Pathway elucidation and microbial synthesis of proaporphine and bis-benzylisoquinoline alkaloids from sacred lotus (*Nelumbo nucifera*)

Michael E. Pyne [a,b,1,*] & Vincent J. J. Martin [a,b,*]

[a] Department of Biology, Concordia University, Montréal, Québec, Canada

[b] Centre for Applied Synthetic Biology, Concordia University, Montréal, Québec, Canada

[1] Present address: Department of Biology, University of Western Ontario, London, Ontario, Canada

*Correspondence and requests for materials should be addressed to:

Michael E. Pyne (email: mpyne3@uwo.ca); and

Vincent J.J. Martin (email: vincent.martin@concordia.ca)

SIGNIFICANCE

We report the elucidation and microbial reconstitution of the aporphine (*Nn*CYP80G) and bis-benzylisoquinoline (*Nn*CYP80A) branches of alkaloid metabolism from sacred lotus (*Nelumbo nucifera*). Our findings demonstrate that both *Nn*CYP80G and *Nn*CYP80A are stereospecific for (*R*)-substrates. Linking an artificial stereochemical inversion pathway to a high-titer (*S*)-norcoclaurine production strain afforded a *de novo* route to the bis-benzylisoquinoline nelumboferine and the proaporphine pronuciferine via (*R*)-intermediates. We also devised a novel path to liensinine, a promising lotus chemotherapeutic, by screening 66 diverse methyltransferases for activity on nelumboferine. Our work highlights the unique benzylisoquinoline metabolism of *N. nucifera* and enables the targeted overproduction of potential lotus pharmaceuticals using engineered microbial systems.

ABSTRACT

Sacred lotus (*Nelumbo nucifera*) has been utilized as a food, medicine, and spiritual symbol for nearly 3,000 years. The medicinal properties of lotus are largely attributed to its unique profile of benzylisoquinoline alkaloids (BIAs), which includes potential anti-cancer, anti-malarial and anti-arrhythmic compounds. BIA biosynthesis in sacred lotus differs markedly from that of opium poppy and other members of the Ranunculales, most notably in an abundance of BIAs possessing the (*R*)-stereochemical configuration and the absence of reticuline, a major branchpoint intermediate in most BIA producers. Owing to these unique metabolic features and the pharmacological potential of lotus, we set out to elucidate the BIA biosynthesis network in *N. nucifera*. Here we show that lotus CYP80G (*Nn*CYP80G) and a superior ortholog from Peruvian nutmeg (*Laurelia sempervirens*; *Ls*CYP80G) stereospecifically convert (*R*)-armepavine to pronuciferine, a proaporphine alkaloid and the presumed precursor to nuciferine. While sacred lotus employs a dedicated (*R*)-route to aporphine alkaloids from (*R*)-norcoclaurine, we implemented an artificial stereochemical inversion approach to flip the stereochemistry of the core BIA pathway. Exploiting the unique substrate specificity of dehydroreticuline synthase from common poppy (*Papaver rhoeas*) and pairing it with dehydroreticuline reductase enabled *de novo* synthesis of (*R*)-armepavine from (*S*)-norcoclaurine and its subsequent conversion to pronuciferine. We leveraged our stereochemical inversion approach to also elucidate the role of *Nn*CYP80A in sacred lotus metabolism, which we show catalyzes the stereospecific formation of the bis-BIA nelumboferine. Screening our collection of 66 plant *O*-methyltransferases enabled conversion of nelumboferine to liensinine, a potential anti-cancer bis-BIA from sacred lotus.

Keywords: Aporphine, benzylisoquinoline, CYP80, lotus, metabolic engineering, norcoclaurine, yeast

INTRODUCTION

As the national flower of both India and Vietnam, sacred lotus (*Nelumbo nucifera*) has played a major historical role in Asian food, medicine, culture, and religion. Seeds of *N. nucifera* have been consumed for approximately 3,000 years, while most parts of the plant have been used in traditional Chinese medicine to treat fever, diarrhea, cholera, inflammation, and other ailments. The potential health benefits of lotus are largely ascribed to the plant's unique collection of benzylisoquinoline alkaloids (BIAs), which includes common 1-benzylisoquinolines, 15 aporphine alkaloids, seven bis-BIAs, and one unusual tri-BIA (1, 2). BIAs produced by sacred lotus exhibit a range of promising pharmacological properties for potential treatment of cancer (neferine and liensinine), malaria (roemerine), arrhythmia (neferine), HIV (norcoclaurine), obesity (pronuciferine), and fungal infection (roemerine). A recent screen of nearly 200 natural products identified the bis-BIA neferine as a pan-coronavirus entry inhibitor with promising *in vitro* activity against the SARS-CoV-2 virus (3).

The sacred lotus BIA metabolic network and its associated metabolites exhibit a number of structural and stereochemical features that are not found in members of the Ranunculales. (*S*)-Reticuline is the major branchpoint intermediate in most BIA-producing plants, giving rise to more than 2,000 derivatized structures within the morphinan, aporphine, protoberberine, phthalideisoquinoline, and benzophenanthridine downstream branches. However, reticuline has not been isolated from sacred lotus, implying that aporphine and bis-BIA pathways branch from *N*-methylcoclaurine or armepavine in *N. nucifera*. Additionally, aporphine alkaloids from *N. nucifera* lack 4′-substitution of the benzyl moiety that derives from 4-hydroxyphenylacetaldehyde in nearly all BIAs elucidated to date (4). An exception is pronuciferine, an unusual proaporphine alkaloid synthesized by lotus, which possesses a 4′-

substituted benzyl group, implying the presence of a novel uncharacterized dehydration reaction involved in the conversion of pronuciferine to nuciferine (1).

