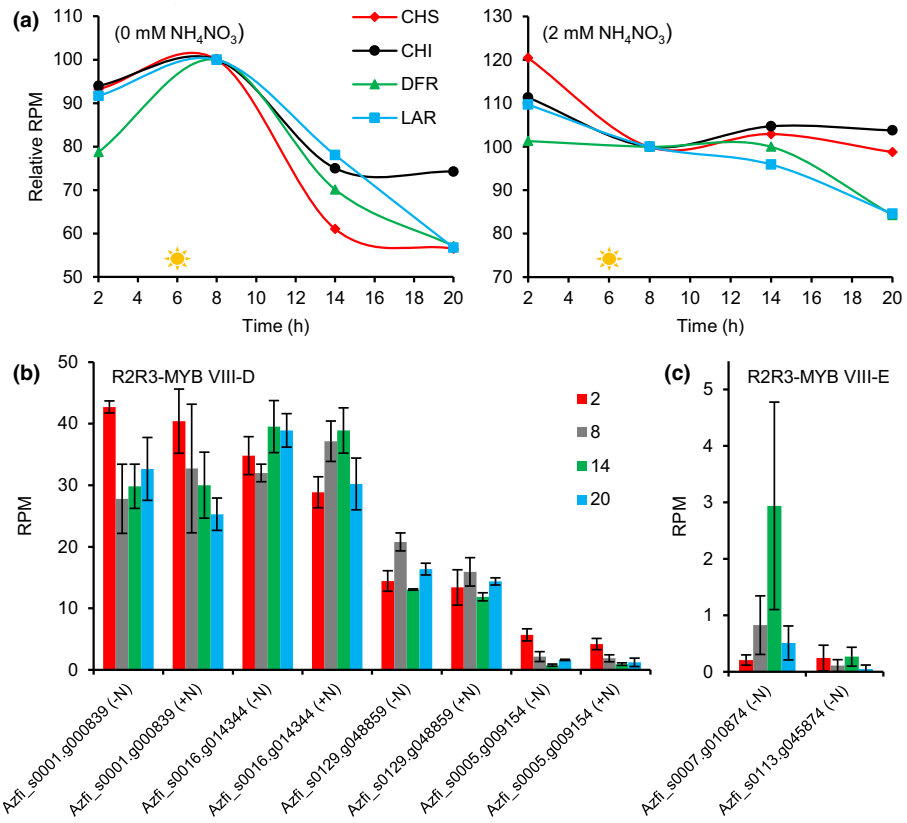


Fig. 6 Changes in transcripts of proanthocyanidin (PA) biosynthesis enzymes and potential R2R3 MYB-regulators during the diel cycle. Raw read data from (Brouwer *et al*., 2017) were mapped to gene models predicted in the *Azolla filiculoides* genome assembly v.1 (Azfi_1; Li *et al*., 2018). Data were available from sporophytes grown without (−N) and with nitrogen (2 mM NH_4NO_3, +N), collected in triplicates at 2 h, 8 h, 14 h and 20 h with the light period of 16 h commencing at 6 h. Transcript accumulation is given as reads per million reads mapping (RPM; Supporting Information Table S2). Changes in transcript accumulation of PA biosynthesis enzymes are relative to 8 h time point in sporophytes without or with nitrogen (a). Transcript accumulation of R2R3-MYB assigned to the VIII classes D (b) and E (c) by Jiang & Rao (2020) are averages from three biological replicates, with SD.

(IFR) and phenylcoumaran benzylic ether reductase (PCBER) (Maugé *et al*., 2010). A phylogenetic tree with *Azfi*LAR-like and other well characterised members of the PIP protein family placed *Azfi*LAR-like at the base of the branch splitting into gymnosperm and angiosperm LAR (Fig. S5; Table S3). *Azfi*LAR-like protein had only 42.6% amino acid identity with *V. vinifera* LAR (*Vv*LAR). 3D-modelling of the proteins revealed that their structures largely overlapped, with minor changes in the active site (data not shown). We therefore proceeded with testing whether recombinant *Azfi*LAR-like protein was able to convert leucoanthocyanidins into *trans*-flavan-3-ols *in vitro*.

*Azfi*LAR-like was cloned, and the protein produced via recombinant protein expression in *E. coli* and purified (Fig. S6). Enzymatic assay was performed in a double reaction with *Vv*DFR because leucoanthocyanidins are highly unstable and not commercially available: *Vv*DFR can produce leucoanthocyanidins from dihydroflavonols such as dihydroquercetin (DHQ) (Petit *et al*., 2007). Products of the enzymatic assay were analysed using LC-MS. Incubation of DHQ with only *Azfi*LAR-like did not yield any new peak (Fig. 7a), while catechin was produced when both *Vv*DFR and *Azfi*LAR-like were present (Fig. 7b). The control incubation of DHQ, with only *Vv*DFR, did not result in the production of catechin (data not shown). Catechin production could therefore only be ascribed to LAR activity by *Azfi*LAR-like (*Azfi*LAR from this point forward).

