

Ancestral sequence reconstruction of the CYP711 family reveals functional divergence in strigolactone biosynthetic enzymes associated with gene duplication events in monocot grasses

Marcos H. Vinde^{1,2,3} , Da Cao⁴ , Rebecca J. Chesterfield^{1,3} , Kaori Yoneyama^{5,6} , Yosephine Gumulya² ,
Raine E. S. Thomson² , Tebogo Matila² , Birgitta E. Ebert¹ , Christine A. Beveridge⁴ ,
Claudia E. Vickers^{6,7,8}  and Elizabeth M. J. Gillam² 

¹Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, St Lucia, Qld 4072, Australia; ²School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Qld 4072, Australia; ³CSIRO Synthetic Biology Future Science Platform, CSIRO Land & Water, EcoSciences Precinct, Dutton Park, Brisbane, Qld 4012, Australia;

⁴School of Biological Sciences, ARC Centre of Excellence for Plant Success in Nature and Agriculture, The University of Queensland, St Lucia, Qld 4072, Australia; ⁵Graduate School of Agriculture, Ehime University, Ehime 790-8566, Japan; ⁶Japan Science and Technology Agency, PRESTO, Saitama 332-0012, Japan; ⁷ARC Centre of Excellence in Synthetic Biology, Queensland University of Technology, Brisbane, Qld 4000, Australia; ⁸Griffith Institute for Drug Design, Griffith University, Nathan, Brisbane, Qld 4111, Australia

Summary

Authors for correspondence:

Elizabeth M. J. Gillam

Email: e.gillam@uq.edu.au

Claudia E. Vickers

Email: claudia.vickers@qut.edu.au

Received: 17 November 2021

Accepted: 9 May 2022

New Phytologist (2022) **235**: 1900–1912

doi: 10.1111/nph.18285

Key words: ancestral sequence reconstruction, CYP711, cytochrome P450, GR24, phytohormone, plant evolution, strigolactone, terrestrialization.

- The strigolactone (SL) class of phytohormones shows broad chemical diversity, the functional importance of which remains to be fully elucidated, along with the enzymes responsible for the diversification of the SL structure. Here we explore the functional evolution of the highly conserved CYP711A P450 family, members of which catalyze several key monooxygenation reactions in the strigolactone pathway.
- Ancestral sequence reconstruction was utilized to infer ancestral CYP711A sequences based on a comprehensive set of extant CYP711 sequences. Eleven ancestral enzymes, corresponding to key points in the CYP711A phylogenetic tree, were resurrected and their activity was characterized towards the native substrate carlactone and the pure enantiomers of the synthetic strigolactone analogue, GR24.
- The ancestral and extant CYP711As tested accepted GR24 as a substrate and catalyzed several diversifying oxidation reactions on the structure. Evidence was obtained for functional divergence in the CYP711A family. The monocot group 3 ancestor, arising from gene duplication events within monocot grasses, showed both increased catalytic activity towards GR24 and high stereoselectivity towards the GR24 isomer resembling strigol-type SLs.
- These results are consistent with a role for CYP711As in strigolactone diversification in early land plants, which may have extended to the diversification of strigol-type SLs.

Introduction

Strigolactones (SLs) are a chemically diverse group of terpene phytohormones involved in developing complex plant architecture (Butler, 1995; Brewer *et al.*, 2013). They also function as soil signalling compounds that initiate symbiotic mycorrhizal colonization and stimulate the germination of parasitic weeds (Cook *et al.*, 1966; Xie *et al.*, 2010; Cardoso *et al.*, 2011; Lumba *et al.*, 2017). The broad spectrum of SL bioactivities indicates their great potential as novel multipurpose agrochemicals (Chesterfield *et al.*, 2020a). Strigolactone synthesis can be traced back to the earliest land plants (Walker *et al.*, 2019), and SL receptors even further back to algae, where trace amounts of the SL sorgolactone also have been found in *Charales* (Delaux *et al.*, 2012). In early land plants, SLs probably assisted the development of crucial root structures necessary for plants to conquer

land (Rensing, 2018). It also is believed that terrestrial fungi colonized land before plants and that the associations between fungi and plants, which are initiated by SLs, may have assisted plants in conquering land (Brundrett, 2002).

However, the need for SLs to be exuded into the rhizosphere to execute their signalling function with mycorrhiza also provided an opportunity for exploitation. Strigolactones have been hijacked as a germination signal for parasitic weeds such as *Orobanche* and *Striga*. Use of SLs by these parasites allows their seeds to lie dormant in the soil until a host plant is within sufficiently close proximity for infection by the germinating seedling (Cardoso *et al.*, 2011). It is thought that these conflicting roles of SLs as soil signalling molecules led to an evolutionary arms race that resulted in the broad chemical diversity observed for SLs today.

Chemical diversity in SLs is achieved through modification and decoration of carlactone (CL), the universal precursor of SLs

(Fig. 1), via enzymes including cytochrome P450 monooxygenases (P450s) from the CYP711, CYP722 and CYP728 families, methyltransferases and 2-oxoglutarate-dependent dioxygenases (Abe *et al.*, 2014; Zhang *et al.*, 2014; Brewer *et al.*, 2016; Yoneyama *et al.*, 2018a; Wakabayashi *et al.*, 2019; Wakabayashi *et al.*, 2021). Carlactone is derived from all-*trans*- β -carotene by the action of the isomerase DWARF27 (D27) and the carotenoid cleavage dioxygenases CCD7 and CCD8 (Jia *et al.*, 2018) (Fig. 1). Although the pathway up to CL is well-described, information about the enzymes involved in the diversification of CL to generate active SLs remains sparse, especially considering how many different SLs have been discovered in nature. It is known that CL is converted into the simplest bioactive SL, carlactonoic acid (CLA) which shows some but not all characteristic SL activities (Mori *et al.*, 2016; Yoneyama *et al.*, 2018b) by the highly conserved P450 subfamily CYP711A (including AtCYP711A1 known as MAX1 in *Arabidopsis thaliana*) (Abe *et al.*, 2014). Thereafter, SL synthesis branches into three separate core structural groups. The canonical SLs, comprising two groups, are identified by a tricyclic lactone-ring structure (ABC ring) (Fig. 1). Canonical SLs are further classified into either strigol-type or orobanchol-type based on the stereochemistry of the C ring (Xie *et al.*, 2013). Finally, noncanonical SLs exist, which lack the ABC ring formation but still retain SL bioactivity (Yoneyama *et al.*, 2018b). These three core structures are further diversified through different decorations, particularly of the A- and B-rings, where additions of methyl, hydroxyl, acetyloxy, epoxy and oxy groups are observed.

Although more than 30 SLs have been identified in nature (Chesterfield *et al.*, 2020a), knowledge about their individual roles in the plant – namely, their structure–function relationships – is still very limited. Many plants produce a variety of SLs, whereas others produce only a few. Moreover, some SLs seem to be exclusive to only a few plant species, yet others are widely distributed (Wang & Bouwmeester, 2018). Observations such as this suggest that there is probably a relationship between the chemical diversity of SLs in each species (the species' 'SL fingerprint') and its evolutionary history, particularly with respect to parasite interactions.

There is still much to be discovered about the enzymes catalyzing the diversifying reactions on the SL structures, yet P450s of three different subfamilies are known to play crucial roles. The CYP711A subfamily is the only member of the CYP711 clan of plant P450s conserved back to primitive land plants such as liverworts and mosses and is the earliest-occurring SL-associated P450 family (Nelson & Werck-Reichhart, 2011; Challis *et al.*, 2013; Hansen *et al.*, 2021). All active CYP711As identified so far catalyze the sequential oxidations of CL into CLA (Yoneyama *et al.*, 2018a). If this is the only activity observed, the CYP711A is classified as an A1-type CYP711A. Those that also oxidize the C18 of CLA to initiate the B-C ring closure of 4-deoxyorobanchol (4DO; Fig. 1) in addition to CLA synthesis, such as the rice CYP711A Os900, are classified as A2-type CYP711As. Those that catalyze the hydroxylation of the B-ring of 4DO, generating the most prevalent canonical SL, orobanchol (Fig. 1), are classified as A3-type (Yoneyama *et al.*, 2018a). No

CYP711As have been identified to date that can catalyze all three reactions.

The other two P450 subfamilies involved in SL biosynthesis, CYP722C and CYP728B, belong to the CYP85 clan of plant P450s and are less ancient than the CYP711A subfamily, being first observed in gymnosperms (Hansen *et al.*, 2021). However, the specific subfamilies, CYP722C and CYP728B, which are involved in SL synthesis, are not conserved throughout the land plant evolutionary tree; for example, the CYP722C subfamily is absent in monocot grasses and *Arabidopsis* (Wakabayashi *et al.*, 2021). The characterization of a variety of CYP722Cs heterologously expressed in a bacterium–yeast consortium (Wu *et al.*, 2021) showed that they possess one of two distinct activities (Fig. 1): direct conversion of CLA into orobanchol (without evidence of 4DO as an intermediate) and CLA conversion to 5-deoxystrigol (5DS) (Wakabayashi *et al.*, 2019; Wakabayashi *et al.*, 2020). In monocot grasses, strigol-type synthesis also occurs, despite the absence of the CYP722C subfamily P450s (Wakabayashi *et al.*, 2021). Instead, a sulfotransferase (LGS1) has been shown to catalyze 4DO and 5DS synthesis in sorghum (Fig. 1) (Wu & Li, 2021; Yoda *et al.*, 2021). Interestingly, within the monocot grasses, gene duplication events of the CYP711A subfamily have delivered multiple subfamily members (as opposed to the one or two CYP711As usually found in other species), including members with A3-type activities that contribute to orobanchol synthesis (Yoneyama *et al.*, 2018a). The third subfamily, CYP728B, is not exclusively involved in SL biosynthesis (Tu *et al.*, 2020), but CYP728B genes are upregulated in cowpea (*Vigna unguiculata*) roots under SL-inducing conditions such as phosphate deficiency, and a CYP728B has been reported to catalyze the hydroxylation of 5DS, yielding sorgomol in *Sorghum* (Wakabayashi *et al.*, 2021) (Fig. 1).