Beyond structural features, BIAs synthesized by sacred lotus exhibit unique stereochemistry compared to other BIA-producing plants. Most BIAs found in nature are (*S*)-conformers owing to the stereospecific synthesis of (*S*)-norcoclaurine by all norcoclaurine synthase (NCS) orthologs characterized to date. In contrast, sacred lotus synthesizes both (*R*)- and (*S*)-conformers of norcoclaurine. Although the *N. nucifera* genome contains at least five NCS isoforms, a dedicated (*R*)- or (*R*,*S*)-yielding NCS has yet to be identified (1). Both (*R*)- and (*S*)-stereoisomers of *N*-methylcoclaurine and armepavine have also been isolated from sacred lotus (5), suggesting a lack of stereospecificity of BIA methyltransferases from *N. nucifera*. Two such lotus methyltransferases involved in the conversion of norcoclaurine to coclaurine (*Nn*6OMT) and *N*-methylcoclaurine to armepavine (*Nn*OMT5) were recently characterized and shown to methylate both (*S*)- and (*R*)-substrates (6). All aporphine alkaloids isolated from *N. nucifera* to date possess the (*R*)-configuration, while bis-BIAs from lotus arise from both (*R*)- and (*S*)-precursors or two (*R*)-monomers (1), highlighting a distinct preference for (*R*)-conformers in downstream branches of the lotus BIA network. These unique structural and stereochemical features of BIA metabolites from *N. nucifera* provide substantial rationale for elucidating BIA biosynthesis in sacred lotus and engineering its production in yeast.

Previously, we constructed a yeast BIA platform with capacity to synthesize more than 4.5 g/L of plant BIAs (4). Owing to the pharmaceutical potential and unique features of BIA metabolism in sacred lotus, we wished to leverage our BIA production platform for the discovery and engineering of lotus BIA biosynthesis in yeast. In this work we design and implement an artificial pathway to invert the stereochemistry of our high titer BIA production platform.

Flipping the stereochemistry of the core BIA pathway enabled access to the bis-BIA and aporphine branches of sacred lotus secondary metabolism. We show that *Nn*CYP80A and *Nn*CYP80G are stereospecific for (*R*)-substrates and catalyze committed steps in the bis-BIA and aporphine alkaloid branches in sacred lotus. Our work sheds light on the unique BIA metabolism of *N. nucifera* and will enable the pharmacological exploration and overproduction of untapped BIA structures synthesized by sacred lotus.

RESULTS

De novo production of armepavine

To begin reconstructing the lotus aporphine biosynthesis pathway in yeast, we selected the (*S*)-norcoclaurine-producing strain LP603 as the basis for our work. This strain harbors an overexpressed and deregulated precursor pathway (*ARO4*^{FBR}, *ARO2*, *ARO7*^{FBR}, *TYR1*, *ARO10*), two copies of an optimized (*S*)-norcoclaurine synthase gene (*CjNCSΔN*), and deletion of several key host genes encoding oxidoreductases (*ari1*Δ *adh6*Δ *ypr1*Δ *ydr541c*Δ *gre2*Δ). LP478, the parent of LP603, produces more than 1 g/L of (*S*)-norcoclaurine in a sucrose pulsed fed-batch fermentation (4).

Armepavine, a key BIA intermediate in sacred lotus, is synthesized from norcoclaurine via successive action of 6-*O*-methyltransferase (6OMT), coclaurine-*N*-methyltransferase (CNMT), and 7-*O*-methyltransferase (7OMT). Recently, OMT5 from *N. nucifera* (*Nn*OMT5) was characterized from sacred lotus and shown to exhibit 7OMT activity on *N*-methylcoclaurine, yielding armepavine (6). To engineer LP603 for high-level armepavine production, we introduced a chimeric armepavine production pathway consisting of 6OMT and CNMT from *Papaver somniferum* (*Ps*6OMT and *Ps*CNMT, respectively), in addition to *Nn*OMT5 (**Fig. 1a**).

6OMT and CNMT variants from opium poppy (*P. somniferum*) dramatically outperformed orthologous isoforms from *N. nucifera* (*Nn*6OMT1-4, *Nn*CNMT1, and *Nn*CNMT3) (**Supplementary Fig. 1**). Strain LP609 synthesized 152 mg/L armepavine in small-scale microtiter plate cultivations, corresponding to a 72% consumption of (*S*)-norcoclaurine (**8**) (**Fig. 1b**). A minor amount of (*S*)-coclaurine (**9**) was produced by strain LP609, while a substantial proportion of (*S*)-*N*-methylcoclaurine (**10**) accumulated, indicating incomplete conversion to armepavine by *Nn*OMT5.