*Azfi*LAR orthologues are found in many ferns and LAR and PCBER evolved from a common ancestor in the last common ancestor of seed plants and ferns

We next wondered whether LAR enzymes were found in further ferns and other lineages of seed-free land plants. The 1KP orthogroup database v.2 (Wong *et al*., 2019) contained an orthogroup of the *Vitis vinifera* LAR (NP_001267887) which we used for phylogenetic analyses. The orthogroup sequences were subsampled by including all transcripts for each species chosen, and concentrating on species with known genomes and PIP proteins. We monitored changes in trees computed after including firstly more algae species and secondly more species from lycophytes and bryophytes. Typically, green algae IFR/PLR-like enzymes clustered and we, therefore, rooted the phylogenetic tree as shown in Fig. 8 on this algal clade. The orthogroup contained sequences of all known PIP-family dehydrogenases. We observed that the PLR clade, well supported by bootstrapping, evolved early during land-plant evolution and that the IGS&EGS enzymes evolved at least twice (Fig. 8).

The clade containing the PCBER/IFR/group_2_IGS&EGS and LAR enzymes was separated from the PLR clade with over 93.8% aLRT bootstrap support and did not contain sequences from lycophytes, bryophytes and algae (Fig. 8). The well supported PCBER/IFR/group_2_IGS&EGS clade may have originated from LAR-like enzymes in the common ancestor of seed

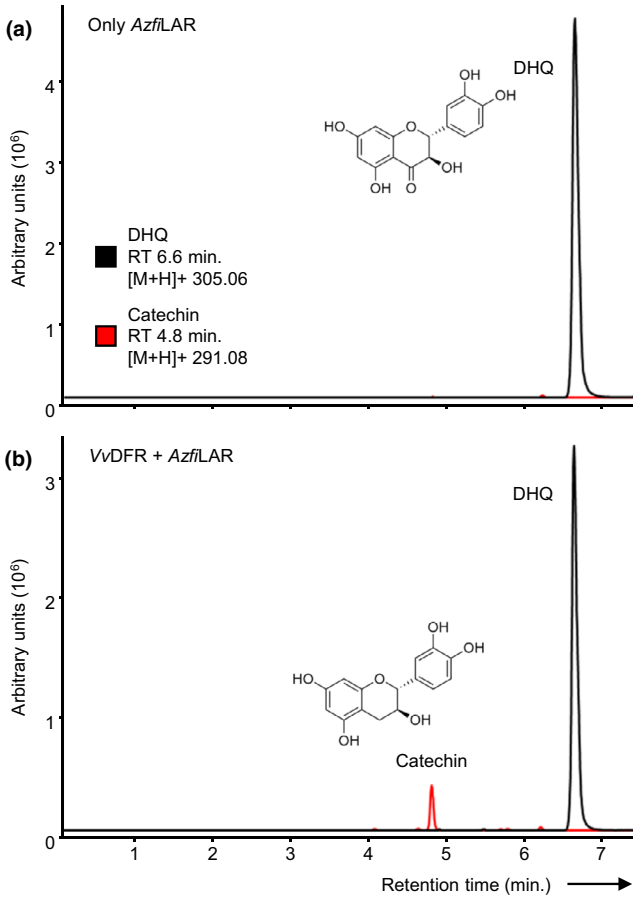


Fig. 7 Leucoanthocyanidin reductase activity of recombinant *Azfi*LAR. (a) Dihydroquercetin (DHQ) was incubated with only recombinant *Azfi*LAR in the presence of NADPH. (b) DHQ was incubated with both recombinant *Azfi*LAR and *Vitis vinifera* DFR (*Vv*DFR) in the presence of NADPH. Representative chromatograms of the catechin and DHQ, respectively. Main ion *m/z* 291.08 and *m/z* 305.06 are shown.

plants and ferns: a cluster of branches with many fern sequences, including *Azfi*LAR and its orthologue in *A. caroliniana*, was found at the base of the well supported seed-plant LAR clade. The *Azfi*LAR clade was well supported with 99.9% aLRT bootstrap and contained sequences from *Adiantum tenereum*, *Ceraptorpteris thalictoides* as well as many other ferns, indicating that LAR activity is more widespread in ferns. Additional fern LAR-like genes were reproducibly found in two separate clades in all trees we computed; we named these clades WannabeLAR_1 (WLAR1) and WannabeLAR_2 (WLAR2) (Fig. 8). Inspection of the alignment of the proteins encoded in *A. caroliniana* transcripts 2016335 and 2016336, *Azfi*LAR and *Vv*LAR, showed that the proteins differed in key amino acids of the active site compared with the known LAR enzymes but not in the NADPH binding site. WLAR1 and WLAR2 enzymes were found in *A. filiculoides* and other ferns (Fig. S7) and are, therefore, not an artefact of 1KP transcript assembly. *Dipteris conjugata*, for example, had transcripts from LAR, WLAR1 and WLAR2 clades (Fig. S7).

Taken together, phylogenetic analysis of the 1KP orthogroup revealed that LAR enzymes are more generally found in ferns and that they evolved in the common ancestor of ferns and seed plants. In addition ferns contain paralogues of the LAR-genes of unknown function.

Discussion

Soluble procyanidins dominated the secondary metabolite content of the biomass from the sporophyte symbiosis. Candidate genes encoding enzymes for PA biosynthesis found in *A. filiculoides* were upregulated in sporophytes and included a LAR, but FLS/ANS-like was not. Recombinant *Azfi*LAR was active *in vitro* and *Azfi*LAR orthologues were commonly found in fern species but were not detected in other seed-free plant lineages. Additional WLAR were found with differing substrate binding sites from LAR.