In this study, we focussed on the critical CYP711A subfamily, seeking to examine its function across evolutionary time. In particular, we were interested in elucidating the probable activity of the ancestral CYP711As from which the extant types arose. These enzymes are particularly interesting owing to their ancient evolutionary position and proposed role in assisting plants in colonizing terrestrial environments. Additionally, we wanted to investigate the evolutionary pathways of the CYP711A family to identify their involvement in SL diversification as a reaction to the evolutionary arms-race between plants and SL-responsive parasites, centring around the SL activities in rhizosphere signalling for both beneficial and detrimental functions. To address these questions, we used ancestral sequence reconstruction (ASR) (Foley *et al.*, 2020) to infer the ancient forms of CYP711A, resurrect them in the laboratory, and investigate their evolved activity while exploring the functional changes to the family along the major evolutionary lineages of land plants.

In addition, ancestral enzymes have been shown to display properties advantageous to application in the industrial production of fine chemicals, such as thermostability and expanded substrate specificity compared to extant enzymes. Therefore, we hypothesized that the resurrection and characterization of ancestors also may reveal opportunities for synthesizing natural and synthetic SLs in quantities suitable for expanded scientific

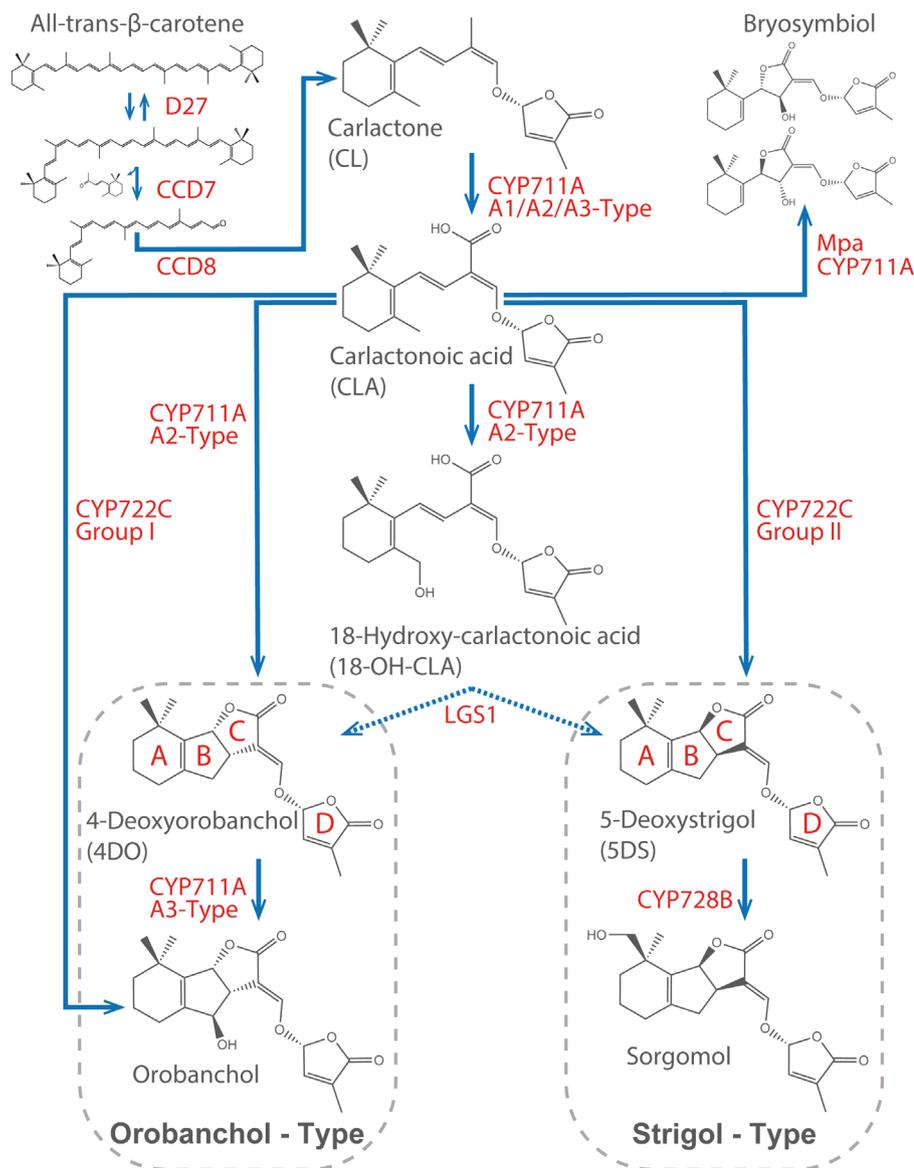


Fig. 1 Strigolactone biosynthetic pathway. Biosynthesis of diverse canonical strigolactones (SLs) starting from all-*trans*-β-carotene being converted to carlactone (CL) by isomerase DWARF27 (D27), and carotenoid cleavage dioxygenases CCD7, CCD8. Converting CL to carlactonoic acid (CLA) is a conserved function of CYP711As. In addition, A2-type CYP711As convert CLA to 18-hydroxycarlactonoic acid (18-OH-CLA) and 4-deoxyorobanchol (4DO), and A3-type CYP711As can convert 4DO to orobanchol. CYP722C group I can convert CLA to orobanchol and CYP722C group II can convert CLA to 5-deoxystrigol (5DS). A CYP728B from sorghum has been shown to convert 5DS to sorgomol. The sulfotransferase LGS1 can convert 18-OH-CLA to 4DO and 5DS. In addition to the canonical SLs with the tricyclic lactone-ring structure (ABC ring), a variety of noncanonical bioactive SLs is found in nature, such as bryosymbiol synthesized by CYP711As in *Marchantia paleacea* (Mpa). Solid arrows indicate where a single pathway is seen; dotted arrows indicate two alternative pathways catalyzed by the same enzyme.

investigations, potentially leading to development of SL biofactories for agricultural use. The latter may be particularly useful for SLs that are difficult to synthesize chemically.

Materials and Methods

Ancestral sequence reconstruction

Extant CYP711A sequences were collected from the databases NCBI, UniProt and the 1KP project (One Thousand Plant Transcriptomes Initiative 2019) using the BASIC LOCAL ALIGNMENT SEARCH TOOL (BLAST) (Altschul *et al.*, 1990) with CYP711A sequences representing various clades in the phylogenetic plant tree (*Arabidopsis thaliana*, BAH19609.1; *Medicago truncatula*, ABC59098.1; *Oryza sativa*, AGI65365.1 - BAS77709.1 - AGI65362.1; *Picea glauca*, AGI65359.1; *Selaginella moellendorffii*, AGI65366.1; *Marchantia paleacea*, KAG6552997.1). All extant CYP711 homologues with $\geq 40\%$ sequence identity were

collected, resulting in 12 121 sequences. Duplicate sequences with $> 99.9\%$ sequence identity were removed using the online tool CD-HIT (<http://weizhong-lab.ucsd.edu/cd-hit/>). Obsolete sequences and sequences containing illegal characters were identified using the online tool SEQSCRUB (<http://www.gabefoley.com/seqscrub/>) and removed before the sequence collection was aligned using the MULTIPLE ALIGNMENT USING FAST FOURIER TRANSFORM (MAFFT) tool with a gap open penalty of 3.0 (<https://www.ebi.ac.uk/Tools/msa/mafft/>). This sequence collection was pruned down to 346 sequences, realigning by MAFFT after each major stage, through the following steps: removing sequences of < 400 amino acids (aa) (average CYP711A length was 529 aa); and removing sequences with deletions in characteristic conserved P450 regions including the proline cluster associated with the microsomal membrane anchor of eukaryotic cytochrome P450 monooxygenases (P450s), the E-X-X-R motif in the K-helix which is important for stabilization of the core structure, and the F-X-X-G-X-R-X-C-X-G region containing the conserved cysteine crucial for heme binding

(Werck-Reichhart & Feyereisen, 2000). Additional irregular, anomalous sequences, consisting mainly of sequences containing apparent indels, were removed after they were identified as clear outliers in the multiple sequence alignment. This multiple sequence alignment was used to build the phylogenetic tree using the RANDOMISED ACCELERATED MAXIMUM LIKELIHOOD (RAxML) code on the Black Box interface provided on the CIPRES Science gateway. The tree was rooted in a group of CYP711A sequences from early land plants. The CYP711A ancestral sequences were inferred by a joint reconstruction using the GRAPHICAL REPRESENTATION OF ANCESTRAL SEQUENCE PREDICTIONS (GRASP) tool (<http://grasp.scmb.uq.edu.au>). The N-termini of the inferred ancestral CYP711A sequences were modified according to von Wachenfeldt *et al.* (1997) using the MAKKTSSKGKL sequence, and the aa sequences were reverse-translated and codon-optimized for expression in *Escherichia coli* using the ThermoFisher Scientific GENEART online tool (<https://www.thermofisher.com/au/en/home/life-science/cloning/c-misc/geneart.html>). The first 96 nucleotides of the ORF were modified to increase the predicted free energy of mRNA secondary structure formation (i.e. reduce the possibility that stable structures would form) using the mRNA optimizer tool (Gaspar *et al.*, 2013).

Homology modelling and prediction of protein attributes

Models of ancestral CYP711A structures were built with Swiss-Model (<https://swissmodel.expasy.org>) using the CYP3A5 structure PDB entry 5v.3.A as a template. This template provided the highest Global Model Quality Estimation of 0.65. Models were visualized using the PYMOL Molecular Graphics System (v.1.2r3pre; Schrödinger, New York, NY, USA). P450 substrate recognition sites were annotated according to Gotoh (1992) and Zawaira *et al.* (2011). Membrane embedded regions of ancestral CYP711As were predicted using the TMHMM-2.0 tool (<https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>). Sequence logos were generated using the WEBLOGO tool (<https://weblogo.berkeley.edu/logo.cgi>).

Plasmid construction

Complementary gBlocks™ gene fragments encoding the inferred ancestral CYP711As, containing a C-terminal 6xHis-tag, and *Arabidopsis thaliana* cytochrome P450 reductase 1 (ATR1) (Urban *et al.*, 1997) were synthesized by Integrated DNA Technologies (IDT, Coralville, IA, USA). Each ancestral CYP711A was assembled with the similarly codon-optimized ATR1 into the pCW expression vector (Parikh *et al.*, 1997) between *Nde*I and *Hind*III using the NEBuilder® HiFi DNA Assembly Cloning Kit (New England BioLabs, Ipswich, MA, USA).

