A flavin-dependent monooxygenase catalyzes the initial step in cyanogenic glycoside synthesis in ferns

Sara Thodberg 1,2,7, Mette Sørensen 1,2,7, Matteo Bellucci 3, Christoph Crocoll 4, Amalie Kofoed Bendtsen 1,2, David Ralph Nelson 5, Mohammed Saddik Motawia 1,2,6, Birger Lindberg Møller 1,2,6 & Elizabeth Heather Jakobsen Neilson 1,2,6

Cyanogenic glycosides form part of a binary plant defense system that, upon catabolism, detonates a toxic hydrogen cyanide bomb. In seed plants, the initial step of cyanogenic glycoside biosynthesis—the conversion of an amino acid to the corresponding aldoxime—is catalyzed by a cytochrome P450 from the CYP79 family. An evolutionary conundrum arises, as no CYP79s have been identified in ferns, despite cyanogenic glycoside occurrence in several fern species. Here, we report that a flavin-dependent monooxygenase (fern oxime synthase; FOS1), catalyzes the first step of cyanogenic glycoside biosynthesis in two fern species (Phlebodium aureum and Pteridium aquilinum), demonstrating convergent evolution of biosynthesis across the plant kingdom. The FOS1 sequence from the two species is near identical (98%), despite diversifying 140 MYA. Recombinant FOS1 was isolated as a catalytic active dimer, and in planta, catalyzes formation of an N-hydroxylated primary amino acid; a class of metabolite not previously observed in plants.

1 Plant Biochemistry Laboratory, Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Copenhagen, Denmark. 2 VILLUM Center for Plant Plasticity, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Copenhagen, Denmark. 3 Novo Nordisk Foundation Center for Protein Research, Protein Production and Characterization Platform, University of Copenhagen, Blegdamsvej 3, 2200 Copenhagen N, Denmark. 4 Section for Plant Molecular Biology, Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Copenhagen, Denmark. 5 Department of Microbiology, Immunology and Biochemistry, University of Tennessee, 858 Madison Ave. Suite G01, Memphis, TN 38163, USA. 6 Center for Synthetic Biology, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Copenhagen, Denmark. 7 These authors contributed equally: Sara Thodberg, Mette Sørensen. *email: en@plen.ku.dk
Plants produce a plethora of natural products (phytochemicals or specialized metabolites) enabling interactions with their biotic and abiotic environment. Cyanogenic glycosides are one such class of amino acid-derived natural products present in more than 3000 plant species, including ferns, gymnosperms, and angiosperms1–3. For example, the cyanogenic glycosides prunasin and amygdalin are responsible for the bitterness of wild almond (Prunus dulcis)1–3. Upon tissue disruption, cyanogenic glycosides are hydrolyzed by specific β-glucosidases resulting in the formation of a hydrogen cyanide bomb as an immediate toxic chemical response e.g. towards chewing herbivores4. More recently, cyanogenic glycosides have been shown to possess alternative functions as remobilizable storage molecules of reduced nitrogen, controllers of bud break and flower induction, and as quenchers of reactive oxygen species5–9.

In higher plants (gymnosperms and angiosperms), the biosynthesis of cyanogenic glycosides is catalyzed by cytochromes P450 (CYPs) and UDP-glucosyltransferases (UGTs)10,11,12. In all cases, initial conversion of the parent amino acid to the corresponding E-oxime is catalyzed by a functionally conserved CYP79 family enzyme. Independently evolved CYP71, CYP706, or CYP736 enzymes convert the oxime into an α-hydroxynitrile10,12,13 that is glycosylated by a UGT85 or UGT94 family member to produce the cyanogenic mono- or diglucosides14,15,16 (Fig. 1).

The CYP79-catalyzed reaction proceeds via two N-hydroxylation events, a decarboxylation, and a dehydration reaction in a single catalytic site15–17. No other enzymes across the plant kingdom are known to catalyze the conversion of an α-amino acid into an oxime. In addition to their involvement as intermediates in cyanogenic glycoside biosynthesis, CYP79-formed oximes are important metabolites in general and specialized metabolism18–20 exemplified by indole-3-acetaldoxime, which is a shared precursor for the phytotoxin auxin (indole-3-acetic acid, IAA), the phytoalexin camalexin and tryptophan-derived glucosinolates19–21. An evolutionary conundrum arises as no gene sequences encoding CYP79s have been found in fern transcriptomes or genomes22,23 despite the occurrence of cyanogenic glycosides in ferns1,24. Due to their significant phylogenetic position, ferns represent an important lineage for studying the evolution of land plants. Fern research has been hampered by the scarcity of genome information22. In total, 11 orders of ferns are known of which four are extant: Psilotopsida, Equisetopsida, Marattiopsida, and Polypoetiopsida25. Polypoetiopsida are termed modern ferns, and ~3% of the species in this order have been reported as cyanogenic1,24,26.

Here, we investigate the cyanogenic glycoside biosynthetic pathway in modern ferns by a differentially expression survey of de novo assembled transcriptomes from Pteridium aquilinum and Pteridium aquilinum. We report biochemical and biological evidence that ferns harbor an N-hydroxylating flavin-dependent monoxygenase that converts phenylalanine to a corresponding oxime via N-hydroxyphenylalanine. This demonstrates convergent evolution at the biochemical pathway level and resolves how ferns produce cyanogenic glycosides in the absence of CYP79 encoding genes.