The procyanidin type PA accumulating in *Azolla* ferns have an optimum degree of polymerisation for protein aggregation and possibly, therefore, a role as antifeedant

PA could be purified with high yield and characterised from both *A. filiculoides* and *A. pinnata*, and their structures were similar: procyanidins with an average DP estimated at seven (Figs 3b, 4; Table 1). Assay of a large diversity of PA for protein aggregation revealed that the DP is a key determinant for the capacity of procyanidin type PA to aggregate proteins in solution, with stiff globular proteins, such as bovine serum albumin, requiring more PA than the more flexible and less compacted gelatin (Ropiak *et al.*, 2017). Moreover, the smallest PA with largest effect on protein aggregation had a DP of seven which happens to also be the average DP of *Azolla* PA. We concluded, therefore, that *Azolla* PA may have evolved to optimally aggregate proteins and thus may have an ecological role as antifeedant. Consistently, PA formed precipitating PA–protein aggregates during protein extraction from *Azolla* biomass which reduced the protein yields from extractions (Leizerovich *et al.*, 1988; Zeller *et al.*, 2015; Brouwer *et al.*, 2019). This also explained why *Azolla* biomass with about 20% DW crude protein may only be included in animal diets in limited proportions without affecting growth (Leterme *et al.*, 2009; Acharya *et al.*, 2015; Brouwer *et al.*, 2019).

Staining with vanillin–hydrochloric acid confirmed localisation of PA in vacuoles of specialised trichomes located inside the leaf cavity where *N. azollae* fix nitrogen (Pereira & Carrapiço, 2007). The PAs therefore may have additional roles in the interaction with symbionts and as redox-active compounds important in the 'after-life' processes in the *Azolla* mats. More research is needed, manipulating levels of individual compounds in the pathways, to evaluate physiological and ecological roles of PA and PA biosynthesis intermediates in *Azolla*. Such manipulations may best be done by specifically altering key biosynthetic enzymes *in vivo*.

The common ancestor of ferns and angiosperms already contained all enzymes of the angiosperm PA biosynthesis pathway

Genes of all enzymes of PA biosynthesis characterised from angiosperms were found in the genome of *A. filiculoides* except