CYP711A expression in *E. coli* and membrane preparation

Expression of the ancestral CYP711As was based on previously published methods (Gillam *et al.*, 1993; Iseki *et al.*, 2018) with changes as described below. Bicistronic pCW plasmids

containing the ancestral CYP711As and ATR1 were transformed into chemically competent DH5 α /IQ™ cells already hosting the pGro7 plasmid (Nishihara *et al.*, 1998) for co-expression of chaperones GroEL and GroES. Transformed cells were applied to lysogeny broth (LB) agar containing 100 μ g ml⁻¹ ampicillin, 20 μ g ml⁻¹ chloramphenicol and 0.2% w/v glucose and incubated overnight at 37°C. Single colonies were inoculated into 5-ml LB starter cultures with the same additives as above and grown overnight at 37°C with shaking at 250 rpm. The starter cultures were used to inoculate the bulk cultures at a 1 : 100 seeding volume ratio. The bulk culture consisted of Terrific Broth (TB) medium containing 100 μ g ml⁻¹ ampicillin, 20 μ g ml⁻¹ chloramphenicol and 1 mM thiamine, supplemented with trace elements (Gillam *et al.*, 1993). Bulk cultures were grown at 37°C with shaking at 250 rpm until an OD₆₀₀ of 0.5 was reached; thereafter expression was induced by adding 1 mM isopropyl β -D-1-thiogalactopyranoside, 0.5 mM δ -aminolevulinic acid and 4 mg ml⁻¹ arabinose (Notley *et al.*, 2002). Cultures were grown at 25°C and shaking at 150 rpm for 72 h. For the preparation of microbial membranes, cells were harvested by centrifugation at 5000 *g* for 10 min and resuspended in homogenization buffer (100 mM potassium phosphate buffer, 6 mM magnesium acetate, 20% glycerol, 0.1 mM dithiothreitol) with the addition of protease inhibitor cocktail (Sigma-Aldrich) and 1 mM phenylmethylsulfonyl fluoride. Cells were lysed using a One Shot high-pressure homogenizer (Constant Systems Ltd, Northants, UK) and lysates were centrifuged at 10 000 *g* for 20 min. The supernatant was collected and centrifuged for an additional 65 min at 180 000 *g* to isolate membrane fractions, which then were resuspended in TES buffer (50 mM Tris acetate, 250 mM sucrose, 0.25 mM EDTA, pH 7.6). The P450 content in the membrane preparations was determined from ferrous CO vs ferrous difference spectra as described by (Guengerich *et al.*, 2009), using an OLIS-Aminco™ DW-2A spectrophotometer.

Carlactone and GR24 assays and LC-MS/MS analysis

The membrane preparations containing the ancestral CYP711As and ATR1 were used for *in vitro* assays. Substrates used for these assays included crude extracts of a CL-producing strain of *E. coli* (the construction of the CL-producing *E. coli* strain is described in Supporting Information, Methods S1), as well as *rac*-GR24 and purified GR24 isomers (provided by Christopher S. P. McErlean, University of Sydney, Australia). Reaction mixtures (total volume 250 μ l) were prepared in 50 mM potassium phosphate (pH 7.4) and contained 30 μ M GR24 substrate and 0.3 μ M ancestral CYP711A in bacterial membranes. When the P450 content was too low to be quantified, 50 μ l of membrane preparation were used. Reactions were started by adding NADPH to 500 μ M and an NADPH-regenerating system consisting of 0.5 units ml⁻¹ glucose-6-phosphate dehydrogenase and 10 mM glucose-6-phosphate and allowed to proceed at 30°C for 1 h. Metabolites were extracted from the incubations by adding 250 μ l ethyl acetate, vortexing for 1 min and centrifuging at 17 000 *g* for 5 min. The organic layer was collected and evaporated under nitrogen (or reduced as much as possible for CL

assays, where substrate is unstable when desiccated completely). Extracts then were dissolved in 30 μ l acetonitrile and subjected to LC–MS/MS analysis.

The ultrahigh performance liquid chromatography (UPLC)–MS/MS system used for product screening was a Nexera X2 UPLC system (Shimadzu Corp., Kyoto, Japan) coupled with a 5500 triple quadrupole linear ion trap (QTRAP) mass spectrometry system equipped with an electrospray ionization source (ESI) (AB Sciex, Framingham, MA, USA). The extracts were separated with a Phenomenex Kinetex C18 reversed phase column (2.1 mm \times 100 mm, 1.7 μ m). The settings and gradients for UPLC were as follows: mobile phase A, 0.1% acetic acid in MilliQ water (v/v); mobile phase B, 0.1% acetic acid in acetonitrile (v/v). The flow rate was 0.5 ml min⁻¹. The programmed step gradient was: 30% B over 0.1 min; 30% to 100% B over 12 min; followed by a clean-up step, 100% to 100% B over 2.5 min, 100% to 30% B over 0.1 min and column wash for 0.5 min. The settings for MS were as follows: the ion source settings were followed as described in (Chesterfield *et al.*, 2020b). The precursor ion scan mode was used to search for putative compounds which can generate the product ion: 97 *m/z*. Scans were performed in positive mode; Q1 scan mass range: 50–350 *m/z*; collision energy: 35 V; dwell time: 35 ms. The accurate mass and fragmentation patterns for selected peaks were confirmed with a ZenoTOF 7600 system (AB Sciex).

Results

Ancestral sequence reconstruction of the CYP711A family

In this study, we used ASR to infer 11 different ancestors from the CYP711A phylogenetic tree (Fig. 2; Table 1). The tree was built from 346 sequences collected across the databases NCBI and Uniprot and the oneKP (1KP) project (<https://sites.google.com/a/ualberta.ca/onekp/>). The majority of the sequences came from angiosperms; 97 monocot sequences were represented in the tree, along with 227 eudicot sequences divided into 119 rosids and 74 asterids. Amborella, chloranthale, magnolid, basal eudicot, caryophyllale, santalale and saxifragale sequences also were represented within the angiosperm clade. In addition to the angiosperm sequences, three gymnosperm sequences were included in the tree, along with eight sequences from early land plants, including a hornwort, three liverworts, a lycophyte, two bryophytes and a fern. The final collection of extant CYP711A sequences broadly represented the phylogenetic tree of land plants and was used to infer ancestral nodes along the evolutionary lineages of the CYP711A subfamily.

In most species, only one or two CYP711A sequences were present, with only a few exemptions having multiple copies of CYP711A. In eudicots, multiple CYP711As (up to four copies) were present, mainly in members of the rosaceae family. The most distinct gene duplication events, however, occurred in monocots.

The monocot sequences clustered into two groups (Fig. 2). The monocot grass species showed particularly interesting

diversity in their CYP711A genes. At least three copies of CYP711A were found in each species, and often more. These genes belonged to three separate groups, named here as monocot G1, G2 and G3. Monocot G1 and G2 corresponded to higher-level monocot groups, whereas monocot G3 evolved from within the monocot G2 group and only consisted of monocot grasses (Fig. 2). Monocot G3 was of particular interest as it contains both A2-type CYP711As, such as Os900 from rice, and A3-type CYP711As, such as Os1400 from rice, in addition to the common A1-type CYP711As (see Fig. 2).

Resurrection and characterization of ancestors

The 11 ancestral nodes selected for reconstruction consisted of six key evolutionary branch points in the land plant evolutionary tree, including the last common ancestor (LCA), and the angiosperm, monocot, eudicot, rosid and asterid ancestors. Additionally, more recent ancestors of particular interest were resurrected, including: the ancestors of the aforementioned two monocot groups (monocot groups 1, 2) and the third group of the grass species (monocot group 3); the rosaceae ancestor, where multiple CYP711As were found in several species; and the solanale ancestor, where synthesis of both strigol type and orobanchol-type strigolactones (SLs) occurs (Xie *et al.*, 2013).

The inferred ancestral CYP711A protein sequences were reverse-translated into DNA sequences, which were synthesized and cloned into the pCW expression vector along with the NADPH-cytochrome P450 reductase from *A. thaliana*, ATR1, for co-expression of the enzymes in *E. coli*. Constructs prepared from the native CYP711A cDNA sequences demonstrated poor expression (data not shown). To improve expression, the sequences were codon-optimized for *E. coli* and the 5' end of the open reading frame (ORF) encoding the N-terminal region of the enzyme upstream of the proline-rich motif was replaced with the MAKKTSSKGKL sequence (von Wachenfeldt *et al.*, 1997), which was previously shown to increase the expression of the CYP711A MAX1 in *E. coli* (Gumulya & Gillam, 2017). Using this approach, most of the ancestral CYP711As were expressed at detectable levels, except for N28 (monocot group 1), N118 (eudicot ancestor) and N294 (solanale ancestor) (Fig. 2). The expression of the two rice CYP711As also was below the sensitivity limit of the spectral assay (Table 1). Variations in expression levels also were observed between the ancestors expressed at detectable levels, from the poorly expressed N0 and N71 to the highest expressed N25 (see Table 1).

These observations are consistent with previous attempts to express eukaryotic P450s in bacteria; N-terminal modifications, particularly truncation of the hydrophobic N-terminal and minimization of the potential for secondary structure formation in the mRNA appear to enhance expression in many cases. It is not clear why certain P450s are easily expressed whereas others are refractory to many different N-terminal modifications. However, the efficiency or fidelity of protein folding and/or heme insertion may be critical as apoprotein production is typically strong even in the absence of a hemoprotein spectrum (results not shown).