**Results**

**Metabolite-guided pathway discovery in ferns.** The two distantly related modern ferns, Phlebodium aureum (Polydaceae) and Pteridium aquilinum (Dennstaedtiaceae) (Fig. 2), produce the phenylalanine-derived cyanogenic monoglucoside prunasin (D-mandelonitrile-β-D-glucopyranoside)1,26,27 and the diglucoside vicianin (6-O-arabinopyranosylglucopyranoside)28, respectively. Targeted metabolite profiling of a population of 25 field-collected *P. aquilinum* (Paq) identified individuals with high and low cyanogenic glycoside containing pinnae ranging from 0.2 to 0.6 mg prunasin g−1 fw (Fig. 2d). Two individuals were selected based on their metabolite content and the quality of RNA extracted. Similarly, analysis of different tissues within a single *Phlebodium aureum* (Pa) fern identified variable levels of vicianin in the different tissue types from negligible levels in the spores, to 5 mg vicianin g−1 fw in the emerging fiddlehead (Fig. 2c, Supplementary Figure 1). For Pa, fiddlehead and young pinna were selected for transcriptome analysis (Fig. 2c). mRNA was isolated from these four tissues to obtain biosynthetic gene candidates using a comparative transcriptomic approach.

BLAST searches of the de novo assembled transcriptomes from Paq and Pa identified 139 CYPs in *Paq* compared to 120 in *Pa* (Supplementary Table 1). Neither genes encoding CYP79s nor CYPs with the CYP79 signature F/H substitution in the PERF region were identified. Additional searches of the gene sequences deposited in the OneKP database31 confirmed the previously reported absence of CYP79 encoding gene sequences in ferns (Fig. 3)22,23. In the absence of CYP candidates, we extended the search to other gene families encoding monooxygenases with focus on genes showing interesting differential expression patterns of gene homologs within and between the two fern species.

The contig Pa22758 encoded a flavin-dependent monooxygenase ORF of 543 aa in accordance with a predicted full-length sequence32. Deep-mining of the transcriptome reads revealed that *Pa* contained variants of Pa22758 (Supplementary Fig. 2). The transcript level of this gene was 24-fold higher (−logFDR of 4.32) in Pa fiddlehead compared to the young pinnae (Fig. 2e). A reciprocal BLAST search between the predicted flavin-dependent monoxygenases of Paq and Pa identified a contig Paq18302 in *Paq* harboring a full-length nucleotide sequence encoding a flavin-dependent monooxygenase identical to the sequence of Pa22758, except for a single amino acid substitution. The expression level of Paq18302 showed a 23% increase (−logFDR of 0.50) in the individual containing higher levels of prunasin (Fig. 2e). The deduced amino acid sequences of the *Pa* and *Paq* flavin-dependent monoxygenases revealed that they contain conserved motifs specific to Class B flavin-dependent monoxygenases (Fig. 2f, Supplementary Fig. 2)33.

A BLAST search in the CATH: Protein Structure Classification Database (cathdb.info, version 4.2) with the Paq18302 protein sequence, classified the fern sequences as belonging to the Class B, flavin-dependent monoxygenases. These enzymes are characterized by being single-component FAD-binding enzymes harboring...
binding sites for the hydride electron donor NAD(P)H and molecular oxygen (CATH code 3.50.50.60).32,34,35

**Biochemical characterization of fern oxime synthase 1 (FOS1) in planta.** Functional characterization of the FMO enzyme Paq18302 (Fig. 4a) was obtained by *Agrobacterium tumefaciens*-mediated transient expression of the encoding gene in *Nicotiana benthamiana*. Leaf discs of agro-infiltrated tissue were harvested after 4 days and subjected to metabolite profiling. Expression of the Paq18302 FMO afforded production of two constituents of \( m/z \) eluting at rt = 8.17 and 8.70 min corresponding to the [M+H]+ adduct of (E)- and (Z)-phenylacetaldoxime, respectively, as verified by co-elution with an authentic standard (Fig. 4a). Two additional constituents were identified as a glucoside of phenylacetaldoxime (\( m/z \) 320, [M+Na]+ at rt = 6.7 min) and as a phenylacetaldoxime glucoside-malonic acid conjugate (\( m/z \) 406, [M+Na]+ at rt = 7.9 min) based on the diagnostic fragments in the MS/MS spectra (Fig. 4c, Supplementary Fig. 3).36–38. Identical oxime derivatives have previously been
Fig. 2 Cyanogenic glycoside content and FMO transcript abundances in tissues from the two modern fern species _Pteridium aquilinum_ and _Phlebodium aureum_. a Two phenylalanine-derived cyanogenic glycosides have been reported from ferns: the monoglucoside prunasin (D-mandelonitrile-β-D-glucopyranoside) and the diglycoside vicianin (6-O-arabinofuranosylglucopyranoside). b Phylogenetic relationship between _Pteridium aquilinum_ (Dennstaedtiaceae) and _Phlebodium aureum_ (Polypodiaceae) showing that these two modern ferns species diversified 140 million years ago (tree adapted from [20]). c Content of vicianin across different tissue types of _P. aureum_ (see also Supplementary Fig. 1), with the blue bars indicating the tissue selected for downstream transcriptomic analysis. d Content of prunasin in the pinnae of a population of 25 field-collected _P. aquilinum_, with the green bars indicating individuals selected for transcriptomic analysis. e Transcript abundance of predicted flavin monoxygenases (FMOs) in tissues of _P. aquilinum_ and _P. aureum_ containing high (gray bars) or low (black bars) cyanogenic glycoside levels. Arrows indicate candidate genes. TPM transcripts per million mapped reads. f Schematic illustration of identity and motifs in the transcriptome-deduced amino acid sequences of Paq18302 (PaqFOS1) and Paq22578. Differences between Paq22578 and the isolated PaqFOS1 are indicated by blue lines. The position of putative binding motifs for FAD and NADPH, the FATGY and FMO-identifier motifs conserved across plant FMOs are highlighted. Supporting alignment is shown in Supplementary Fig. 1.