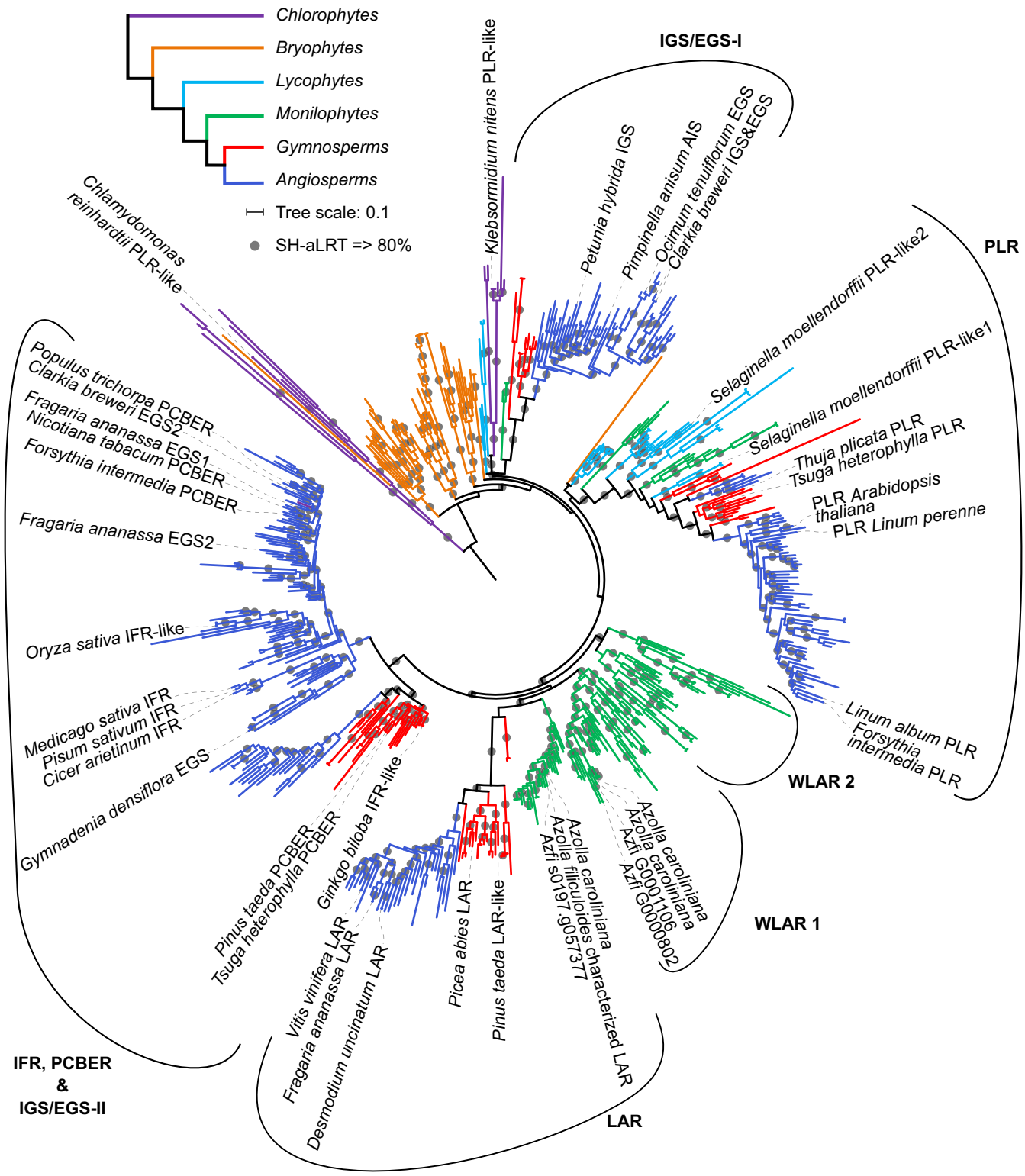


Fig. 8 Phylogeny of LAR and other PIP-family enzymes across land-plant lineages. Protein sequences were retrieved as a single orthogroup containing the *Vitis vinifera* LAR from the 1KP orthogroup database_2 (Wong *et al*., 2019), subsampled and supplemented with guide and *Azolla filiculoides* LAR-like sequences. The 808 sequences were aligned with MAFFT-linsi (Katoh *et al*., 2019), then trimmed using trimAL (Capella-Gutiérrez *et al*., 2009) and, using iQ-TREE (Nguyen *et al*., 2015), the phylogenetic tree computed with a resulting 305 parsimony informative sites, a best-fit substitution model LG + R7 and 1000 bootstrap determined by SH-aLRT (Kalyaanamoorthy *et al*., 2017). The tree was annotated in iTOL (Letunic & Bork, 2019; details of all sequences are given in Supporting Information Fig. S8). Nodes with bootstrap support equal or > 80% SH-aLRT are indicated by circles. EGS and IGS clustered in two groups: IGS/EGS-I and IGS/EGS-II. Abbreviations: EGS, eugenol synthase; IFR, isoflavone reductase; IGS, isoeugenol synthase; LAR, leucoanthocyanidin reductase; PCBER, phenylcoumaran benzylic ether reductase; PLR, pinoresinol-lariciresinol reductase; WLAR, fern-specific LAR-like.

for F3′5′H, in line with the lack of flavan-3-ols with three hydroxyl groups attached to the flavonoid B-ring, and ANR (Fig. 6). Presence of either F3′5′H, or ANR in *Azolla* cannot be completely excluded, however, due to the incomplete nature of our first fern assembly and annotation reported in Li *et al.*, 2018. Results are consistent with our chemical analyses, and earlier examples that detected flavone and flavonol derivatives in various *Azolla* species (Teixeira *et al.*, 2001; Qian *et al.*, 2020). *Azolla* was shown, in addition, to synthesise red-coloured 3-deoxyanthocyanins correlating with snail and tadpole deterrence (Cohen *et al.*, 2002a); the fern turned completely red upon 3-deoxyanthocyanin accumulation, for example when exposed to cadmium that led to CHS and DFR-like transcript accumulation in *Azolla imbricata* (Dai *et al.*, 2012). Steps leading to final 3-deoxyanthocyanin formation remain uncharacterised, final steps might be followed by DFR-like reduction or ANS-like dehydrogenation/dehydroxylation reactions, and subsequent glycosylation (Liu *et al.*, 2010).

The phenylpropanoid, flavonol and lignin biosynthesis pathways are widely distributed in seed-free plant lineages (Tohge *et al.*, 2013; De Vries *et al.*, 2017). Specifically, the products of PLR, the lignans, have been described in bryophytes and lycophytes, as well as other vascularised plants (Koeduka, 2018; Markulin *et al.*, 2019) consistent with results in Fig. 8. By contrast, PA appear to be absent in bryophytes and lycophytes (Davies *et al.*, 2020). LAR, therefore, could have evolved from a PLR.

Recent analyses aimed at evaluating the production of antioxidants in cyanobacteria, revealed the accumulation of catechin and epicatechin in several *Nostoc* and *Anabaena* species, with catechin present in all filamentous cyanobacteria tested (Blagojević *et al.*, 2018). Moreover, a DFR was annotated from a *N. punctiforme* strain associated with lichen (Tabuchi Yagui., NCBI-accession RCJ41608). In our hands, searches for DFR identified the *N. punctiforme* gene reproducibly, but returned none from the *N. azollae* genome, searches for LAR were fruitless. We concluded, therefore, that enzymes until DFR were likely to have already evolved in cyanobacteria, but that *N. azollae* had lost the capability.

Interestingly, synthesis of some of the flavonols was repressed when cyanobacteria were grown under conditions in which they needed to fix nitrogen (Blagojević *et al.*, 2018), suggesting that they may be important for the N_2-fixation process, possibly as antioxidants. Synthesis of flavonols for *N. azollae* that are locked into nitrogen fixation in the leaf cavities, may thus have been taken over by the fern host in the *Azolla* symbiosis. Some fungi synthesise leucoanthocyanidins, suggesting that these metabolites are not restricted to cyanobacteria and plant lineages (Bu *et al.*, 2020). A thorough analysis of cyanobacterial FBP enzymes will reveal whether enzymes such as DFR appeared in the plant, fungal and cyanobacterial lineages by convergent evolution or horizontal gene transfer.

*Azfi*LAR evolved in the common ancestor of ferns and seed plants, but do fern LAR also control the DP of PA?

Although ferns were known to accumulate PA, neither LAR nor ANR had been identified in this lineage before (Alonso-Amelot *et al.*, 2004; Nierop *et al.*, 2011). The *Azfi*LAR recombinant protein exhibited LAR activity *in vitro* and phylogenetic analyses placed it at the base of the LAR leaf containing characterised LAR from seed plants (Figs 8, S5). Apparent absence of PA in bryophytes and lycophytes, and our phylogenetic analyses suggested that LAR could be a key evolutionary innovation in the last common ancestor (LCA) of ferns and seed plants. This was suggested in spite of the fact that enzyme activity of PIP proteins from a single clade may not be predicted with certainty (Koeduka, 2018; Markulin *et al.*, 2019).