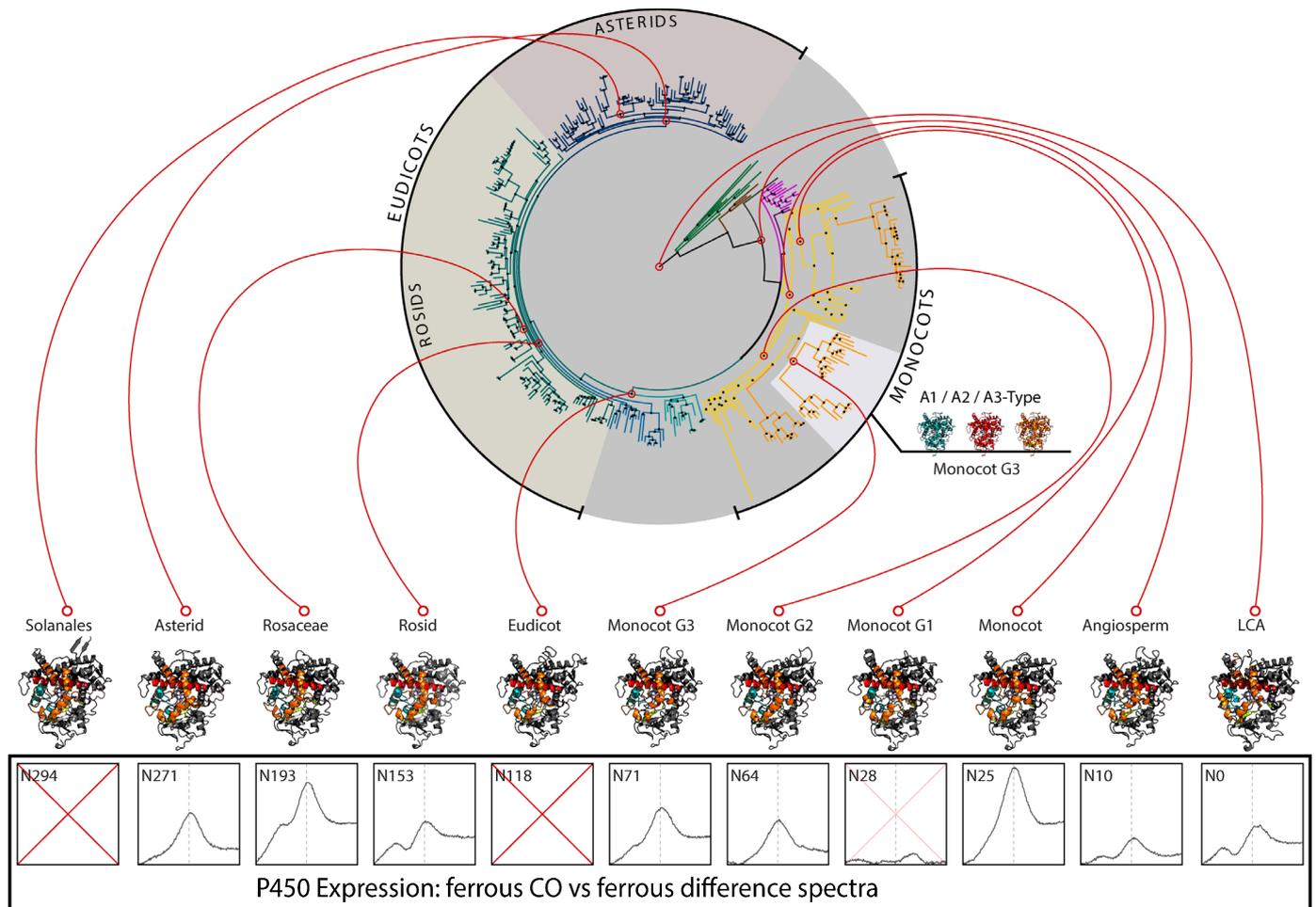


Fig. 2 Phylogenetic tree of the CYP711 family built using RANDOMISED AXELERATED MAXIMUM LIKELIHOOD (RAXML). CYP711As families are shown as different colours: green, early land plants; brown, gymnosperms; magenta, magnoliids; yellow, monocots; orange, monocot grasses; teal, core eudicots – rosids; blue, core eudicots – asterids; light teal, basal eudicots; light blue, caryophyllales, santalales and saxifragales. Monocot group three (G3), which contains CYP711As with A1, A2 and A3-type activity, is highlighted with teal, red, and orange CYP711A structures respectively. Beneath the phylogenetic tree, structural models of the 11 inferred ancestors are depicted, along with the ferrous CO vs ferrous difference spectra of purified membranes isolated from *Escherichia coli* expressing the respective ancestral enzymes, as an indication of the different expression levels observed for the ancestors in *E. coli*. LCA, last common ancestor.

Development of an *E. coli* system for production of CL

In order to test the activity of CYP711A ancestors on their native substrate CL, a CL-producing *E. coli* strain was developed (Methods S1; Table S1). Carlactone is not commercially available, is challenging to produce via chemical synthesis and has poor stability, making long-term storage challenging. For these reasons, a simple, on-demand production method for CL is valuable for SL research applications. To this end, an engineered carotenogenic *E. coli* was used as a background strain for the introduction of three enzymes that catalyze the conversion of β -carotene to the SL precursor, CL: *O. sativa* isomerase DWARF27 (D27) lacking its plastid transit peptide (OsD27 Δ TP); and carotenoid cleavage dioxygenases CCD7 (PsCCD7) and CCD8 (PsCCD8) from *Pisum sativum*.

The expression yield and solubility of a recombinant protein can often be enhanced by expressing it as a fusion with the ORF of a highly expressed domain (fusion partners), and

this may augment enzyme activity in whole-cell expression systems. Maltose binding protein (MBP) and thioredoxin (Trx) have been shown to improve the solubility of a wide range of protein partners (Costa *et al.*, 2014). *Escherichia coli* glycerol-conducting channel protein (GlpF) previously had been shown to enhance production through two other pathways originating from β -carotene, probably due to its ability to localize fusion partners to the *E. coli* inner membrane (Ye *et al.*, 2018). Therefore, to enhance CL production, the ORFs for the CL biosynthetic enzymes were fused to MBP, Trx or GlpF and assembled, expressed and screened combinatorically in carotenogenic *E. coli* along with ‘no fusion partner’ constructs. A strain comprising Trx fusions of OsD27 Δ TP and PsCCD7, and a maltose-binding protein fusion of PsCCD8, produced 20-fold higher titres of CL than strains harbouring the enzymes without fusion partners. A maximum measured CL titre of $112 \pm 22 \mu\text{g l}^{-1}$ (mean \pm SD) was obtained after 48 h (Figs 3, S1).

Table 1 Characteristics of the ancestral nodes selected for reconstruction, including the number of extant sequences in the alignment descending from the given ancestor and the expression yields of the ancestral CYP711As per litre of culture.

Ancestral node	Description	No. of sequences in reconstruction	Expression yield (nmol l ⁻¹ ; mean ± SD)
N0	Last Common Ancestor	346	8.6 ± 2.3
N10	Angiosperms	335	12.0 ± 0.6
N25	Monocots	94	45.3 ± 8.5
N28	Monocots Group 1	37	nd ^a
N64	Monocots Group 2	55	26.2 ± 1.6
N71	Monocots Group 3	26	7.4 ± 1.2
N118	Eudicots	227	nd
N153	Rosids	119	13.3 ± 1.7
N193	Rosaceae	19	26.7 ± 3.5
N271	Asterids	74	29.0 ± 2.5
N294	Solanales	16	nd
Os900	Extant rice form	1	nd
Os1400	Extant rice form	1	nd

^and, no P450 haemoprotein could be detected.

Conversion of CL to CLA by a subset of ancestral CYP711As

The conversion of CL to CLA is a well-conserved activity reported for all active CYP711As characterized so far. Although it has been speculated that CL was not the native substrate for the P450s ancestral to the CYP711 family (Wu *et al.*, 2021), it is the substrate for MpaCYP711A, which catalyzes the synthesis of bryosymbiol, an SL discovered in the early land plant *M. paleacea* (Kodama *et al.*, 2021) (Fig. 1).

The focus of this study was to explore the involvement of ancestral CYP711As in further diversification of the SL structure. We particularly wanted to identify where A2-type activity had evolved to convert CLA to 4-deoxyorobanchol (4DO). Using crude extracts of the CL-producing *E. coli* strain as substrate, we tested whether N25 (the monocot ancestor) and N71 (the monocot group 3 ancestor) converted CL to CLA and/or 4DO, as extant forms from these groups are known to possess A2-type activity. Both ancestors could convert CL to CLA as expected, but neither 18-OH-CLA or 4DO were identified as products (Fig. 4).

Hydroxylation of GR24 as a surrogate substrate for 4DO and 5-deoxystrigol

Many canonical SLs contain one or more oxygen atoms inserted into the core SL structure by P450 enzymes as part of their biosynthesis, such as the reactions catalyzed by the A3-type CYP711As and CYP722Cs that produce orobanchol, and the reactions catalyzed by CYP728B that produce sorgomol (Fig. 1). Given that the ancestors failed to show A2-type activity, the next step in investigating their role in the diversification of the SL structure was to assess their activity on the undecorated canonical SLs, 4DO and 5DS, to see if they could insert oxygen into any of these structures, as seen in orobanchol synthesis by A3-type CYP711As.

We wanted to investigate when A3-type activity had evolved in the CYP711A lineage and if the ancestral CYP711As were capable of inserting oxygen into any other sites on the core canonical SL structures 4DO and 5DS.

Unfortunately, no sufficiently pure 5DS and 4DO was available for use as substrates to test the ancestral CYP711As. Therefore, the SL analogue GR24 was investigated as a surrogate substrate in place of 4DO/5DS. GR24 is known to be hydroxylated when fed to sorghum (Ueno *et al.*, 2015), and although the enzyme involved was not identified, we hypothesized that analysis of GR24 metabolism could provide information on what reactions the ancestral CYP711As were capable of catalyzing, as well as an indication of the substrate promiscuity of the enzyme in regard to a non-native substrate.

Membranes isolated from *E. coli* expressing the ancestral CYP711As with ATR1 were incubated with *rac*-GR24, and the formation of GR24 + 16 *m/z* products was monitored by LC-MS/MS. The solanale (N294) and eudicot (N118) ancestors were omitted from this comparison because they failed to show detectable P450 expression. However, the extant CYP711As from rice (*O. sativa*), Os900 and Os1400, and the monocot group 1 ancestor (N28), which expressed poorly compared to the other ancestors but better than N294 and N118, were included in the experiment.