Fig. 3 Schematic diagram illustrating the occurrence of CYP families across plant taxa, based on the known CYP families present in eudicots. The diagram is based on analysis of the OneKP database [3], which includes transcriptomes from 74 ferns. The presence of 8930 cytochromes P450 fern sequences were predicted and sorted into families in accordance with nomenclature. The analyzed fern transcriptomes harbor at least 81 different P450 families of which 49 (60%) are novel fern-specific families. Approximately half of the CYP families present in higher plants are also found in ferns. The CYP79 family is present from gymnosperm to eudicots, but based on the >40% sequence identity, the CYP79 is absent in ferns. This also applies to the other known CYP families involved in cyanogenic glucoside biosynthesis in plants: the CYP71, CYP706, and CYP736 families.

observed in _N. benthamiana_ in response to expression of CYP79 enzymes [10]. Guided by Pa22758, the FMO encoding sequence from _Pa_ was isolated from cDNA and shown by transient expression to be functionally equivalent to the Paq18302 FMO (Fig. 4, Supplementary Fig. 4). We designate the orthologous FMO proteins as “FOS1”.Transient expression of FOS1 in _N. benthamiana_ did not give rise to formation of other oximes or additional products, demonstrating that FOS1 has phenylalanine as its specific amino acid substrate. Independent functional characterization of the FOS1 genes was obtained in _N. benthamiana_ by co-expression with CYP71AN24 and UGT785A19, encoding the last two enzymes in the prunasin biosynthetic pathway in almond ( _Prunus dulcis_)[4,14]. This resulted in production of prunasin as demonstrated by the formation of a constituent comigrating with an authentic standard with an extracted ion chromatogram (EIC) of m/z 318 corresponding to the [M+Na]+ adduct of prunasin (Fig. 4b, Supplementary Fig. 4). Neither phenylacetaldoloxime nor prunasin are endogenous constituents of _N. benthamiana_.

In vitro assays revealed existence of _N_-hydroxyphenylalanine. Of the limited number of FMOs characterized from plants, only YUCCA6 from _A. thaliana_ and AsFMO from garlic ( _Allium sativum_ ) have been successfully expressed and purified in enough quantity for downstream biochemical analyses [33,39,40]. Here, we expressed the PaqFOS1 protein with an N-terminal 6xHIS tag using _Escherichia coli_ as a heterologous host and isolated PaqFOS1 by immobilized metal affinity chromatography followed by size exclusion chromatography (SEC) with a yield of 0.7 mg/L culture (Fig. 5). The isolated PaqFOS1 protein binds the cofactors FAD and NADPH as demonstrated by absorbance spectrometry, and migrates with an apparent molecular mass of 60 kDa on SDS-PAGE (Fig. 5) in agreement with a calculated molecular mass of 62.5 kDa. Upon SEC, PaqFOS1 eluted with a mass of 120 kDa suggesting that the native protein is a homodimer. In vitro assays of the isolated PaqFOS1 protein followed by LC–MS analyses confirmed that FOS1 can catalyze the conversion of phenylalanine to phenylacetaldoloxime (Fig. 6). Concomitant production of _N_-hydroxyphenylalanine was also observed and verified using a chemically synthesized standard (Supplementary Fig. 8). Surprisingly, targeted LC–MS analysis of fiddleheads of _Pa_ as well as of _N. benthamiana_ leaves transiently expressing FOS1 demonstrated the in vivo presence of _N_-hydroxyphenylalanine in these tissues (Fig. 6, Supplementary Fig. 5).

Discussion

In this report, we identify and functionally characterize a flavin-dependent monoxygenase designated FOS1 from two modern fern species, _Pa_ and _Paq_. Metabolite profiling of _N. benthamiana_ leaves following transient expression of PaqFOS1 revealed the production of phenylacetaldoloxime, which was also confirmed by the targeted in vitro experiments. When jointly expressed with CYP71AN24 and UGT85A19 from almond ( _Prunus dulcis_), the entire prunasin pathway was established.