***Nelumbo nucifera* CYP80G stereospecifically converts (*R*)-armepavine to pronuciferine**

The proaporphine alkaloid pronuciferine and the aporphine alkaloid nuciferine exhibit the same methylation pattern as armepavine, suggesting that armepavine is the physiological substrate of the proposed aporphine synthase from *N. nucifera* (*Nn*CYP80G). Having engineered *de novo* synthesis of armepavine, we set out to characterize *Nn*CYP80G activity in yeast. However, introduction of *Nn*CYP80G and a cytochrome P450 reductase (CPR) from *Arabidopsis thaliana* (*At*ATR2; strain LP624) failed to alter the metabolite profile of LP609. Swapping *At*ATR2 for a CPR from *Hypericum calycinum* (*Hc*CPR) or including a cytochrome *b5* variant from *Artemisia annua* (*Aa*CYB5) also failed to yield new metabolite peaks or consumption of (*S*)-armepavine.

Sacred lotus synthesizes both (*S*)- and (*R*)-conformers of armepavine (1), leading us to speculate that *Nn*CYP80G might be specific for one stereoisomer. Our armepavine production strain (LP609) employs *Cj*NCSΔN from *Coptis japonica*, which we showed yields exclusively (*S*)-norcoclaurine (4), implying that armepavine produced by LP609 is the (*S*)-conformer. To clarify the stereochemistry of BIAs synthesized by strain LP609, we performed chiral analysis on norcoclaurine and armepavine, which were found to be exclusively (*S*)-conformers (**Fig. 1c**). We

also detected both (*S*)- and (*R*)-stereoisomers of norcoclaurine and armepavine from commercial extracts of lotus, which is consistent with previous metabolite characterization studies (1).

Because LP609 synthesizes exclusively (*S*)-armepavine, yet *N. nucifera* produces both (*S*)- and (*R*)-stereoisomers, we speculated that *Nn*CYP80G is stereospecific for (*R*)-armepavine. To establish the stereospecificity of *Nn*CYP80G, we introduced a partial aporphine production pathway (*Ps*6OMT-*Ps*CNMT-*Nn*OMT5-*Nn*CYP80G-*At*ATR2) to an unmodified BY4741 laboratory strain lacking an NCS enzyme and a dopamine biosynthesis pathway (yielding strain LP626). LP626 is unable to synthesize BIAs *de novo*, thus enabling its use as a host for feeding various stereoisomers of norcoclaurine. Feeding strain LP626 with (*R*,*S*)-norcoclaurine (*rac*-norcoclaurine), generated by the *in vitro* phosphate-catalyzed condensation of dopamine (**1**) and 4-hydroxyphenylacetaldehyde (**2**) (7), yielded (*R*)-armepavine, in addition to (*S*)-armepavine, demonstrating three-step methylation of (*R*)-norcoclaurine (**3**) to (*R*)-armepavine (**6**) by *Ps*6OMT, *Ps*CNMT, and *Nn*OMT5 (**Fig. 1c**). Importantly, feeding *rac*-norcoclaurine to strain LP626 resulted in the formation of a new LC-MS metabolite peak consistent with pronuciferine (**7**) ($[M + H]^+$ = 312.1600 Da) (**Fig. 1d**; dark blue trace). Omitting *Nn*CYP80G failed to generate the tentative pronuciferine LC-MS product. Unexpectedly, feeding (*S*)-norcoclaurine standard to strain LP626 also generated pronuciferine, yet with a peak area approximately 98% smaller than the corresponding peak derived from supplemented *rac*-norcoclaurine (**Fig. 1d**; red trace). Chiral analysis of commercial (*S*)-norcoclaurine revealed that the standard was contaminated with a small quantity of (*R*)-norcoclaurine (**Fig. 1c**), as also shown in a prior study (4), providing explanation for the low level of pronuciferine synthesized by LP626 from supplemented (*S*)-norcoclaurine. Supplementing phosphate buffer in place of *rac*-norcoclaurine and (*S*)-norcoclaurine failed to generate the tentative pronuciferine LC-MS product (**Fig. 1d**).

Although an authentic standard of pronuciferine is not available, we detected the same tentative LC-MS pronuciferine peak from commercial lotus leaf extract. Pronuciferine from lotus extract exhibited the same retention time and exact mass as the tentative peak derived from supernatants of strain LP626 supplemented with *rac*-norcoclaurine (**Fig. 1d**; light blue trace). Targeted fragmentation of the tentative pronuciferine peak from lotus extract and supernatants of LP626 fed with *rac*-norcoclaurine yielded identical fragmentation spectra (**Fig. 1e**), which were in agreement with published spectra derived from authentic pronuciferine (8, 9). Collectively, these data demonstrate that *Nn*CYP80G stereospecifically converts (*R*)-armepavine to pronuciferine in the committed step of the sacred lotus aporphine biosynthesis pathway.

Based on BIA alkaloids isolated from *N. nucifera*, synthesis of nuciferine arises from an unknown dehydration reaction from pronuciferine (1). Recently, a study of alkaloids from *N. nucifera* identified a cluster of 90 genes closely associated with alkaloid content, including *Nn*CYP80G and several BIA methyltransferases (10). In an attempt to convert pronuciferine to nuciferine, we selected 10 of these candidate biosynthetic genes from lotus (NNU05340, NNU08935, NNU17596, NNU01031, NNU20422, NNU17669, NNU24387, NNU12203, NNU07402, and NNU20868), which were integrated into a pronuciferine-producing strain (LP634). However, none of the candidate lotus genes facilitated accumulation of nuciferine or consumption of pronuciferine.