Surprisingly, multiple GR24 + 16 *m/z* product peaks were detected: a single peak with a retention time (RT) of 2.45 min and a complex peak consisting of at least three separate peaks with a RT of around 3.35 min (Fig. 5). The GR24 + 16 *m/z* peaks were confirmed as derivatives of GR24 based on exact mass MS2 fragmentation patterns (Fig. S2). All ancestors, as well as the two extant forms Os900 and Os1400, were capable of producing the [GR24 + 16 Da] RT 3.35 products, whereas the [GR24 + 16 Da] RT 2.45 product, which was of relatively low abundance for some ancestors, was detected at significant levels only in assays with the well-expressed ancestors (although a small nonquantifiable peak was present in the N28 product). Therefore, it cannot be excluded that the low-abundance ancestors can catalyze this reaction. No GR24 + 16 *m/z* products were detected in the samples from negative control reactions containing membranes from cells transformed with the empty pCW vector.

In both CL and GR24 assays, it was clear that the N71 (monocot G3) ancestor generated more product than several other ancestors, even though its expression yield was lower. To elucidate whether the N71 ancestor had evolved increased catalytic activity, the GR24 assays were repeated with an equal concentration of P450 (0.3 μM) in the assay to compare the relative activity between the ancestors. As the extant forms Os900, Os1400 and the N28 ancestor were all poorly expressed, it was not possible to use the same concentration of P450 in reactions with these forms; although they were still included in the assay, the relative activity of these enzymes cannot be compared to the rest. As earlier results indicated, the N71 ancestor was indeed more active than the other ancestors producing around three times more of the [GR24 + 16 Da] RT 3.35 product than the second most active ancestor N271 (Fig. 6). By contrast, the N193 and N271 ancestors were more active in producing the [GR24 + 16 Da] RT

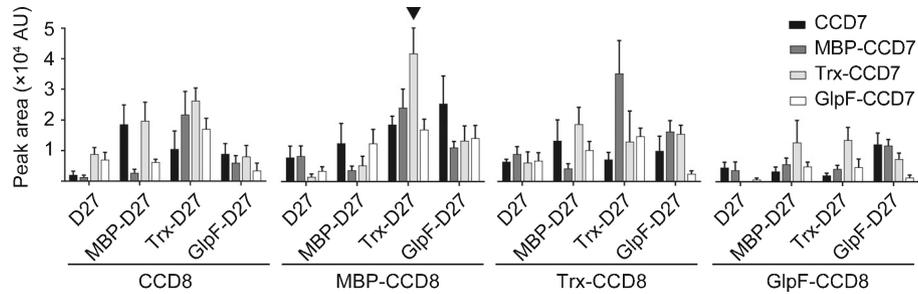


Fig. 3 Carlactone (CL) production in a combinatorial screen of maltose binding protein (MBP), thioredoxin (Trx) and glycerol-conducting channel protein (GlpF) fusion partners for OsD27ΔTP, PsCCD7 and PsCCD8 (D27, isomerase DWARF27; CCD, carotenoid cleavage dioxygenase). Peak area is given for the 301 → 97 *m/z* transition. Black arrow indicates the best-performing strain, denoted BL21-EcoFlex58. Data represent the mean ± SE of triplicate experiments.

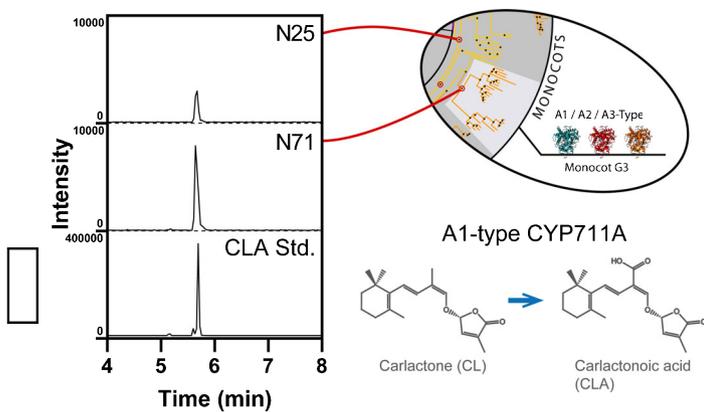


Fig. 4 Hydroxylation of carlactone. Mass spectrometry analysis showed formation of carlactonic acid (CLA) (331 → 113 *m/z* transition, 5.7 min retention time) by both the ancestral CYP711As, N25 and N71. Neither of these enzymes showed A2-type activity even though this activity is known to occur in extant CYP711As within these groups. CYP711As from monocots are shown in yellow, and those from monocot grasses are shown in orange. Red circles indicate the nodes chosen for resurrection.

2.45 product, accumulating around 2.5-fold more than the third most active ancestor, N71. This result suggests that these two eudicot ancestors may have evolved a preference for hydroxylation at a different site on the SL structure than the N71 ancestor, and the noticeably higher activity of the N71 ancestor suggests that a functional change could have occurred along this lineage of the CYP711A tree.

Substrate stereo-specificity of ancestral CYP711As

The [GR24 + 16 Da] RT 2.45 and [GR24 + 16 Da] RT 3.35 products were assumed to be formed from hydroxylation at two separate sites of the GR24 structure or different faces of the same site, yet the multiple peaks with retention times around 3.35 min could have been caused by hydroxylation at the same site of different GR24 enantiomers in the racemic mixture. To clarify whether these peaks were formed from different isomers or from hydroxylation at different sites or different faces on the same enantiomer, the assays were repeated with the highly active N71 ancestor using each of the purified isomers of GR24: (+)-GR24, mimicking the strigol-type SLs like 5DS; (–)-*ent-epi*-GR24,

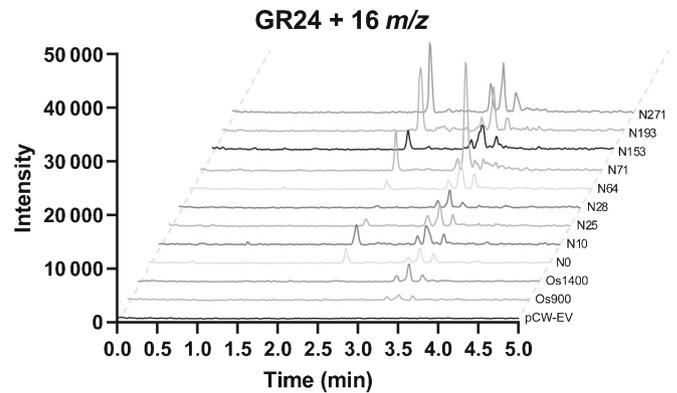


Fig. 5 Hydroxylation of racemic GR24. Multiple reaction monitoring (MRM) chromatograms showing formation of the GR24 + 16 *m/z* products, [GR24 + 16 Da] Retention time (RT) 2.45 (315 → 97 *m/z* transition, 2.45 min retention time) and [GR24 + 16 Da] RT 3.35 (315 → 97 *m/z* transition, 3.35 min retention time). Representative peaks are shown from three replicate samples. Purified membranes from *Escherichia coli* expressing the ancestral or extant CYP711As along with ATR1 or the empty pCW vector were used for these GR24 assays.

mimicking orobanchol-type SLs like 4DO; and the corresponding isomers with the same B/C-ring conformation but an unnatural conformation of the D-ring, (+)-*epi*-GR24 and (–)-*ent*-GR24, respectively. This experiment also may suggest whether the ancestral CYP711As could have participated in diversifying strigol-type SLs in addition to the known orobanchol synthesis by A3-type CYP711As. It was of interest to assess the N294 ancestor (solanales) because of the known production of both strigol- and orobanchol-type SLs in species within this order. However, no P450 holoenzyme expression could be detected for the N294 ancestor. Therefore, the older N271 (asterid) ancestor was tested against the different GR24 isomers instead, as the reconstruction of the ancestor of this clade included the solanale sequences as well.

The N71 ancestor again was more active than others towards the [GR24 + 16 Da] RT 3.35 products (Fig. 7). Even when using the purified (+)-GR24 isomer as substrate, two major peaks and a minor peak were detected from N71 with retention times around 3.35 min, showing that (+)-GR24 was indeed hydroxylated at several sites and/or alternative faces by the ancestors. Surprisingly, N71 only accepted the 5DS-like (+)-GR24 isomer,

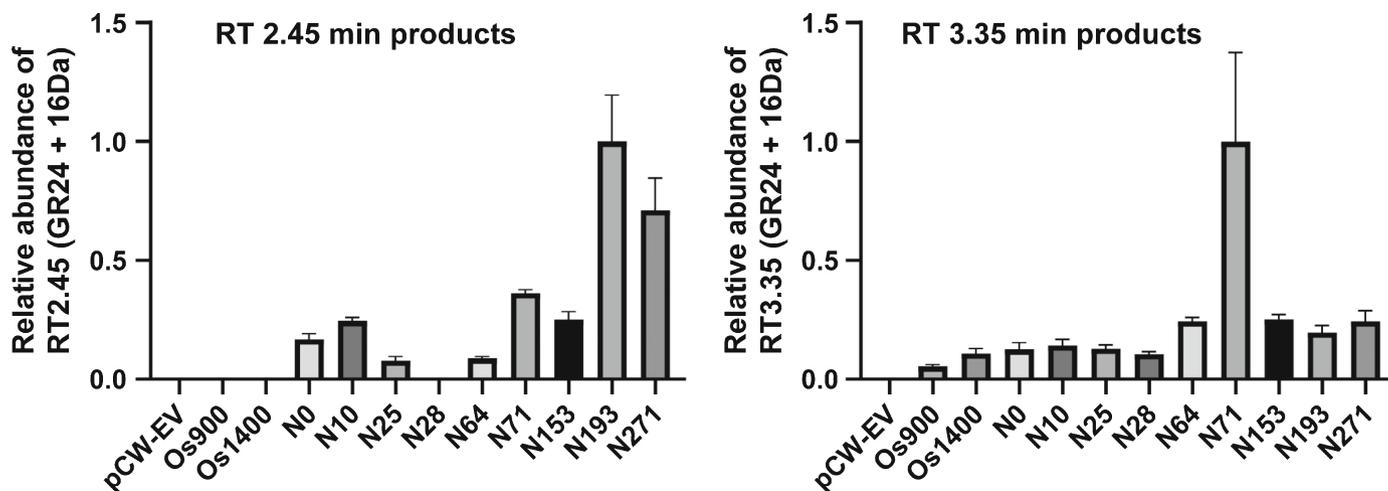


Fig. 6 Relative activity of ancestral CYP711As towards formation of (a) [GR24 + 16 Da] retention time (RT) 2.45 and (b) [GR24 + 16 Da] RT 3.35 products from the ancestral CYP711As is shown, as determined from multiple reaction monitoring (MRM) chromatogram peak areas (315 → 97 *m/z* transition). A consistent amount of 0.3 μM P450 was used for each assay except for the empty pCW vector negative control, Os900, Os1400 and N28. Data are the average \pm SD of three independent biological replicates and normalized to the highest accumulation of product observed.

suggesting that may have been involved in diversification of strigol-type SLs. This observation has not been reported for any other CYP711A so far.