Additionally, two paralogous clades of the *Azfi*LAR containing WLAR1 and WLAR2 were identified in ferns. Given their differing substrate binding site from genuine LAR and PCBER (Fig. S7), more work is needed to understand the function of these novel fern-specific reductases. These enzymes are particularly interesting, however, because seed-plant LAR enzymes catalyse two different reactions, possibly catalysed by distinct enzymes in lineages that evolved earlier in land-plant evolution. *Vv*LAR and *Mt*LAR convert Cys-EP into epicatechin as well as leucoanthocyanidins into *trans*-flavan-3-ols (Liu *et al.*, 2016; Yu *et al.*, 2019).

In vitro, epicatechin produced by LAR from Cys-EP is required as starter unit to which epicatechin carbocations condense sequentially, bringing about polymerisation. The LAR regulate the DP of PA, therefore, by influencing the amount of starter. *Azfi*LAR may similarly encode an enzyme cleaving Cys-EP. Consistent with the hypothesis is the low DP of PA and the absence of a highly expressed candidate gene for ANR in *Azolla*. We concluded that the high *Azfi*LAR expression in *A. filiculoides* may maintain the DP optimum for antifeedant function and will need to test this hypothesis in the near future. Sadly, we have yet to establish mutagenesis or genome editing protocols for *Azolla* to verify the roles of specific enzymes and regulators of the PA pathway *in vivo*.

Coordinate expression of the FBP enzymes is also found in seed plants and suggests a conserved mode of regulation

The coordinate expression observed during the diel cycle of isoenzymes of the FBP (Fig. 6) would be consistent with a regulating transcription factor complex controlling anthocyanin accumulation, as it exists in dicotyledonous plants (Albert *et al.*, 2014; Albert, 2015). Light regulation of anthocyanin biosynthesis in *A. thaliana* is well known and mediated by regulating the stability of the activating R2R3-MYB transcription factors PAP1 and PAP2 that are part of the MYB-bHLH-WD40 (MBW) complexes more generally regulating differentiation of trichomes, as well as expression of genes in secondary metabolism (Bulgakov *et al.*, 2017). Class VIII-E R2R3MYB were, however, much less expressed than class VIII-D, the MIXTA1-type known to specify glandular trichome development (Shi *et al.*, 2018; Jiang & Rao, 2020). Glandular trichomes are abundant and critical to attract *N. azollae* in the specialised structures of the *Azolla* fern host (Peters & Perkins, 1993).

That transcripts of the FBP enzymes begin accumulating in the dark when nitrogen is supplied, or soon after the start of the

light period when *N. azollae* uses light energy to fixate N_2 is consistent with observations in cyanobacteria in which flavonol biosynthesis is upregulated when nitrogen is available (Brouwer *et al.*, 2017; Blagojević *et al.*, 2018). Identification of the MYB and bHLH transcription factors that control the FBP pathway will be of particular use to engineer ferns with altered levels of PA in order to study the specific relation of PA characteristics and quantities with symbiont development, ecological fitness and biomass digestibility. Such studies have begun successfully in symbiotic seed-plant species, such as species of clover (Liu *et al.*, 2018).

Acknowledgements

We thank Jean Chaudière from Bordeaux University for sharing the *Vitis vinifera* DFR recombinant expression vector. We thank Adrie van der Werf from Wageningen University & Research for recruiting IvdM and RdV for the task of running the LC-MS analyses of methanolic extracts from our ferns, and Caroline M. Preston (Pacific Forestry Centre, Victoria BC, Canada) for providing the purified black spruce tannins. We further thank three anonymous reviewers for their very helpful comments. We acknowledge funding for EB from the NWO-TTW grant AZOPRO (Project no. 16294), PB by the LPP foundation and LWD by the NWO-ALW project 2016/ALW/00127599.

Author contributions

PB grew plants, purified tannins and analysed the fractions with the help of KN, RdV and IvdM. PB, EG and HS analysed genes of the condensed tannin biosynthesis pathways, DCS subcloned *Azfi*LAR, EG expressed and tested recombinant *Azfi*LAR, helped by JST for MS analyses of the products. LWD, EG and HS computed PIP protein phylogenies. EG, PB and HS wrote the manuscript with input from all other authors; all authors agreed to the list of authors and the identified contributions of those authors.

ORCID

Laura W. Dijkhuizen https://orcid.org/0000-0002-4628-7671
Henriette Schluepmann https://orcid.org/0000-0001-6171-3029

References

Acharya P, Mohanty GP, Pradhan CR, Mishra SK, Beura NC, Moharana B. 2015. Exploring the effects of inclusion of dietary fresh *Azolla* on the performance of White Pekin broiler ducks. *Veterinary world* **8**: 1293–1299.

Albert NW. 2015. Subspecialization of R2R3-MYB repressors for anthocyanin and proanthocyanidin regulation in forage Legumes. *Frontiers of Plant Science* **6**: 1165.

Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, Brendolise C, Boase MR, Ngo H, Jameson PE, Schwinn KE. 2014. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *The Plant Cell* **26**: 962–980.