An artificial *de novo* route to (*R*)-armepavine using reticuline epimerase

Having established a role for *Nn*CYP80G in proaporphine alkaloid biosynthesis using supplementation assays, we aimed to engineer yeast for *de novo* synthesis of pronuciferine. Because lotus ostensibly synthesizes pronuciferine from (*R*)-norcoclaurine in a dedicated (*R*)-

specific pathway (**Fig. 1a**) (1), reconstructing the natural lotus route in yeast would entail screening NCS orthologs from *N. nucifera* and swapping *Cj*NCSΔN for an (*R*)-specific isoform. However, *Cj*NCSΔN dramatically outperformed all NCS variants screened from a diverse array of plants (11-13), enabling high titers (> 4.5 g/L) of BIAs (4).

We searched for an alternative route to (*R*)-armepavine (**6**) from (*S*)-norcoclaurine (**8**) and shifted our focus to reticuline epimerase (REPI), an enzyme involved in the committed step of the morphinan BIA pathway (14-16). REPI catalyzes the two-step stereochemical inversion of (*S*)- to (*R*)-reticuline via the intermediate 1,2-dehydroreticuline. REPI is composed of dehydroreticuline synthase (DRS) and dehydroreticuline reductase (DRR) domains, which catalyze the respective synthesis and reduction of 1,2-dehydroreticuline. While DRS from opium poppy (*P. somniferum*; *Ps*DRS) is specific for (*S*)-reticuline, DRS from common poppy (*P. rhoeas*; *Pr*DRS) exhibits unique preference for (*S*)-*N*-methylcoclaurine (14), the precursor to armepavine in sacred lotus. We speculated that screening diverse pairs of DRS and DRR orthologs might identify a synthetic DRS-DRR couple capable of catalyzing the stereochemical inversion of (*S*)- to (*R*)-armepavine (**Fig. 2a**). A concurrent study employed a similar approach for the inversion of (*S*)- to (*R*)-*N*-methylcoclaurine, albeit using supplementation with L-DOPA (17). Instead, we wished to develop a fully *de novo* route to (*R*)-armepavine and pronuciferine without supplementation of BIA pathway precursors.

To engineer a *de novo* stereochemical inversion path to pronuciferine, we first screened several DRS variants for activity on (*S*)-*N*-methylcoclaurine and (*S*)-armepavine. This screen included *Pr*DRS, *Ps*DRS, DRS from *Chelidonium majus* (*Cm*DRS) and four isoforms from Iranian poppy (*P. bracteatum*; *Pb*DRS1-4). *Cm*DRS and *Pb*DRS1-4 share 86-99% identity with the canonical *Ps*DRS variant, whereas *Pr*DRS shares only 79% identity to *Ps*DRS (**Fig. 2b**). In

agreement with a prior study (14), *Pr*DRS exhibited remarkable activity on (*S*)-*N*-methylcoclaurine, yielding 1,2-dehydro-*N*-methylcoclaurine (**12**) ($[M + H]^+$ = 298.1443 Da) (**Fig. 2c**).

In addition to its reported activity on (*S*)-*N*-methylcoclaurine, we screened *Pr*DRS and other DRS variants on (*S*)-armepavine. Remarkably, *Pr*DRS was the only DRS variant that converted (*S*)-armepavine to 1,2-dehydroarmepavine (**13**) ($[M + H]^+$ = 312.1600 Da) (**Fig. 2c**). Implementation of *Pr*DRS converted nearly 70% of total (*S*)-substrate [combined (*S*)-*N*-methylcoclaurine and (*S*)-armepavine] to dehydro-intermediates. Assuming equivalent LC-MS response factors, the strain harboring *Pr*DRS produced 45-fold more 1,2-dehydro-*N*-methylcoclaurine than 1,2-dehydro-armepavine, although consumption of (*S*)-*N*-methylcoclaurine reduces the pool of available (*S*)-armepavine. Aside from *Pr*DRS, *Ps*DRS consumed 29% of (*S*)-substrates, while the remaining DRS orthologs converted a very small proportion of available (*S*)-conformers to the corresponding dehydro-derivatives.

Using *Pr*DRS for synthesis of 1,2-dehydro-BIAs, we next screened DRR variants for reduction of 1,2-dehydro-*N*-methylcoclaurine and 1,2-dehydroarmepavine. Our screen included *Pr*DRR, *Ps*DRR, and two isoforms of DRR from *P. bracteatum* (*Pb*DRR1 and *Pb*DRR2) (**Fig. 2d**). Introduction of all four DRR orthologs facilitated the near-complete consumption of 1,2-dehydroarmepavine. While none of the DRR variants completely consumed 1,2-dehydro-*N*-methylcoclaurine, strains harboring *Ps*DRR or *Pb*DRR2 each reduced 68% of available substrate. Chiral LC-MS analysis of supernatants derived from strains harboring *Pr*DRS and *Ps*DRR revealed synthesis of (*R*)-armepavine, while strains lacking a DRR domain produced exclusively (*S*)-armepavine (**Fig. 2e**). Inclusion of *Nn*CYP80G and ATR2 along with *Pr*DRS and *Ps*DRR facilitated production of pronuciferine (**Fig. 2d, e**), further establishing the involvement

of (*R*)-armepavine in its synthesis. Collectively, these modifications establish an artificial *de novo* route to (*R*)-armepavine and pronuciferine via (*S*)-norcoclaurine.