In addition to the [GR24 + 16 Da] RT 3.35 products, N71 also produced the [GR24 + 16 Da] RT 2.45 product from the (+)-GR24 isomer (Fig. 7). The N271 ancestor also converted a small amount of the (+)-GR24 isomer but generally preferred the (–)-*ent-epi*-GR24 and (–)-*ent*-GR24 isomers resembling 4DO in the B/C-ring conformation. [GR24 + 16 Da] RT 2.45 was the major product of the N271 ancestor using *rac*-GR24 as substrate and also was the major product using the (–)-*ent*-GR24 substrate with the unnatural D-ring conformation. However, [GR24 + 16 Da] RT 2.45 was not a product of the (–)-*ent-epi*-GR24 substrate, showing that the unnatural conformation of the D-ring is crucial for N271 to catalyze this hydroxylation. Neither of the ancestors showed any activity towards the (+)-*epi*-GR24 isomer, which is known to be unable to activate SL signalling.

LC–MS/MS of the GR24 products at collision energies that retain the hydroxyl group failed to differentiate these products so full structural identification of these metabolites will require isolation of significant amounts of these metabolites for characterization by NMR.

Discussion

Here we show that resurrected ancestral CYP711As possess varied catalytic activities as well as divergent preferences towards the stereo- and regio-specific hydroxylation of the SL analogue GR24. By exploring the structural differences between the ancestral CYP711As, it is possible to speculate about the changes in the enzyme that caused this functional divergence. The gene duplication events resulting in separation of CYP711A groups within monocot grasses, were of particular interest given that the clade descending from the monocot G3 contained several A2- and A3-type CYP711As. Structural models were built for all of the ancestral CYP711As, and substrate recognition sites (SRS)

were annotated and examined for mutations that potentially could lead to the observed altered activities (Fig. S3). The N71 ancestor stood out as having a lower sequence identity in the SRSs than the other ancestors, except for N0, the most evolutionary distant ancestor (Fig. S4). The sequence logos of extant sequences within the monocot groups also showed greater sequence flexibility in the SRS of members in monocot group 3. For example, two otherwise highly conserved sites in SRS5 and SRS6 often were mutated in the monocot group 3 sequences, resulting in the inferred N71 ancestor sequence containing the unique mutations V377I and L489V (Fig. 8). Although V → I and L → V substitutions should not cause major disruption in the substrate pocket, these specific sites are closely associated with the heme prosthetic group, where even minor modifications could potentially change the characteristics of the enzyme (Fig. 8a). Likewise the E482Q and E484Q substitutions in N71 relative to all other sequences except N0 could be significant in the divergent properties of this ancestor. However residues at these positions are less likely to interact with the substrate.

Another area of interest outside the SRSs is the F-G loop between SRS2 and SRS3, which borders the substrate access channel into the P450 active site in close association with the membrane (Werck-Reichhart & Feyereisen, 2000). Mutations changing the hydrophobicity in this region are known to alter both enzyme efficiency and specificity (Zawaira *et al.*, 2011; Forman *et al.*, 2018), which have led to the suggestion that the original definition of SRSs should be expanded to include this region as part of SRS2 and SRS3 (Gotoh, 1992; Zawaira *et al.*, 2011). In the middle of this region, the N71 ancestor contains a L222Q mutation that should decrease hydrophobicity, reducing the predicted interaction between the F-G loop and the membrane (Fig. 8d).

These mutations are thus interesting candidates for the observed changes in the substrate regio- and stereospecificity and the increased catalytic activity of the N71 ancestor relative to the other ancestors. Examination of these and other mutations in the

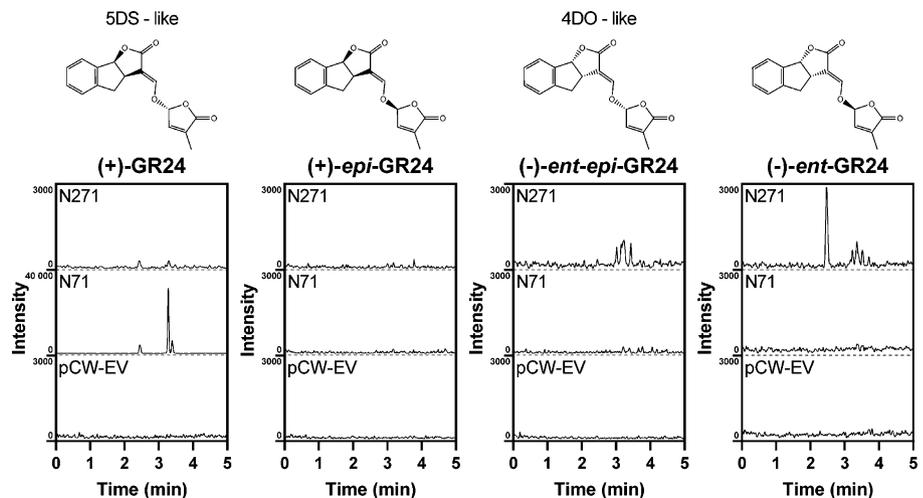


Fig. 7 Hydroxylation of GR24 isomers. Multiple reaction monitoring (MRM) chromatograms (315 → 97 m/z transition) showing the formation of [GR24 + 16 Da] retention time (RT) 2.45 and [GR24 + 16 Da] RT 3.35 products from the ancestral CYP711As using the four different GR24 isomers as substrate. Representative peaks are shown from three replicate samples. Purified membranes from *Escherichia coli* expressing the ancestors N71 or N271 along with ATR1 or the empty pCW vector were used for these GR24 assays.

future through mutational analysis will shed further light on this. Obtaining a substrate-bound structure for one or more of the ancestral enzymes will confirm whether candidate residues and regions hypothesized from necessarily imperfect homology models do in fact participate in interactions with the substrates.

Resurrecting and characterizing the inferred ancestral proteins allows a deeper understanding of how different traits and functions of an enzyme have evolved to adapt to changing environments over time. Stereoselective SL biosynthesis may have evolved as a consequence of the varying contributions of different SLs to parasitic weed germination. Here, sequence flexibility within SRSs in members of monocot group 3 has potentially led to increased activity towards the diversification of the canonical SL structure as well as towards substrates with the strigol-type conformation. It is known that the sulfotransferase LGS1 controls the formation of either orobanchol-type or strigol-type SLs (Fig. 1) in *Sorghum* which can lead to *Striga* resistance (Gobena *et al.*, 2017; Wu & Li, 2021; Yoda *et al.*, 2021). Results from this study suggest that the N71 ancestor can accept 5DS-like structures as substrates and this opens the possibility of CYP711A forms being involved in further diversification of strigol-type SLs downstream of 5DS. Previous studies have shown that CYP711As are involved in the conversion of the universal precursor, CL, to CLA, but also downstream hydroxylations to form the orobanchol-type canonical SLs: hydroxylation of C18 of CLA to initiate the B-C ring closure forming 4DO; and the hydroxylation of the B-ring of 4DO, generating orobanchol (Fig. 1). In addition, the present study suggests that CYP711A forms also may catalyze regioselective modifications on the strigol-type SL scaffolds, which previously were believed to be catalyzed only by the CYP728Bs, which appear to have evolved only in angiosperms.

The results presented here indicate that CYP711As probably played a fundamental role in the structural diversification of SLs before the development of angiosperms, a conclusion that is supported by the observations that CYP711As are responsible for 4DO synthesis in spikemoss (*Selaginella moellendorffii*) (Yoneyama *et al.*, 2018a) and the synthesis of bryosymbiol in the

bryophyte *M. paleacea* (Kodama *et al.*, 2021). Notably all ancestors characterized here generated multiple hydroxylation products from GR24 (Fig. 5) reflecting an inherent promiscuity of the CYP711 ancestors in line with the well-known catalytic versatility of P450 enzymes (Guengerich, 2001; Nelson & Werck-Reichhart, 2011; Hansen *et al.*, 2021). Given that even the oldest ancestor (N0) produced multiple peaks from GR24, it is possible that significant chemical diversification of the SL scaffold occurred before the advent of parasitic angiosperms suggesting the establishment of beneficial mycorrhizal interactions as a key driver of SL diversification. With all the enzymes required for CL biosynthesis already present in ancient land plants, CYP711 diversification could have provided the catalytic versatility necessary to generate a diverse range of noncanonical and canonical SLs before the development of angiosperms. By extension, CYP711s should be considered as likely candidates for the elaboration of a broad range of alternative SL structures in extant plants, including angiosperms, given their much higher degree of conservation than either CYP722Cs or CYP728Bs. Gene duplication in parallel with the development of parasitic angiosperms and the resulting evolutionary arms race between them and their host plants, may then have led to specialization of particular CYP711s to specific activities as well as the co-option of other enzymes (e.g. LSG1) and P450 subfamilies (e.g. CYP722C and CYP728B) to catalyze specialist niche reactions, resulting in the pattern of activities that we see today amongst CYP711s.