Alonso-Amelot ME, Oliveros A, Calcagno-Pisarelli MP. 2004. Phenolics and condensed tannins in relation to altitude in neotropical *Pteridium* spp. *Biochemical Systematics and Ecology* **32**: 969–981.

Arnoldi J-F, Coq S, Kéfi S, Ibanez S. 2020. Positive plant–soil feedback trigger tannin evolution by niche construction: a spatial stoichiometric model. *Journal of Ecology* **108**: 378–391.

Blagojević D, Babić O, Rašeta M, Šibul F, Janjušević L, Simeunović J. 2018. Antioxidant activity and phenolic profile in filamentous cyanobacteria: the impact of nitrogen. *Journal of applied phycology* **30**: 2337–2346.

Brillouet J-M, Romieu C, Schoefs B, Solymosi K, Cheynier V, Fulcrand H, Verdeil J-L, Conéjéro G. 2013. The tannosome is an organelle forming condensed tannins in the chlorophyllous organs of Tracheophyta. *Annals of Botany* **112**: 1003–1014.

Brinkhuis H, Schouten S, Collinson ME, Sluijs A, Damsté JSS, Dickens GR, Huber M, Cronin TM, Onodera J, Takahashi K *et al.* 2006. Episodic fresh surface waters in the Eocene Arctic Ocean. *Nature* **441**: 606–609.

Brouwer P, Bräutigam A, Buijs VA, Tazelaar AOE, van der Werf A, Schlüter U, Reichart G-J, Bolger A, Usadel B, Weber APM *et al.* 2017. Metabolic adaptation, a specialized leaf organ structure and vascular responses to diurnal N_2 fixation by *Nostoc azollae* sustain the astonishing productivity of *Azolla* ferns without nitrogen fertilizer. *Frontiers of Plant Science* **8**: 442.

Brouwer P, Nierop KGJ, Huijgen WJJ, Schluepmann H. 2019. Aquatic weeds as novel protein sources: alkaline extraction of tannin-rich *Azolla*. *Biotechnology Reports* **24**: e00368.

Brouwer P, Schluepmann H, Nierop KGJ, Elderson J, Bijl PK, van der Meer I, de Visser W, Reichart G-J, Smeekens S, van der Werf A. 2018. Growing *Azolla* to produce sustainable protein feed: the effect of differing species and CO_2 concentrations on biomass productivity and chemical composition. *Journal of the Science of Food and Agriculture* **98**: 4759–4768.

Bu C, Zhang Q, Zeng J, Cao X, Hao Z, Qiao D, Cao Y, Xu H. 2020. Identification of a novel anthocyanin synthesis pathway in the fungus *Aspergillus sydowii* H-1. *BMC Genomics* **21**: 29.

Bulgakov VP, Avramenko TV, Tsitsiashvili GS. 2017. Critical analysis of protein signaling networks involved in the regulation of plant secondary metabolism: focus on anthocyanins. *Critical Reviews in Biotechnology* **37**: 685–700.

Butler LG, Price ML, Brotherton JE. 1982. Vanillin assay for proanthocyanidins (condensed tannins): modification of the solvent for estimation of the degree of polymerization. *The Journal of Agricultural and Food Chemistry* **30**: 1087–1089.

Campanella JJ, Smalley JV, Dempsey ME. 2014. A phylogenetic examination of the primary anthocyanin production pathway of the Plantae. *Botanical Studies* **55**: 1–10.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Cohen MF, Meziane T, Tsuchiya M, Yamasaki H. 2002a. Feeding deterrence of *Azolla* in relation to deoxyanthocyanin and fatty acid composition. *Aquatic Botany* **74**: 181–187.

Cohen MF, Sakihama Y, Takagi YC, Ichiba T, Yamasaki H. 2002. Synergistic effect of deoxyanthocyanins from symbiotic fern *Azolla* spp. on *hrma* gene induction in the cyanobacterium *Nostoc punctiforme*. *Molecular Plant-Microbe Interactions* **15**: 875–882.

Coolen S, Proietti S, Hickman R, Davila Olivas NH, Huang P-P, Van Verk MC, Van Pelt JA, Wittenberg AHJ, De Vos M, Prins M *et al.* 2016. Transcriptome dynamics of Arabidopsis during sequential biotic and abiotic stresses. *The Plant Journal* **86**: 249–267.

Dai L-P, Dong X-J, Ma H-H. 2012. Molecular mechanism for cadmium-induced anthocyanin accumulation in *Azolla imbricata*. *Chemosphere* **87**: 319–325.

Davies KM, Jibran R, Zhou Y, Albert NW, Brummell DA, Jordan BR, Bowman JL, Schwinn KE. 2020. The evolution of flavonoid biosynthesis: a bryophyte perspective. *Frontiers of Plant Science* **11**: 7.

De Vos RCH, Moco S, Lommen A, Keurentjes JJB, Bino RJ, Hall RD. 2007. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols* **2**: 778–791.

Dijkhuizen LW, Brouwer P, Bolhuis H, Reichart G-J, Koppers N, Huettel B, Bolger AM, Li F-W, Cheng S, Liu X *et al.* 2018. Is there foul play in the leaf pocket? The metagenome of floating fern *Azolla* reveals endophytes that do not fix N_2 but may denitrify. *New Phytologist* **217**: 453–466.

Dixon RA, Xie D-Y, Sharma SB. 2004. Proanthocyanidins - a final frontier in flavonoid research? *New Phytologist* **165**: 9–28.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**: 307–321.