Enhancing CYP80G-CPR coupling and probing activity of *Nn*CYP80A

After establishing the role of *Nn*CYP80G in proaporphine alkaloid biosynthesis, we attempted to improve its activity through better coupling to a CPR partner. We assembled a combinatorial CYP80-CPR-CYB5 library by pairing different combinations of plant CPRs and CYB5s with *Nn*CYP80G or *Ls*CYP80G, a predicted CYP80G ortholog from *Laurelia sempervirens* identified from our in-house collection of plant CYPs. We also included in this assay *Nn*CYP80A, a lotus enzyme possessing 80% identity to *Nn*CYP80G. Genes encoding *NnCYP80G* and *NnCYP80A* are adjacent within the lotus genome, suggesting tandem duplication and a potential role for *Nn*CYP80A in BIA biosynthesis (1). Previous studies have implicated *Nn*CYP80A in bis-BIA biosynthesis, yet its activity has not been experimentally validated. Because all bis-BIAs characterized from lotus derive from at least one (*R*)-monomer, our ability to synthesize both (*S*)- and (*R*)-conformers using *Pr*DRS and DRR provides an opportunity to fully explore *Nn*CYP80A activity.

We first assembled a CYP80-CPR-CYB5 library consisting of 24 combinations of one CYP80 (*Nn*CYP80A, *Nn*CYP80G, and *Ls*CYP80G), each paired with one of four CPRs and one of four CYB5s (**Fig. 3a**). We then transformed our 24 strain CYP80-CPR-CYB5 library with a plasmid that expresses both *PrDRS* and *PsDRR* genes for conversion of (*S*)-intermediates to (*R*)-conformers. As we demonstrated above (**Fig. 1d**), all *Nn*CYP80G strains expressing *PrDRS* and *PsDRR* produced pronuciferine (**Fig. 3b**). However, all strains harboring *Ls*CYP80G synthesized more pronuciferine than isogenic strains possessing *Nn*CYP80G. Selection of both CPR and

CYB5 variant had a significant effect on pronuciferine production, whereby pronuciferine output was greatest when *Ls*CYP80G was coupled with *Cc*CPR2 and *At*CYB5D.

Expression of *PrDRS* and *PsDRR* in strains containing *Nn*CYP80A and different combinations of CPR and CYB5 triggered the formation of a new LC-MS metabolite peak ([M + H]$^+$ = 597.2965 Da) consistent with a high molecular weight bis-BIA product (**Fig. 3c, d**). Selection of both CPR and CYB5 variant had a significant effect on production of the tentative bis-BIA product. The bis-BIA metabolite(s) eluted as a double peak and its formation required *Nn*CYP80A as well as *Pr*DRS and *Ps*DRR (**Fig. 3e**; dark pink trace), indicating a requirement for (*R*)-substrates.

At least seven bis-BIAs have been isolated from *N. nucifera* (1), of which four, nelumborine, nelumboferine, 6-hydroxynorisoliensinine, and *N*-norisoliensinine have a molecular weight and chemical formula ($C_{36}H_{40}N_2O_6$) consistent with the metabolite peak generated by yeast strains harboring *Nn*CYP80A, *Pr*DRS and *Ps*DRR. Of these tentative products, nelumboferine (**14**) and nelumborine (**15**) possesses a methylation profile consistent with methyltransferases produced by our combinatorial library, as 6-hydroxynorisoliensinine and *N*-norisoliensinine lack methyl groups furnished by *Ps*6OMT and *Ps*CNMT, respectively. Nelumboferine and nelumborine both derive from the coupling of two monomers of *N*-methylcoclaurine (18), indicating that *Nn*OMT5 is not involved in their formation (**Fig. 3d**). To clarify the role of *Nn*OMT5 in the synthesis of the tentative bis-BIA metabolite, we deleted *NnOMT5* from the best performing *Nn*CYP80A strain harboring *At*ATR2 and *Pb*CYB5. In agreement with the presumed routes to nelumboferine and nelumborine in *N. nucifera*, deletion of *NnOMT5* had no effect on synthesis of the bis-BIA metabolite (**Fig. 3e**; light pink trace).

To further probe the identity of the tentative bis-BIA metabolite, we analyzed sacred lotus extracts for the presence of high molecular weight bis-BIA metabolites. We identified a metabolite peak with the same exact mass ($[M + H]^+$ = 597.2965 Da) and retention time in lotus leaf extract as the tentative bis-BIA product synthesized by yeast strains harboring *Nn*CYP80A, *Pr*DRS and *Ps*DRR (**Fig. 3e**; green trace). In line with yeast strains expressing *NnCYP80A*, *PrDRS* and *PsDRR*, the lotus bis-BIA metabolite eluted as a double peak in commercial lotus extract. Targeted fragmentation of the tentative bis-BIA product from lotus extract and yeast supernatants yielded several common fragment ions, notably $[M + H]^+$ = 192.1, 475.2, 554.2, and 566.2 Da (**Fig. 3f**).

We also screened the CYP80-CPR-CYB5 combinatorial strain library in the absence of a DRS-DRR pair to search for tentative products arising from exclusively (*S*)-metabolites. In agreement with known stereochemistry of lotus alkaloids (1), no tentative bis-BIA or proaporphine metabolite peaks were identified, confirming an inability of *Nn*CYP80A, *Nn*CYP80G, and *Ls*CYP80G to turnover substrates possessing exclusively (*S*)-configurations.