In addition to evolutionary studies, ASR also can be used to engineer industrially relevant enzymes, because resurrected ancestral proteins have shown a number of desirable traits, particularly enhanced thermostability (Gumulya & Gillam, 2017; Gumulya *et al.*, 2018). Indeed, some of these traits were evident in this study: most of the ancestral CYP711As expressed better than the extant CYP711As, the N71 ancestor possessed improved catalytic activity, and CYP711A ancestors with varied substrate stereospecificity were found. The N271 ancestor even accepted (–)-ent-GR24, with an unnatural D-ring, as substrate. This opens up the possibility of synthesizing a selection of both native and new-to-nature SLs, in the form of various hydroxylated GR24s, which

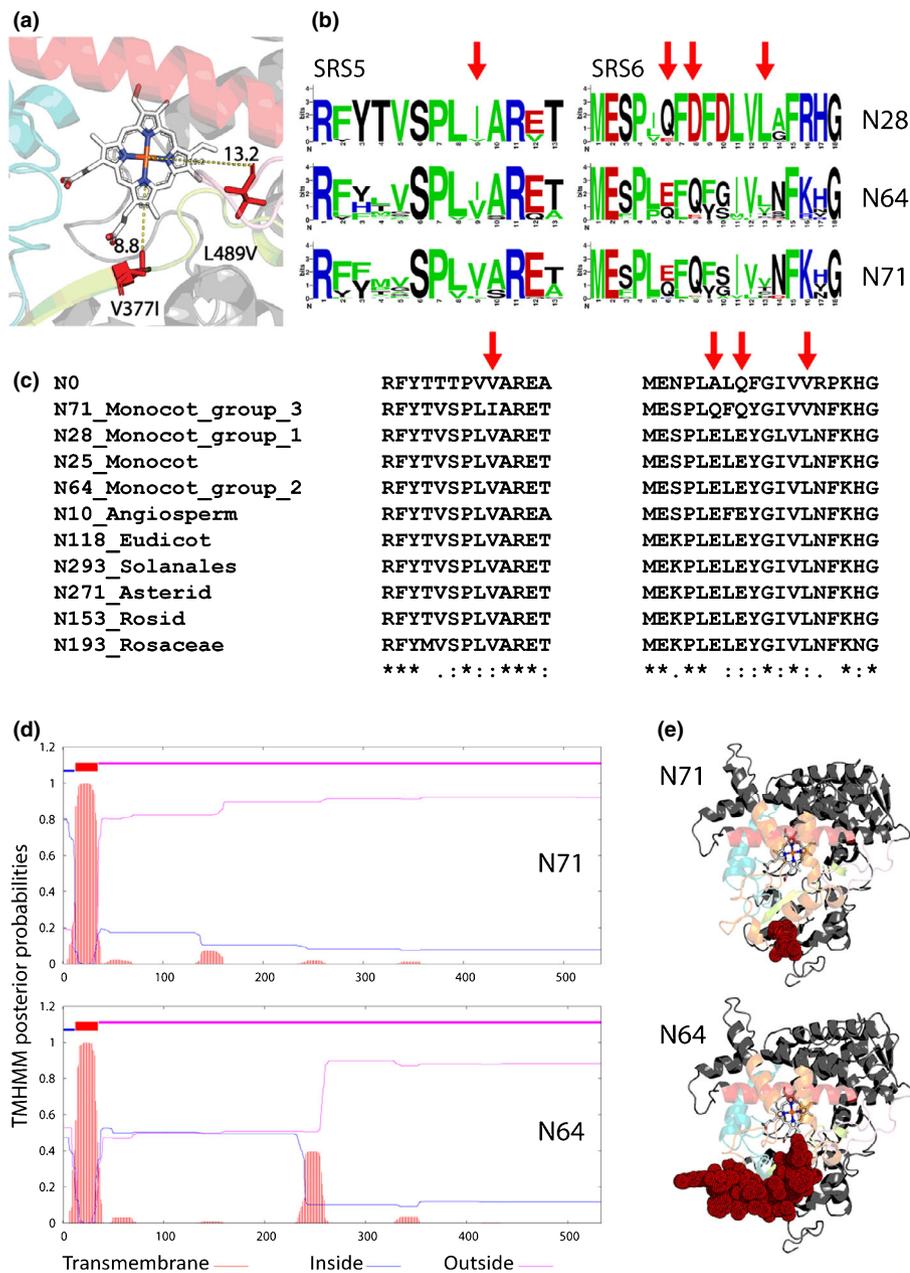


Fig. 8 Structure–function relationships in ancestral CYP711A. (a) View of the active site in the homology model of N71, looking from the distal side of the heme. V377I and L489V mutations of conserved sites observed in substrate recognition sequence (SRS) 5 and SRS6 of the N71 ancestor are in close proximity to the heme reactive centre. (b) Consensus sequence logos representing the diversity in SRS5 and SRS6 in extant sequences descending from the ancestors indicated. (c) Alignment of SRS5 and SRS6 regions found in ancestors resurrected here with residues of interest marked with arrows. (d) Predicted membrane interaction of the holoenzyme N71 and N64 ancestors, impacted by a L222Q substitution in the N71 ancestor. Each residue marked in red represents an amino acid predicted to interact with the membrane. Blue line represents the probability that a certain region is in the membrane; pink line represents the probability that a region is present outside the membrane. (e) Structural models of N71 and N64 with membrane interacting regions marked in red. The view shown is looking down on the active site from the distal side of the heme.

could be tested in bioassays and potentially implemented as novel agrochemicals.

In conclusion, this study has revealed the functional effects of gene duplication events in the monocot grasses, leading to a branch of CYP711As (monocot group 3) derived from an efficient ancestral enzyme that could have been involved in the diversification of strigol-type SLs, a novel property compared to what is known to date for extant CYP711As. These resurrected ancestral CYP711As show potential for further development as biocatalysts for diversification of the SLs structure. Such efforts would be augmented by the availability of co-culture systems for the production of the natural substrates for these enzymes, as demonstrated in an elegant recent study (Wu *et al.*, 2021). Future analysis using these proteins

may provide insights into the exact amino acid changes responsible for specific catalytic activities.

Acknowledgements

The authors gratefully acknowledge the assistance of Amanda Nouwens and Peter Josh of the School of Chemistry and Molecular Biosciences Mass Spectrometry Facility at The University of Queensland. We thank Christopher S. P. McErlean (The University of Sydney) for providing compounds used in this study. MHV was supported by a CSIRO Synthetic Biology Future Science Platform PhD Top-Up Scholarship and an Australian Government Research Training Program Stipend. DC was supported by a China Science Council Scholarship. Australian

Research Council support included FL180100139 and CE200100015. Open Access Funding provided by The University of Queensland.

Author contributions

MHV designed and performed experiments and analyzed data with input from DC, RJC, KY, EMJG and CEV; DC developed methods for analysis of CL and GR24 hydroxylations; RJC developed the CL producing *E. coli* strain; YG and EMJG developed the expression procedure for CYP711As in *E. coli*; REST designed and constructed vectors for co-expression of ATR1; TM performed the initial ASR of the CYP711A family of which the current ASR was based; MHV interpreted the data and wrote the manuscript with input from EMJG, CEV, DC and RJC; EMJG, CEV, BEE, and CAB critically reviewed the manuscript and advised on study design; and EMJG, CAB and CEV conceived the project.

ORCID

Christine A. Beveridge  <https://orcid.org/0000-0003-0878-3110>

Da Cao  <https://orcid.org/0000-0001-7939-1048>

Rebecca J. Chesterfield  <https://orcid.org/0000-0001-7900-6998>

Birgitta E. Ebert  <https://orcid.org/0000-0001-9425-7509>

Elizabeth M. J. Gillam  <https://orcid.org/0000-0003-0378-793X>

Yosephine Gumulya  <https://orcid.org/0000-0002-0529-404X>

Tebogo Matila  <https://orcid.org/0000-0003-4067-984X>

Raine E. S. Thomson  <https://orcid.org/0000-0003-3545-6692>

Claudia E. Vickers  <https://orcid.org/0000-0002-0792-050X>

Marcos H. Vinde  <https://orcid.org/0000-0002-6103-316X>

Kaori Yoneyama  <https://orcid.org/0000-0003-3965-1647>

Data availability

The data that support the findings of this study are available in the supporting information or from the corresponding authors upon reasonable request.

References

- Abe S, Sado A, Tanaka K, Kisugi T, Asami K, Ota S, Ii KH, Yoneyama K, Xie X, Ohnishi T *et al.* 2014. Carlactone is converted to carlactonoic acid by MAX1 in *Arabidopsis* and its methyl ester can directly interact with AtD14 in vitro. *Proceedings of the National Academy of Sciences, USA* 111: 18084–18089.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.
- Brewer PB, Koltai H, Beveridge CA. 2013. Diverse roles of strigolactones in plant development. *Molecular Plant* 6: 18–28.
- Brewer PB, Yoneyama K, Filardo F, Meyers E, Scaffidi A, Frickey T, Akiyama K, Seto Y, Dun EA, Cremer JE *et al.* 2016. Lateral branching oxidoreductase acts in the final stages of strigolactone biosynthesis in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA* 113: 6301–6306.
- Brundrett MC. 2002. Coevolution of roots and mycorrhizas of land plants. *The New Phytologist* 154: 275–304.
- Butler L. 1995. Chemical communication between the parasitic weed *Striga* and its crop host. In: Inderjit A, Dakshini K, Einhellig F, eds. *Allelopathy: organisms, processes and applications, ACS symposium series*. Washington, DC, USA: American Chemical Society, 158–168.
- Cardoso C, Ruyter-Spira C, Bouwmeester HJ. 2011. Strigolactones and root infestation by plant-parasitic *Striga*, *Orobanchae* and *Phelipanche* spp. *Plant Science* 180: 414–420.
- Challis RJ, Hepworth J, Mouchel C, Waites R, Leyser O. 2013. A role for more axillary growth1 (MAX1) in evolutionary diversity in strigolactone signaling upstream of MAX2. *Plant Physiology* 161: 1885–1902.
- Chesterfield RJ, Vickers CE, Beveridge CA. 2020a. Translation of strigolactones from plant hormone to agriculture: achievements, future perspectives, and challenges. *Trends in Plant Science* 25: 1087–1106.
- Chesterfield RJ, Whitfield JH, Pouvreau B, Cao D, Alexandrov K, Beveridge CA, Vickers CE. 2020b. Rational design of novel fluorescent enzyme biosensors for direct detection of strigolactones. *ACS Synthetic Biology* 9: 2107–2118.
- Cook CE, Whichard LP, Turner B, Wall ME, Egley GH. 1966. Germination of witchweed (*Striga lutea* Lour.): isolation and properties of a potent stimulant. *Science* 154: 1189–1190.
- Costa S, Almeida A, Castro A, Domingues L. 2014. Fusion tags for protein solubility, purification, and immunogenicity in *Escherichia coli*: The novel Fh8 system. *Frontiers in Microbiology* 5: 63.
- Delaux PM, Xie X, Timme RE, Puech-Pages V, Dunand C, Lecompte E, Delwiche CF, Yoneyama K, Bécard G, Séjalon-Delmas N. 2012. Origin of strigolactones in the green lineage. *The New Phytologist* 195: 857–871.
- Foley G, Mora A, Ross CM, Bottoms S, Sützl L, Lamprecht ML, Zaugg J, Essebie A, Balderson B, Newell R *et al.* 2020. Identifying and engineering ancient variants of enzymes using GRAPHICAL REPRESENTATION OF ANCESTRAL SEQUENCE PREDICTIONS (GRASP). *BioRxiv*. doi: 10.1101/2019.12.30.891457.
- Forman V, Bjerg-Jensen N, Dyekjær JD, Møller BL, Pateraki I. 2018. Engineering of CYP76AH15 can improve activity and specificity towards forskolin biosynthesis in yeast. *Microbial Cell Factories* 17: 181.
- Gaspar P, Moura G, Santos MAS, Oliveira JL. 2013. mRNA secondary structure optimization using a correlated stem–loop prediction. *Nucleic Acids Research* 41: e73.
- Gillam EMJ, Baba T, Kim BR, Ohmori S, Guengerich FP. 1993. Expression of modified human cytochrome P450 3A4 in *Escherichia coli* and purification and reconstitution of the enzyme. *Archives of Biochemistry and Biophysics* 305: 123–131.
- Gobena D, Shimels M, Rich PJ, Ruyter-Spira C, Bouwmeester H, Kanuganti S, Mengiste T, Ejeta G. 2017. Mutation in sorghum *Low Germination Stimulant 1* alters strigolactones and causes *Striga* resistance. *Proceedings of the National Academy of Sciences, USA* 114: 4471–4476.
- Gotoh O. 1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *The Journal of Biological Chemistry* 267: 83–90.
- Guengerich FP. 2001. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chemical Research in Toxicology* 14: 611–650.
- Guengerich FP, Martin MV, Sohl CD, Cheng Q. 2009. Measurement of cytochrome P450 and NADPH-cytochrome P450 reductase. *Nature Protocols* 4: 1245–1251.
- Gumulya Y, Baek JM, Wun SJ, Thomson RES, Harris KL, Hunter DJB, Behrendorff JBYH, Kulig J, Zheng S, Wu X *et al.* 2018. Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nature Catalysis* 1: 878–888.
- Gumulya Y, Gillam EMJ. 2017. Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the 'retro' approach to protein engineering. *The Biochemical Journal* 474: 1–19.
- Hansen CC, Nelson DR, Møller BL, Werck-Reichhart D. 2021. Plant cytochrome P450 plasticity and evolution. *Molecular Plant* 14: 1244–1265.
- Iseki M, Shida K, Kuwabara K, Wakabayashi T, Mizutani M, Takikawa H, Sugimoto Y. 2018. Evidence for species-dependent biosynthetic pathways for converting carlactone to strigolactones in plants. *Journal of Experimental Botany* 69: 2305–2318.