The FOS1 enzyme belongs to FMO class B, which are single-component FAD-binding enzymes that also harbor binding sites for the hydride electron donor NAD(P)H and molecular oxygen [34,41]. PaqFOS1 was isolated and purified as a functionally active homodimer. In vitro assays demonstrated that FOS1 is able to convert its substrate phenylalanine into two products, _N_-hydroxyphenylalanine and phenylacetaldoloxime. The first product is obtained by a single _N_-hydroxylation reaction, the second by two consecutive _N_-hydroxylations followed by decarboxylation and dehydration reactions. In plants, the conversion of an amino acid to the corresponding aldoxime has only been reported as catalyzed by P450s from the CYP79 family. It is to be noted that an FMO from the actinomycete fungus _Streptomyces coelicolor_ A3 can convert tryptophan and C5 prenylated tryptophan into their
corresponding aldoximes. Free N-hydroxyphenylalanine was also present in the fern tissue of Paq. N-hydroxylated protein amino acids have to our knowledge not previously been detected in biological tissues, and their functional roles remains unknown.

The 98% conservation of the FOS1 encoding gene sequences of Paq and Paq is remarkable. Among the total number of 55,000 and 63,000 gene sequences present in the transcriptomes of Paq and Paq, respectively, only 2% (1297 transcripts) share a sequence identity higher than 95%. Modern ferns diversified into the Pteridoids (including Paq) and Eupolypoids (including Paq) 140 million years ago (Fig. 2b). A blast search of the FOS1 gene sequence against the 68 fern transcriptomes available in the oneKP identified the presence of an identical transcript in Phlebodium pseudouareum, a close relative to Paq. Deep mining of the transcriptomic resources documented the presence of homologous sequences across the evolutionary gap between these species (Supplementary Tables 2 and 3). If a common ancestral FOS1 sequence was present in a progenitor to the derived ferns, corresponding aldoximes. Free N-hydroxyphenylalanine was also present in the fern tissue of Paq. N-hydroxylated protein amino acids have to our knowledge not previously been detected in biological tissues, and their functional roles remains unknown.

The 98% conservation of the FOS1 encoding gene sequences of Paq and Paq is remarkable. Among the total number of 55,000 and 63,000 gene sequences present in the transcriptomes of Paq and Paq, respectively, only 2% (1297 transcripts) share a sequence identity higher than 95%. Modern ferns diversified into the Pteridoids (including Paq) and Eupolypoids (including Paq) 140 million years ago (Fig. 2b). A blast search of the FOS1 gene sequence against the 68 fern transcriptomes available in the oneKP identified the presence of an identical transcript in Phlebodium pseudouareum, a close relative to Paq. Deep mining of the transcriptomic resources documented the presence of homologous sequences across the evolutionary gap between these species (Supplementary Tables 2 and 3). If a common ancestral FOS1 sequence was present in a progenitor to the derived ferns,
representative functionally characterized FMOs: AsFMO1 from garlic (A. sativum) catalyzing S-oxygenation of allyl-mercaptoacetate, the A. thaliana YUCCAs involved in the biosynthesis of the phytormone auxin35, the AtFMOGS-OX1,35,36 performing S-oxygenation of methylthioalkyl glucosinolates35, and A. thaliana AfFMO1 that catalyzes N-hydroxylation of piperocolic acid to form N-hydroxypiperocic acid, the critical signaling molecule in systemic acquired resistance34,35,36. The analysis shows that plant FMOs cluster in three phylogenetically distinct groups (Fig. 7), with each harboring members from all evolutionary distinct species from Selaginella (moss) to A. thaliana (angiosperm). This suggests an evolutionary split of the FMOs prior to emergence of the early land plants.

By adding fern FMOs to a plant FMO-specific phylogenetic tree, the role of the other putative fern monooxygenases can be hypothesized (Fig. 2e, Supplementary Fig. 6). The analysis identified six transcripts, three orthologs from each fern species, to cluster in the YUCCA clade. To date all characterized YUCCAs are involved in auxin biosynthesis, as they are proposed to catalyze a decarboxylation of indole-3-pyruvate acid to form IAA. An additional role of thiol reductase activity has been linked to these enzymes33,35. The establishment of a YUCCA-like mediated auxin function for fern FMOs would contribute to the evolutionary perspectives of hormone biosynthesis and signaling. PaFOS1 and PaqFOS1 group together with two additional FMO contigs (Pa22435 and Paq33416; Fig. 2e, Supplementary Fig. 6). This group of FMOs also encapsulates the AtFMO1 catalyzing N-hydroxylation of piperocolic acid34 (Fig. 7). Based on the phylogeny and functional characterization, we suggest that the group encapsulating PaFOS1, PaqFOS1, and AtFMO1 catalyze N-hydroxylation reactions. The PaqFOS1 and the AtFMO1 share 40% amino acid sequence identity (209/518) with a similarity score of 61% (318/518). As suggested by the SEC-elution profile, PaqFOS1 elutes solely as a homodimer. In parallel to PaqFOS1, we also expressed and isolated AtFMO1 and demonstrated that it also elutes as a homodimer (Supplementary Fig. 7). A dimeric FMO protein has previously been isolated and crystallized from the methylotrophic bacterium Methylobacterium58. Most recently, the crystal structures and proposed dimeric arrangement of class B FMOs from multicellular organisms were reported for the pyrrolizidine alkaloid N-oxygenase (ZvPNO) from the Locust grasshopper (Zonocerus variegatus)59 and for ancestral reconstructed mammalian FMOs. Based on the sequence similarities and crystal structures obtained, and in agreement with the isolation of PaqFOS1 and AtFMO1 as stable homodimers, the tertiary structure of the class B FMOs are predicted to be conserved across the kingdoms of life.