Guattegaumerine (**16**) is a bis-BIA metabolite synthesized by *Berberis stolonifera*. Both guattegaumerine and nelumboferine derive from two units of (*R*)-*N*-methylcoclaurine, yet guattegaumerine derives from 3′-*O*-4′ phenol coupling, while nelumboferine arises from 7-*O*-3′ coupling (**Fig. 3d**). Recently, yeast was engineered to synthesize guattegaumerine from (*R*)-*N*-methylcoclaurine by employing CYP80A1 from *B. stolonifera* (*Bs*CYP80A1), the product of which was validated using authentic guattegaumerine standard (17). Because guattegaumerine, nelumboferine, and nelumborine are structural isomers ($C_{36}H_{40}N_2O_6$) generated through different phenol coupling reactions, we reasoned that our *Nn*CYP80A-derived bis-BIA product would yield a similar LC-MS retention time and MS/MS fragmentation spectra to that of *Bs*CYP80A1-

derived guattegaumerine. To first demonstrate synthesis of guattegaumerine, we paired *Bs*CYP80A1 with four CPR variants in our (*R*)-*N*-methylcoclaurine host lacking *NnOMT5*. All *Bs*CYP80A1 strains facilitated the synthesis of a new LC-MS peak with the same exact mass as guattegaumerine ($[M + H]^+$ = 597.2965 Da) (**Fig. 3e**; black trace). Pairing of *Bs*CYP80A1 with *Hc*CPR yielded approximately 45-fold more guattegaumerine than an isogenic strain harboring *Cc*CPR2 (**Supplementary Figure 2**). Production of guattegaumerine required expression of *PrDRS* and *PsDRR*, corroborating the requirement for (*R*)-*N*-methylcoclaurine by *Bs*CYP80A1 (17). In contrast to the *Nn*CYP80A-derived bis-BIA product, guattegaumerine eluted as a single LC-MS peak, yet with the same retention time as the *Nn*CYP80A product, presumably due to the different phenol coupling reactions involved in the synthesis of guattagaumerine, nelumboferine, and nelumborine. We compared MS/MS fragmentation of *Nn*CYP80A- and *Bs*CYP80A1-derived products (**Fig. 3f**), which yielded the same major 597→192 transition generated by fragmentation of authentic guattegaumerine standard (17). Collectively, these data are consistent with a tentative assignment of nelumboferine or nelumborine for the identity of the *Nn*CYP80A-derived bis-BIA product.

Derivatization of nelumboferine by screening a library of BIA *O*-methyltransferases

In sacred lotus, nelumboferine (**14**) is converted to liensinine (**17**) or isoliensinine (**18**) via methylation by 7OMT or 4′OMT, respectively (18). One or both of these compounds are subsequently methylated to give neferine (**19**) (**Fig. 4a**). We screened our in-house library of BIA methyltransferase enzymes for activity on nelumboferine. In total we introduced 66 putative *O*-methyltransferases from 19 BIA-producing plant species to our best *Nn*CYP80A strain (**Fig. 4b**). Our search included seven predicted *O*-methyltransferases from *N. nucifera* (*Nn*6OMT1-4

and *Nn*7OMT1-3). We analyzed supernatants of methyltransferase-expressing strains for new peaks consistent with single ($[M + H]^+$ = 611.3121 Da) or double methylation ($[M + H]^+$ = 625.3278 Da) of the *Nn*CYP80A-derived bis-BIA product ($[M + H]^+$ = 597.2965 Da). One *O*-methyltransferase candidate (*Ec*OMT1 from *Eschscholzia californica*) generated a peak ($[M + H]^+$ = 611.3121 Da) consistent with single methylation (**Fig. 4c**). Based on a prior study, *Ec*OMT1 is a 7OMT that catalyzes the conversion of reticuline to laudanine (19). In line with its reported activity, *Ec*OMT1 clusters with other enzymes exhibiting confirmed or predicted 7OMT activity (**Fig. 4b**), including *Ec*OMT2, a scoulerine 9OMT (SOMT) that also possesses 7OMT activity on reticuline (20).

Based on the reported 7OMT activity of *Ec*OMT1, the identity of the new methylated LC-MS product is expected to be liensinine ($C_{37}H_{42}N_2O_6$) (**Fig. 4a**), a major bis-BIA in lotus. We obtained authentic liensinine standard, which generated a mass ($[M + H]^+$ = 611.3121 Da) and retention time (RT = 2.16 min) in agreement with the LC-MS product derived from *Ec*OMT1 activity on nelumboferine (**Fig. 4c**). We also compared MS/MS fragmentation of the tentative liensinine metabolite peak with authentic liensinine standard, which generated identical fragmentation spectra (**Fig. 4d**). Specifically, the major fragment ion with mass $[M + H]^+$ = 206.1 was previously used to differentiate liensinine from its isomer isoliensinine ($C_{37}H_{42}N_2O_6$) (21), which arises from 4′-*O*-methylation of nelumboferine in place of 7-*O*-methylation. The fragment ion with mass $[M + H]^+$ = 206.1 also differentiates liensinine from its nelumboferine precursor ($C_{36}H_{40}N_2O_6$), which yields a corresponding fragment ion with mass $[M + H]^+$ = 192.1 (**Fig. 3f, g**), indicating *O*-methylation of the free tetrahydroisoquinoline moiety of nelumboferine. Importantly, synthesis of liensinine by yeast strains harboring *Nn*CYP80A, ATR2, and *Ec*OMT1 confirms that nelumboferine is a product of CYP80A from *N. nucifera*.