- Jia K-P, Baz L, Al-Babili S. 2018. From carotenoids to strigolactones. *Journal of Experimental Botany* 69: 2189–2204.
- Kodama K, Rich MK, Yoda A, Shimazaki S, Xie X, Akiyama K, Mizuno Y, Komatsu A, Luo Y, Suzuki H *et al.* 2021. An ancestral function of strigolactones as symbiotic rhizosphere signals. *BioRxiv*. doi: 10.1101/2021.08.20.457034.
- Lumba S, Holbrook-Smith D, McCourt P. 2017. The perception of strigolactones in vascular plants. *Nature Chemical Biology* 13: 599–606.
- Mori N, Nishiuma K, Sugiyama T, Hayashi H, Akiyama K. 2016. Carlactone-type strigolactones and their synthetic analogues as inducers of hyphal branching in arbuscular mycorrhizal fungi. *Phytochemistry* 130: 90–98.
- Nelson D, Werck-Reichhart D. 2011. A P450-centric view of plant evolution. *The Plant Journal* 66: 194–211.
- Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. 1998. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese Cedar Pollen, Cryj2, in *Escherichia coli*. *Applied and Environmental Microbiology* 64: 1694–1699.
- Notley LM, De Wolf CJF, Wunsch RM, Lancaster RG, Gillam EMJ. 2002. Bioactivation of tamoxifen by recombinant human cytochrome P450 enzymes. *Chemical Research in Toxicology* 15: 614–622.
- Parikh A, Gillam EMJ, Guengerich FP. 1997. Drug metabolism by *Escherichia coli* expressing human cytochromes P450. *Nature Biotechnology* 15: 784–788.
- Rensing SA. 2018. Great moments in evolution: the conquest of land by plants. *Current Opinion in Plant Biology* 42: 49–54.
- Tu L, Su P, Zhang Z, Gao L, Wang J, Hu T, Zhou J, Zhang Y, Zhao Y, Liu Y *et al.* 2020. Genome of *Tripterygium wilfordii* and identification of cytochrome P450 involved in triptolide biosynthesis. *Nature Communications* 11: 971.
- Ueno K, Ishiwa S, Nakashima H, Mizutani M, Takikawa H, Sugimoto Y. 2015. Regioselective and stereospecific hydroxylation of GR24 by *Sorghum bicolor* and evaluation of germination inducing activities of hydroxylated GR24 stereoisomers toward seeds of *Striga* species. *Bioorganic & Medicinal Chemistry* 23: 6100–6110.
- Urban P, Mignotte C, Kazmaier M, Delorme F, Pompon D. 1997. Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5. *The Journal of Biological Chemistry* 272: 19176–19186.
- von Wachenfeldt C, Richardson TH, Cosme J, Johnson EF. 1997. Microsomal P450 2C3 is expressed as a soluble dimer in *Escherichia coli* following modifications of its N-terminus. *Archives of Biochemistry and Biophysics* 339: 107–114.
- Wakabayashi T, Hamana M, Mori A, Akiyama R, Ueno K, Osakabe K, Osakabe Y, Suzuki H, Takikawa H, Mizutani M *et al.* 2019. Direct conversion of carlactonoic acid to orobanchol by cytochrome P450 CYP722C in strigolactone biosynthesis. *Science Advances* 5: eaax9067.
- Wakabayashi T, Ishiwa S, Shida K, Motonami N, Suzuki H, Takikawa H, Mizutani M, Sugimoto Y. 2021. Identification and characterization of sorgomol synthase in sorghum strigolactone biosynthesis. *Plant Physiology* 185: 902–913.
- Wakabayashi T, Shida K, Kitano Y, Takikawa H, Mizutani M, Sugimoto Y. 2020. CYP722C from *Gossypium arboreum* catalyzes the conversion of carlactonoic acid to 5-deoxystrigol. *Planta* 2515: 1–6.
- Walker CH, Siu-Ting K, Taylor A, O'Connell MJ, Bennett T. 2019. Strigolactone synthesis is ancestral in land plants, but canonical strigolactone signalling is a flowering plant innovation. *BMC Biology* 17: 70.
- Wang Y, Bouwmeester HJ. 2018. Structural diversity in the strigolactones. *Journal of Experimental Botany* 69: 2219–2230.
- Werck-Reichhart D, Feyereisen R. 2000. Cytochromes P450: a success story. *Genome Biology* 1: REVIEWS3003.
- Wu S, Li Y. 2021. A unique sulfotransferase-involving strigolactone biosynthetic route in sorghum. *Frontiers in Plant Science* 12: 793459.
- Wu S, Ma X, Zhou A, Valenzuela A, Zhou K, Li Y. 2021. Establishment of strigolactone-producing bacterium-yeast consortium. *Science Advances* 7: 4048–4065.
- Xie X, Yoneyama K, Kisugi T, Uchida K, Ito S, Akiyama K, Hayashi H, Yokota T, Nomura T, Yoneyama K. 2013. Confirming stereochemical structures of strigolactones produced by rice and tobacco. *Molecular Plant* 6: 153–163.
- Xie X, Yoneyama K, Yoneyama K. 2010. The strigolactone story. *Annual Review of Phytopathology* 48: 93–117.
- Ye L, Zhu X, Wu T, Wang W, Zhao D, Bi C, Zhang X. 2018. Optimizing the localization of astaxanthin enzymes for improved productivity. *Biotechnology for Biofuels* 11: 278.
- Yoda A, Mori N, Akiyama K, Kikuchi M, Xie X, Miura K, Yoneyama K, Sato-Izawa K, Yamaguchi S, Yoneyama K *et al.* 2021. Strigolactone biosynthesis catalyzed by cytochrome P450 and sulfotransferase in sorghum. *The New Phytologist* 232: 1999–2010.
- Yoneyama K, Mori N, Sato T, Yoda A, Xie X, Okamoto M, Iwanaga M, Ohnishi T, Nishiwaki H, Asami T *et al.* 2018a. Conversion of carlactone to carlactonoic acid is a conserved function of MAX1 homologs in strigolactone biosynthesis. *The New Phytologist* 218: 1522–1533.
- Yoneyama K, Xie X, Yoneyama K, Kisugi T, Nomura T, Nakatani Y, Akiyama K, McErlean CSP. 2018b. Which are the major players, canonical or non-canonical strigolactones? *Journal of Experimental Botany* 69: 2231–2239.
- Zawaira A, Ching LY, Coulson L, Blackburn J, Chun Wei Y. 2011. An expanded, unified substrate recognition site map for mammalian cytochrome P450s: analysis of molecular interactions between 15 mammalian CYP450 isoforms and 868 substrates. *Current Drug Metabolism* 12: 684–700.
- Zhang Y, van Dijk ADJ, Scaffidi A, Flematti GR, Hofmann M, Charnikhova T, Verstappen F, Hepworth J, van der Krol S, Leyser O *et al.* 2014. Rice cytochrome P450 MAX1 homologs catalyze distinct steps in strigolactone biosynthesis. *Nature Chemical Biology* 10: 1028–1033.

Supporting Information

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

Fig. S1 Carlactone production in engineered *Escherichia coli*.

Fig. S2 Exact mass MS2 fragmentation spectra of GR24 and novel GR24 + 16 *m/z* compounds.

Fig. S3 Substrate recognition sites of ancestral CYP711As.

Fig. S4 Sequence identity of substrate recognition sites.

Methods S1 Development of a carlactone-producing strain of *Escherichia coli*.

Table S1 Cloning details for development of the CL-producing *Escherichia coli* strain.

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.