Here, we show that an FMO catalyzing N-hydroxylation of an α-amino acid plays a key role in the convergent evolution of cyanogenic glycoside biosynthesis in ferns and higher plants. In all currently investigated gymnosperms and angiosperms, oxime formation from amino acids is catalyzed by a cytochrome P450 from the CYP79 family18. Our study demonstrates that the introduction of cyanogenic glycoside biosynthesis in ferns was based on independent recruitment of a unique class B FMO protein.

In addition to lacking a CYP79, ferns also lack the other cyanogenic glycoside-related CYP families: CYP71, CYP736, and CYP706 (Fig. 3). These families are all members of the large 71 clan, and selected members of these families catalyze conversion of oximes into the cyanohydrin intermediate in seed plants (Fig. 1). This opens speculation on the possible identity of the remaining biosynthetic pathway members in ferns. Based on less than 40% amino acid sequence identity criteria, many of the CYPs identified in the Pa and Paq transcriptomes did not correspond to any previously named and characterized P450
This study therefore unmasks a treasure trove of what is to our knowledge new CYP families and possible pathway candidates. Interestingly, other families of the 71 clan are also highly abundant in fern species, and would be possible targets for gene discovery. The diversity of previously unnamed CYP families accentuates convergent evolution of cyanogenic glycoside biosynthesis as well as other metabolic pathways in ferns. Our comparative transcriptomic strategy to identify FOS1 was highly successful, and would provide a robust approach to identify the remaining pathway members.

The evolution of the cyanogenic glycoside pathway is quite dynamic. Recently, the classical three-step pathway was revised, with the discovery that sugar gum (Eucalyptus cladocalyx) harbors four genes that catalyze the conversion of amino acid to

---

**Fig. 6** The formation of N-hydroxyphenylalanine (NOH-phe). LC-MS chromatograms from targeted analysis of samples compared with the authentic standard, derived from a transient expression of FOS1 in N. benthamiana showing the p19 as a negative control, b heterologous expression of FOS1 in E. coli using the absence of NADPH as a negative control, and c the presence of N-hydroxyphenylalanine in P. aureum tissue. The chromatographic trace represents the abundance of the fragment ion of N-hydroxyphenylalanine (precursor ion → fragment ion of 182.1 → 136.0; see Supplementary Table 4).

d The hypothesized biosynthetic route from phenylalanine to phenylacetaldoxime, as catalyzed by FOS1. e In vitro activity assay of recombinant PaqFOS1 using L-phenylalanine (Phe) as a substrate, with different combinations of the necessary cofactors FAD and NADPH. Bars represent mean ± SE (n ≥ 3).

f The presence of N-hydroxyphenylalanine in P. aureum fiddlehead and young pinnae metabolite extracts (n = 3).
Fig. 7 Phylogenetic tree of the flavin monoxygenase (FMO) superfamily containing all predicted full-length FMOs from ferns (P. aureum and P. aquilinum) together with FMOs from eight higher plant species. As all species are represented in each of the tree clades, the phylogenetic analysis suggests an early diversification of the groups prior to species differentiation. Employed sequence IDs are compiled in Supplementary Data 2. Characterized FMOs (or clusters of all characterized FMOs such as the Arabidopsis YUCCAs) are indicated by a star.

cyanogenic glycoside. It is therefore also a possibility that more (or less) biosynthetic steps and intermediates might be present in ferns. Further, the cyanogenic glycoside biosynthetic pathway is shown to act as a dynamic metabolon, ensuring channeling of intermediates. Here, Sorghum bicolor metabolons encounter the soluble UGT into tight organization. A likewise orchestration of FOS1 into a membrane-bound complex could be plausible, and soluble UGT into tight organization. A likewise orchestration of N-hydroxy amino acid intermediate. Furthermore, this example of convergent evolution by the recruitment of different enzyme families—specifically showing that a soluble FMO can catalyze the same reaction as a membrane-bound cytochrome P450—opens opportunities for industrial applications in the future.

Methods

Plant material. Plant material for metabolite and RNA extraction was obtained from P. aureum (P. aureum and P. aquilinum) grown in glass house at the Botanical Garden of Copenhagen (Plant ID E615) and from Pteridium aquilinum (Paq) collected at Dronningens Borge, Erslev Lake, Nødskov, on June 17, 2015 (55°39′ 49.97′N, 11°22′ 57.47′E). Tissues were snap-frozen in liquid nitrogen and stored at −80 °C until further analyses.

Metabolite profiling of fern tissue. Plant tissue (~30 mg) was weighed and boiled in 85% methanol (v/v, 300 μL) for 5 min. The vial was transferred to an ice bath and the material macerated with a small pestle. The supernatant obtained after centrifugation (13,000×g, 1 min) was filtered (0.45 μm low-binding Durapore membrane) and diluted 1:5 in water prior to LC–MS analysis.