We also identified liensinine in commercial lotus extract (RT = 2.16 min), although an additional larger peak (RT = 2.25 min) with the same mass was also observed (**Fig. 4c**). As expected, MS/MS fragmentation of the minor lotus peak yielded the key 611 → 206 transition characteristic of liensinine (**Fig. 4e;** dark green trace). In contrast, fragmentation of the dominant lotus product (RT = 2.25 min) yielded a major 611 → 192 transition (**Fig. 4e;** light green trace), which is in agreement with isoliensinine (21).

We also introduced our panel of 66 methyltransferases to our guattegaumerine-producing strain harboring *Bs*CYP80A1 and *Hc*CPR, yet were unable to detect methylated derivatives ($[M + H]^+$ = 611.3121 Da or 625.3278 Da). Our screen identified several methyltransferases with tentative 7OMT or 4′OMT activities on *N*-methylcoclaurine (**Supplementary Table 1**).

DISCUSSION

In contrast to most BIA-producing plants, in which the core BIA pathway proceeds through exclusively (*S*)-conformers, sacred lotus synthesizes (*R*)-*N*-methylcoclaurine and (*R*)-armepavine in a dedicated (*R*)-stereochemical path from (*R*)-norcoclaurine. In this study, we designed and implemented a synthetic strategy to flip the stereochemistry of our BIA platform strain by exploiting the unique substrate preference of DRS from common poppy (*Pr*DRS). Pairing of *Pr*DRS with several DRR variants enabled stereochemical inversion of (*S*)-*N*-methylcoclaurine and (*S*)-armepavine via 1,2-dehydro-*N*-methylcoclaurine and 1,2-dehydro-armepavine intermediates. Production of (*R*)-conformers enabled the characterization of *Nn*CYP80G and *Nn*CYP80A as respective stereospecific proaporphine and bis-BIA synthases involved in the committed steps of the aporphine and bis-BIA branches of BIA metabolism in sacred lotus.

NnCYP80A and *NnCYP80G* genes are adjacent in the genome of *N. nucifera* and the corresponding enzymes share a high degree of sequence similarity, suggesting a tandem duplication event and subsequent divergence of enzymatic activity. Our findings firmly establish roles for *Nn*CYP80G and *Nn*CYP80A in propaporphine and bis-BIA biosynthesis, respectively. Specifically, we show that *Nn*CYP80G is a proaporphine synthase that catalyzes the stereospecific conversion of (*R*)-armepavine to pronuciferine. Formation of pronuciferine was dependent on implementation of *Pr*DRS-*Ps*DRR or exogenous supplementation of (*R,S*)-norcoclaurine, clearly establishing the stereospecificity of *Nn*CYP80G for (*R*)-armepavine. Our findings, coupled with the recent discovery and characterization of *Nn*OMT5 (6), the dedicated 7OMT responsible for armepavine production, have unlocked the aporphine branch of BIA biosynthesis in sacred lotus. Nuciferine, a major alkaloid in *N. nucifera* and a metabolite with potential anti-cancer properties, possesses the same methylation pattern as pronuciferine, indicating that a hitherto unidentified enzyme is responsible for conversion of pronuciferine to nuciferine. In an attempt to identify the unknown activity, we implemented 10 lotus metabolic enzymes with expression profiles similar to *Nn*CYP80G and other core BIA pathway enzymes. However, we were unable to identify the elusive missing enzyme responsible for the conversion of pronuciferine to nuciferine. More sophisticated enzyme discovery strategies, such as those employed in the recent elucidation of colchicine and strychnine biosynthetic pathways (22, 23), will be required to complete the lotus pathway to nuciferine.

*Nn*CYP80A was previously implicated in bis-BIA biosynthesis, while other studies have proposed it to be an alternative isoform of *Nn*CYP80G (1). We leveraged our artificial stereochemical inversion strategy to clarify the role of *Nn*CYP80A in sacred lotus metabolism. We show that *Nn*CYP80A is the dedicated bis-BIA synthase involved in the committed step of

the *N. nucifera* bis-BIA biosynthesis pathway. To date seven high molecular weight bis-BIA metabolites have been isolated from sacred lotus, in addition to an unusual tri-BIA (neoliensinine) (1). Of the characterized lotus bis-BIAs, only nelumboferine and nelumborine possess the exact mass ($[M + H]^+$ = 597.2965 Da; $C_{36}H_{40}N_2O_6$) and methylation profile consistent with the *Nn*CYP80A-catalyzed metabolite. Nelumborine and nelumboferine possess the same methylation pattern, yet derive from distinct phenol coupling reactions. Whereas nelumboferine and other lotus bis-BIAs derive from a C7-*O*-C3′ phenol coupling, nelumborine arises from an unusual C8-C3′ coupling. We showed that formation of the *Nn*CYP80A bis-BIA product was independent of *Nn*OMT5, yet required the presence of *Pr*DRS-*Ps*DRR, implicating involvement of (*R*)-*N*-methylcoclaurine in its synthesis. Consistent with this observation, nelumboferine arises from two units of (*R*)-*N*-methylcoclaurine (24), while to our knowledge the stereochemistry of *N*-methylcoclaurine monomers in the synthesis of nelumborine has not been deduced.