Heckman CW. 1997. Ecoclimatological survey of the wetland biota in the tropical wet-and-dry climatic zone. *Global Ecology and Biogeography* **6**: 97.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Le SV. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**: 518–522.

van der Hooft JJJ, Vervoort J, Bino RJ, de Vos RCH. 2012. Spectral trees as a robust annotation tool in LC–MS based metabolomics. *Metabolomics* **8**: 691–703.

Jiang CK, Rao GY. 2020. Insights into the diversification and evolution of R2R3-MYB transcription factors in plants. *Plant physiology* **183**: 637–655.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A, Jermiin LS. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587–589.

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

Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Khanbabaee K, van Ree T. 2001. Tannins: Classification and definition. *Natural Product Reports* **18**: 641–649.

Koeduka T. 2018. Functional evolution of biosynthetic enzymes that produce plant volatiles. *Bioscience, Biotechnology and Biochemistry* **82**: 192–199.

Leizerovich I, Fleminger G, Kardish N, Frensdorff A, Galun M. 1988. Polyphenols, and not lectins, are responsible for hemagglutinating activity in extracts of *Azolla filiculoides* Lam. *Symbiosis* **5**: 209–222.

Leterme P, Londoño AM, Muñoz JE, Súarez J, Bedoya CA, Souffrant WB, Buldgen A. 2009. Nutritional value of aquatic ferns (*Azolla filiculoides* Lam. and *Salvinia molesta* (Mitchell) in pigs. *Animal Feed Science and Technology* **149**: 135–148.

Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* **47**: W256–W259.

Li F-W, Brouwer P, Carretero-Paulet L, Cheng S, de Vries J, Delaux P-M, Eily A, Koppers N, Kuo L-Y, Li Z *et al.* 2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants* **4**: 460–472.

Liu C, Ha CM, Dixon RA. 2018. Functional genomics in the study of metabolic pathways in *Medicago truncatula*: an overview. *Methods in Molecular Biology* **1822**: 315–337.

Liu C, Wang X, Shulaev V, Dixon RA. 2016. A role for leucoanthocyanidin reductase in the extension of proanthocyanidins. *Nature Plants* **2**: 16182.

Liu H, Du Y, Chu H, Shih CH, Wong YW, Wang M, Chu IK, Tao Y, Lo C. 2010. Molecular dissection of the pathogen-inducible 3-deoxyanthocyanidin biosynthesis pathway in sorghum. *Plant and Cell Physiology* **51**: 1173–1185.

Lohse M, Nagel A, Herter T, May P, Schroda M, Zrenner R, Tohge T, Ar Fernie, Stitt M, Usadel B. 2014. Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, Cell & Environment* **37**: 1250–1258.

Markham KR. 1988. Distribution of flavonoids in the lower plants and its evolutionary significance. In: Harborne JB, ed. *The flavonoids*. Boston, MA, USASpringer, 427–468.

Markulin L, Corbin C, Renouard S, Drouet S, Gutierrez L, Mateljak I, Auguin D, Hano C, Fuss E, Lainé E. 2019. Pinoresinol–lariciresinol reductases, key to the lignan synthesis in plants. *Planta* **249**: 1695–1714.

Maugé C, Granier T, D'Estaintot BL, Gargouri M, Manigand C, Schmitter J-M, Chaudière J, Gallois B. 2010. Crystal structure and catalytic mechanism of leucoanthocyanidin reductase from *Vitis vinifera*. *Journal of Molecular Biology* **397**: 1079–1091.

Metzgar JS, Schneider H, Pryer KM. 2007. Phylogeny and divergence time estimates for the fern genus *Azolla* (Salviniaceae). *International Journal of Plant Sciences* **168**: 1045–1053.

Neville LA, Grasby SE, McNeil DH. 2019. Limited freshwater cap in the Eocene Arctic Ocean. *Scientific Reports* **9**: 1–6.

Nguyen LT, Schmidt HA, Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution* **32**: 268–274.

Nierop KGJ, Preston CM, Kaal J. 2005. Thermally assisted hydrolysis and methylation of purified tannins from plants. *Analytical Chemistry* **77**: 5604–5614.

Nierop KGJ, Speelman EN, de Leeuw JW, Reichart G-J. 2011. The omnipresent water fern *Azolla caroliniana* does not contain lignin. *Organic Geochemistry* **42**: 846–850.

Pereira AL, Carrapiço F. 2007. Histochemistry of simple hairs from the foliar cavities of *Azolla filiculoides*. *Plant Biosystems* **141**: 323–328.

Peters GA, Perkins SK. 1993. The *Azolla-Anabaena* symbiosis: endophyte continuity in the *Azolla* life-cycle is facilitated by epidermal trichomes: II. Re-establishment of the symbiosis following gametogenesis and embryogenesis. *New Phytologist* **123**: 65–75.

Petit P, Granier T, D'Estaintot BL, Manigand C, Bathany K, Schmitter J-M, Lauvergeat V, Hamdi S, Gallois B. 2007. Crystal structure of grape dihydroflavonol 4-reductase, a key enzyme in flavonoid biosynthesis. *Journal of Molecular Biology* **368**: 1345–1357.

Qian W, Wu W, Kang Y, Wang Y, Yang P, Deng Y, Ni C, Huang J. 2020. Comprehensive identification of minor components and bioassay-guided isolation of an unusual antioxidant from *Azolla imbricata* using ultra-high performance liquid chromatography—quadrupole time-of-flight mass spectrometry combined with multicomponent knoc. *Journal of Chromatography A* **1609**: 460435.