Analytical LC–MS was carried out using an Agilent 1100 Series LC (Agilent Technologies, Germany) coupled to an HCT Ultra ion trap mass spectrometer running in positive electrospray ionization (ESI) ultra-scan mode (Bruker, Bremen, Germany). The LC was fitted with a Zorbax SB-C18 column (2.1×50 mm, 1.8 μm; Agilent Technologies) and operated at 35 °C, with a flow rate of 0.2 mL min−1. The mobile phases were: (a) 0.1% HCOOH (v/v) and 50 μM NaCl; and (b) 0.1% HCOOH in MeCN (v/v). The gradient program was: 0–0.5 min, isocratic phase 2% B; 0.5–7.5 min, linear gradient 2–40% B; 7.5–8.5 min, linear gradient 40–90% B; 8.5–11.5 min, isocratic phase 90% B; 11.5–18 min, isocratic phase 2% B. The flow rate was raised to 0.3 mL min−1 in the interval 11.2–13.5 min. Traces of total ion current and of extracted ion currents for specific [M+Na]+ and [M+H]+ adduct ions were used to identify the eluted constituents using Compass DataAnalysis software (version 4.2, Bruker Daltonics). See below for targeted analytical LC–MS analysis of N-hydroxyphenylalanine from Paq tissue.

RNA isolation and transcriptome mining. Total RNA was prepared from 30 to 50 mg of plant tissue using the Spectrum Plant Total RNA Kit (Sigma-Aldrich, US). Transcriptomes were prepared from mRNA isolated from Paq frond tips containing high and low cyanogenic glycoside levels. Transcriptome sequencing was carried out by Macrogen (Seoul, South Korea) using an Illumina HiSeq2000 sequencer (Illumina, San Diego, CA) to generate paired-end libraries. The reads were de novo assembled and relative transcript abundance estimated by Sequester BioSeq (www.sequentiobiotech.com) using the Trinity pipeline.

The expression level of all transcripts was quantified. Identification of gene families was performed using the OrthoMCL pipeline (http://orthomcl.org/orthomcl/about.do). Differential gene expression analyses based on the Pa and Paq transcriptomes were carried out usingExpress and the data transferred into R and analyzed with the package NOISeq.

Isolation and transient expression of FMO candidate genes. The full-length sequence of the predicted ORF PaqF13802 encoding gene (most upstream methionine) without codon optimization was fitted with attB1 and attB2 Gateway cloning sites: attB1: gggccagctgtacacaaagacgct, attB2: gggccagcttgtacacaaagacgct. The full length of the FMO sequence from Pa cdNA using gene specific primers flanked by gateway sites (Forward: gggccagctgtacacaaagacgctgctatatctatgctatatgatgtctatgcatgtctatggtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatgcatgtctatg
kanamycin and chloramphenicol) with 20 mL of bacterial culture into four 2.5 L. Ultra Yield flasks fitted with AutOtop enhanced seals (Thompson Instrument, Germantown, Wisconsin). Preculture flasks were incubated at 37 °C and 225 rpm to reach OD 1.2–1.5 before induction with IPTG (final concentration 0.5 mM). The expression culture was grown O/N at 18 °C and the E. coli cells sedimented (5000 × g, 10 min) and stored at –20 °C until used.

E. coli cultures were thawed and resuspended in lysis buffer (1 g cells per 5 mL buffer; 100 mM HEPES pH 7.5, 500 mM NaCl, 1 mM MgSO4, 0.5 mM TCEP, Benzonase (25 μL/mL; Sigma-Aldrich)) and lysed in a high-pressure homogenizer (Avestin Emulsiflex D20, 40 psi). The lysate was clarified by centrifugation (11,000 rpm, 40 min), filtered (0.22 μm) and applied to two 5-μL His-Trap FF columns (GE Healthcare) connected in line and pre-equilibrated with binding buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM Imidazole, 0.5 mM TCEP). Columns were washed with ten column volumes of wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 30 mM Imidazole, 0.5 mM TCEP) and protein eluted using a ten column volume 0–100% gradient of elution Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 500 mM Imidazole, 0.5 mM TCEP). The eluate was collected in 1.5 mL fractions in 96-wells plates. Fractions of interest were analyzed by SDS-PAGE using 4–12% NuPage gradient protein gels (Thermo Scientific). Fractions containing the target protein were concentrated to a final total volume of 5 mL by centrifugation using pre-equilibrated membrane filters (30 kDa cut off; Thermo scientific) and applied to a HiLoad 16/60 Superdex 200 column (120 mL, GE Healthcare). Before usage, the isolated protein was frozen in liquid nitrogen and stored at −80 °C.

Recombinant enzyme assays. The activity of the isolated recombinant FOS1 protein was determined in vitro in assay mixtures (total volume: 50 μL) containing buffer; 100 mM HEPES pH 7.5, 500 mM NaCl, 1 mM MgSO4, 0.5 mM TCEP, and stored at 40 °C. The mobile phases were: (a) HCOOH (0.05%, v/v); and (b) 3.0 mm, 1.8 μm, Agilent Technologies, Germany) with a column temperature maintained at 30 °C, 300 rpm), the reaction was stopped by addition of following incubation (1 h, 30 °C, 300 rpm), the reaction was stopped by addition of Avestin EmulsiFlex D20, 40 psi). The lysate was clarified by centrifugation (11,000 rpm, 40 min), filtered (0.22 μm) and applied to two 5-μL His-Trap FF columns (GE Healthcare) connected in line and pre-equilibrated with binding buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM Imidazole, 0.5 mM TCEP). Columns were washed with ten column volumes of wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 30 mM Imidazole, 0.5 mM TCEP) and protein eluted using a ten column volume 0–100% gradient of elution Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 500 mM Imidazole, 0.5 mM TCEP). The eluate was collected in 1.5 mL fractions in 96-wells plates. Fractions of interest were analyzed by SDS-PAGE using 4–12% NuPage gradient protein gels (Thermo Scientific). Fractions containing the target protein were concentrated to a final total volume of 5 mL by centrifugation using pre-equilibrated membrane filters (30 kDa cut off; Thermo scientific) and applied to a HiLoad 16/60 Superdex 200 column (120 mL, GE Healthcare). Before usage, the isolated protein was frozen in liquid nitrogen and stored at −80 °C.

Recombinant enzyme assays. The activity of the isolated recombinant FOS1 protein was determined in vitro in assay mixtures (total volume: 50 μL) containing 30 μL of diluted FOS1 protein (0.345 mg mL−1) and protein was determined in vitro in assay mixtures (total volume: 50 μL) containing Recombinant enzyme assays and stored at −80 °C.

Analysis of recombinant PpaFOs1 and Pa tissue. Enzyme reaction mixtures (50 μL aliquots) were diluted with 50 μL Milli-Q grade water and filtered (Durapore 0.22 μm PVDF filter plates, Merck Millipore, Tullagreen, Ireland) together with filtered and diluted MeOH extracts of Pa fiddlehead and young pinna (see above). Sample pH was adjusted to 6.5 with NaOH. The mixture was pre-equilibrated with 4°C. Before usage, the isolated protein was frozen in liquid nitrogen and stored at −80 °C.

Recombinant enzyme assays. The activity of the isolated recombinant FOS1 protein was determined in vitro in assay mixtures (total volume: 50 μL) containing 30 μL of diluted FOS1 protein (0.345 mg mL−1) and protein was determined in vitro in assay mixtures (total volume: 50 μL) containing Recombinant enzyme assays and stored at −80 °C.

Analysis of recombinant PpaFOs1 and Pa tissue. Enzyme reaction mixtures (50 μL aliquots) were diluted with 50 μL Milli-Q grade water and filtered (Durapore 0.22 μm PVDF filter plates, Merck Millipore, Tullagreen, Ireland) together with filtered and diluted MeOH extracts of Pa fiddlehead and young pinna (see above). Sample pH was adjusted to 6.5 with NaOH. The mixture was pre-equilibrated with 4°C. Before usage, the isolated protein was frozen in liquid nitrogen and stored at −80 °C.

Recombinant enzyme assays. The activity of the isolated recombinant FOS1 protein was determined in vitro in assay mixtures (total volume: 50 μL) containing 30 μL of diluted FOS1 protein (0.345 mg mL−1) and protein was determined in vitro in assay mixtures (total volume: 50 μL) containing Recombinant enzyme assays and stored at −80 °C.

References
1. Harper, N. L., Cooper-Driver, G. A. & Swain, T. A survey for cyanogenesis in ferns and gymnosperms. Phys. Plant. 15, 1764–1767 (1976).
2. Gedrew, R. M. & Boler, B. L. Functional diversification in ferns and gymnosperms. Phys. Plant. 15, 1764–1767 (1976).
3. Luck, K. et al. CYP79 P450 monoxygenases in gymnosperms: CYP79A118 is associated with the formation of taxiphylin in Taxus baccata. Plant Mol. Biol. 95, 169–180 (2017).
4. Thodberg, S. et al. Elucidation of the amygdalin pathway reveals the metabolic basis of bitter and sweet almonds (Praunus dulcis). Plant Physiol. 178, 1096–1111 (2018).
5. Sanchez-Perez, R. et al. Mutation of a BHLH transcription factor allowed almond domestication. Science 364, 1095–1098 (2019).
6. Møller, B. L. Functional diversifications of cyanogenic glucosides. Curr. Opin. Plant Biol. 13, 338–347 (2010).
7. Picmanova, M. et al. A recycling pathway for cyanogenic glucosides evidenced by the comparative metabolic profiling in three cyanogenic plant species. Biochem J. 469, 375–389 (2015).
64. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512 (2013).

65. Roberts, A. & Pachter, L. Streaming fragment assignment for real-time analysis of sequencing experiments. Nat. Methods 10, 71 (2012).

66. Tarazona, S. et al. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic Acids Res. 43, e140 (2015).

67. Sainsbury, F., Thueenersmann, E. C. & Lomonossoff, G. P. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. Plant Biotechnol. J. 7, 682–693 (2009).

68. Voinnet et al. Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants. Proceedings of the National Academy of Sciences Nov 1999, 96, 14147–14152. https://doi.org/10.1073/pnas.96.24.14147.

69. Voinnet et al. Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants. Proceedings of the National Academy of Sciences Aug 2015, 112, E4812. https://doi.org/10.1073/pnas.1513950112.