Ran L, Larsson J, Vigil-Stenman T, Nylander JA, Ininbergs K, Zheng WW, Lapidus A, Lowry S, Haselkorn R, Bergman B. 2010. Genome erosion in a nitrogen-fixing vertically transmitted endosymbiotic multicellular cyanobacterium. *PLoS ONE* **5**: e11486.

Ropiak HM, Lachmann P, Ramsay A, Green RJ, Mueller-Harvey I. 2017. Identification of structural features of condensed tannins that affect protein aggregation. *PLoS ONE* **12**: e0170768.

Shi P, Fu X, Shen Q, Liu M, Pan Q, Tang Y, Jiang W, Lv Z, Yan T, Ma Y *et al.* 2018. The roles of Aa MIXTA 1 in regulating the initiation of glandular trichomes and cuticle biosynthesis in *Artemisia annua*. *New Phytologist* **217**: 261–76.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systems Biology* **56**: 564–577.

Teixeira G, Carrapiço F, Gomes ET. 2001. C-glycosylflavones in the genus Azolla. *Plant Biosystems* **135**: 233–237.

Tohge T, Watanabe M, Hoefgen R, Fernie AR. 2013. The evolution of phenylpropanoid metabolism in the green lineage. *Critical Reviews in Biochemistry and Molecular Biology* **48**: 123–152.

Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* **44**: W232–W235.

de Vries J, de Vries S, Slamovits CH, Rose LE, Archibald JM. 2017. How embryophytic is the biosynthesis of phenylpropanoids and their derivatives in streptophyte algae? *Plant and Cell Physiology* **58**: 934–945.

Wagner GM. 1997. *Azolla*: a review of its biology and utilization. *Botanical Review* **63**: 1–26.

Wang P, Liu Y, Zhang L, Wang W, Hou H, Zhao Y, Jiang X, Yu J, Tan H, Wang Y *et al.* 2020. Functional demonstration of plant flavonoid carbocations proposed to be involved in the biosynthesis of proanthocyanidins. *The Plant Journal* **101**: 18–36.

Watanabe I, Roger PA, Ladha JK, Van Hove C. 1992. *Biofertilizer germplasm collections at IRRI*. Los Baños, PhilippinesIRRI.

Waterman PG, Mole S. 1994. *Analysis of phenolic plant metabolites. (Methods in Ecology)*. Oxford, UKBlackwell Scientific Publications.

Wong GK-S, Soltis DE, Leebens-Mack J, Wickett NJ, Barker MS, de Peer YV, Graham SW, Melkonian M. 2019. Sequencing and analyzing the

transcriptomes of a thousand species across the tree of life for green plants. *Annual Review of Plant Biology* **18**: 71.

Yonekura-Sakakibara K, Higashi Y, Nakabayashi R. 2019. The origin and evolution of plant flavonoid metabolism. *Frontiers in Plant Science* **10**: 943.

Yu K, Jun JH, Duan C, Dixon RA. 2019. VvLAR1 and VvLAR2 are bifunctional enzymes for proanthocyanidin biosynthesis in grapevine. *Plant Physiology* **180**: 1362–1374.

Zeller WE, Sullivan ML, Mueller-Harvey I, Grabber JH, Ramsay A, Drake C, Brown RH. 2015. Protein precipitation behavior of condensed tannins from *Lotus pedunculatus* and *Trifolium repens* with different mean degrees of polymerization. *Journal of Agricultural and Food Chemistry* **63**: 1160–1168.

Zhao J, Pang Y, Dixon RA. 2010. The mysteries of proanthocyanidin transport and polymerization. *Plant Physiology* **153**: 437–443.

Żyszka B, Anioł M, Lipok J. 2017. Modulation of the growth and metabolic response of cyanobacteria by the multifaceted activity of naringenin. *PLoS ONE* **12**: e0177631.

Żyszka-Haberecht B, Poliwoda A, Lipok J. 2018. Biocatalytic hydrogenation of the C=C bond in the enone unit of hydroxylated chalcones—process arising from cyanobacterial adaptations. *Applied Microbiology and Biotechnology* **102**: 7097–7111.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Photographs of typical *Azolla* sporophytes and their continuous culturing setup.

Fig. S2 Partial THM-GC-MS chromatograms of the main (poly)phenolic standards.

Fig. S3 Partial THM-GC-MS chromatograms of whole *A. filiculoides* and *A. pinnata* biomass.

Fig. S4 THM-GC-MS of fractions from polyphenol extractions for *A. filiculoides* and *A. pinnata*.

Fig. S5 Phylogenetic relationship of *A. filiculoides* LAR and characterised PIP enzymes.

Fig. S6 SDS-PAGE of the proteins used in the enzymatic assay of Fig. 7.

Fig. S7 Alignments of LAR, PCBER and related sequences (WLAR) from ferns and *V. vinifera*.

Fig. S8 Phylogeny of LAR and other PIP-family enzymes across land-plant lineages.

Methods S1 Cloning, expression and purification of recombinant *Azfi*LAR.

Table S1 Compounds detected with THM-GC-MS in *Azolla* biomass during this study.

Table S2 Transcript accumulation of phenylpropanoid/flavonoid isoenzymes and relevant R2R3-MYB transcription factors in the *A. filiculoides* genome.

Table S3 PIP enzymes used to generate the phylogenetic tree in Fig. S4.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.