70. Yoshimoto, N. et al. Identification of a flavin-containing monoxygenase involved in allidin biosynthesis in garlic. Plant J. 83, 941–951 (2015).

71. Ferreira-Silva, B., Lavandera, I., Kern, A., Faber, K. & Kroutil, W. Chemopromiscuity of alcohol dehydrogenases: reduction of phenylacetaldoxime to the alcohol. Tetrahedron 66, 3410–3414 (2010).

72. Möller, B. L., Olsen, C. E. & Motawia, M. S. General and stereocontrolled approach to the chemical synthesis of naturally occurring cyanogenic glycosides. J. Nat. Prod. 79, 1198–1202 (2016).

73. Oppolzer, W., Tamura, O. & Deereberg, J. Asymmetric synthesis of α-amino acids and α-N-hydroxyamino acids from N-acetylornithine-10,2-sultams-1-chloro-1-nitrosocyclohexane as a practical [NH₂]+ equivalent. Helv. Chim. Acta 75, 1965–1978 (1992).

74. Koch, B., Sillesen, O., Halkier, B. A., Svendsen, I. & Møller, B. L. The primary sequence of cytochrome P450tyr, the multifunctional N-hydroxylase catalyzing the conversion of L-tyrosine to p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dchurrin in Sorghum bicolor (L.) Moench. Arch. Biochem. Biophys. 323, 177–186 (1995).

75. Bak, S., Kahn, R. A., Nielsen, H. L., Møller, R. L. & Halkier, B. A. Cloning of three A-type cytochromes P450, CYP71E1, CYP98, and CYP99 from Sorghum bicolor (L.) Moench by a PCR approach and identification. J. Biol. Chem. 275, 19790–19796 (2000).

76. Jones, P. R., Møller, B. L. & Hoj, P. B. The UDP-glucose-hydroxymandelonitrile-O-glucosyltransferase that catalyzes the last step in synthesis of the cyanogenic glucoside dhurrin in Sorghum bicolor. Isolation, cloning, heterologous expression, and substrate specificity. J. Biol. Chem. 274, 35483–35491 (1999).

77. Andersen, M. D., Busk, P. K., Svendsen, I. & Møller, B. L. Cytochromes P-450 from cassava (Manihot esculenta Crantz) catalyzing the first steps in the biosynthesis of the cyanogenic glucosides linamarin and lotaustralin. Cloning, functional expression in Pichia pastoris, and substrate specificity of the isolated recombinant enzymes. J. Biol. Chem. 275, 1966–1975 (2000).

78. Jørgensen, K. et al. Biosynthesis of the cyanogenic glucosides linamarin and lotaustralin in cassava: isolation, biochemical characterization, and expression pattern of CYP71E7, the oxime-metabolizing cytochrome P450 enzyme. Plant Physiol. 155, 282–292 (2011).

79. Kannangara, R. et al. Characterization and expression profile of two UDP-glucosyltransferases, UGT85K4 and UGT85K5, catalyzing the last step in cyanogenic glucoside biosynthesis in cassava. Plant J. 68, 287–301 (2011).

Acknowledgements
The authors gratefully acknowledge the Protein Production and Characterization Platform at the Novo Nordisk Foundation Center for Protein Research for offering instruments and protocols for expression and purification. Gardeners Sue Dix and Jimmy Oluf Olsen, The Botanical Garden, University of Copenhagen, are thanked for taking good care of Ps. We acknowledge Søren Bak, Fernando Geu-Hores, and the Theme Group students over the years for their shared enthusiasm in fenns metabolic evolution. Laboratory technicians Susanne Bistrup, Lene Dalsten, and Theme Group students in 2018 are thanked for assistance in RNA extraction, cloning, and sequencing. Ricardo Aiione Cigliano from Sequenica Biotech is thanked for assistance with transcriptomic assembly and bioinformatics analyses. This work was supported by the VILLUM Center for Plant Plasticity (VKR023054) (B.L.M.); the European Research Council Advanced Grant (ERC-2012-ADG_20130414) (B. L.M.); VILLUM Young Investigator Grant (VKR01367) (E.H.J.N.); a Danish Independent Research Council Sapere Aude Research Talent Post-Doctoral Stipend (6111-003798) (E.H.J.N.); and a Novo Nordisk Emerging Investigator Grant (Grant No. 0054890) (E.H.J.N.). The financial support is gratefully acknowledged.

Author contributions
S.T., B.L.M., D.R.N., and E.H.J.N. initiated the study; S.T. and A.K.B. collected plant material and performed metabolite analysis; S.T. prepared RNA, performed transcriptomic analysis, and identified candidate genes. S.T., M.S., and A.K.B. carried out transient expression experiments; S.T., M.S., and M.B. carried out protein isolation; S.T. and M.S. carried out enzyme assays; S.T., M.S., and C.C. carried out LC-MS analyses; M.S.M. conducted chemical synthesis; S.T. carried out phylogenetic analyses; D.R.N. performed data mining from transcriptomes and the short read archive; S.T., M.S., E.H.J.N., and B.L.M. wrote the manuscript with input from all other authors.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s42003-020-01224-5.

Correspondence
Correspondence and requests for materials should be addressed to E.H.J.N.

Reprints and permission information
Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2020, corrected publication 2025