Based on the methylation profile of BIAs isolated from *N. nucifera*, nelumboferine gives rise to neferine, the fully methylated bis-BIA end product in lotus, via convergent routes from liensinine or isoliensinine (18) involving unidentified 7OMT and 4′OMT enzymes. We introduced our library of 66 plant methyltransferases into our top bis-BIA-producing strain, including seven predicted OMTs from *N. nucifera*, and identified *Ec*OMT1 from California poppy (*Eschscholzia californica*) as a promiscuous 7OMT with activity on nelumboferine. Implementation of *Ec*OMT1 in a strain containing *Nn*CYP80A, ATR2, *Pr*DRS, and *Ps*DRR facilitated synthesis of liensinine, which was confirmed using an authentic liensinine standard. Liensinine is a major bis-BIA metabolite in sacred lotus with promising anti-cancer activity (25) and our findings establish an artificial *de novo* path for microbial overproduction of liensinine.

While we did not identify the dedicated *N. nucifera* bis-BIA 7OMT involved in liensinine formation, our findings might aid in the identification of the elusive enzyme by using *Ec*OMT1 to query the lotus genome, which encodes at least 60 candidate *O*-methyltransferase genes (26). Moreover, repeating our screen of diverse plant *O*-methyltransferases might unveil a novel 4′OMT variant with activity on liensinine, thus enabling synthesis of neferine, another prospective anti-cancer pharmaceutical.

Formation of liensinine by the concerted activity of *Nn*CYP80A and *Ec*OMT1 confirms nelumboferine as a product of CYP80A activity in *N. nucifera*. We compared *Nn*CYP80A activity to that of *Bs*CYP80A1, a recently characterized bis-BIA synthase from *Berberis stolonifera*. *Bs*CYP80A1 catalyzes the 3′-4′ C-O phenol coupling of two units of (*R*)-*N*-methylcoclaurine to give guattegaumerine or one unit each of (*R*)- and (*S*)-*N*-methylcoclaurine, yielding berbamunine (17). Interestingly, *Nn*CYP80A shares a higher degree of sequence similarity with *Nn*CYP80G than with *Bs*CYP80A1 (80% and 45%, respectively), despite confirmed roles of the CYP80A variants in bis-BIA biosynthesis. Products of *Bs*CYP80A1 and *Nn*CYP80A possess the same mass ($[M + H]^+$ = 597.2965 Da; $C_{36}H_{40}N_2O_6$) and expected methylation profile, yet the *Nn*CYP80A product eluted as two distinct LC-MS peaks, while strains harboring *Bs*CYP80A1 produced a single peak. Although we confirmed nelumboferine as a product of *Nn*CYP80A based on conversion to liensinine, it is possible that *Nn*CYP80A is involved in the biosynthesis of both nelumboferine and nelumborine by catalyzing dual coupling reactions (C7-*O*-C3′ and C8-C3′, respectively), which might provide justification for the two LC-MS peaks derived from *Nn*CYP80A activity. However, the biosynthetic origin of nelumborine in *N. nucifera* remains unclear.

In addition to *Nn*CYP80A and *Bs*CYP80A1 variants, Asian moonseed (*Menispermum dauricum*) and Canadian moonseed (*M. canadense*) synthesize dauricine, a potential anti-cancer bis-BIA that arises from the C3′-*O*-C4′ coupling of two units of (*R*)-armepavine. A previous study screened five plant CYP80A orthologs in an (*R*)-armepavine-producing strain, yet did not observe production of dauricine (17). We attempted to devise an artificial route to dauricine through direct methylation of guattegaumerine by screening our library of 66 plant methyltransferases in a strain harboring *Bs*CYP80A1 and *Hc*CPR. However, we were also unsuccessful in synthesizing dauricine or methylated derivatives of guattegaurmerine. A more exhaustive screen of diverse plant CYP80A orthologs on stereoisomers of both *N*-methylcoclaurine and armepavine might unveil new coupling activities, such as the C3′-*O*-C4′ coupling route of (*R*)-armepavine in the biosynthesis of dauricine.

Finally, our study highlights the utility of high-titer microbial strains in plant enzyme discovery and pathway elucidation. BIA output from our (*S*)-norcoclaurine strain exceeds 1.5 g/L (4), which provides an ideal background for enzyme characterization, particularly those with poor catalytic properties or off-target activities. By repurposing our BIA platform, we achieved *de novo* synthesis of 152 mg/L armepavine, a 14-fold improvement over a previous study (11.1 mg/L armepavine) in which L-DOPA precursor supplementation was employed to boost BIA synthesis (17). Relatedly, guattegaumerine output varied by a factor of 45 across all four of our *Bs*CYP80A1-CPR strains, while an earlier study was unable to detect guattegaumerine in two out of four CPR pairings (17). We also detected pronuciferine or bis-BIA synthesis by all 24 strains of our combinatorial CYP80-CPR-CYB5 library, again showcasing the utility of our high titer BIA platform in the discovery and elucidation of plant secondary metabolic pathways. Our characterization of both *Nn*CYP80A and *Nn*CYP80G herein will enable the exploration and

overproduction of potential aporphine and bis-BIA therapeutics from sacred lotus, such as nuciferine, liensinine, and neferine.

Acknowledgments

We thank Nicholas Gold and the Concordia Genome Foundry for assistance with combinatorial strain construction, and Marcos DiFalco and Lauren Narcross for assistance with LC-MS analyses. This study was financially supported by an NSERC-Industrial Biocatalysis Network (IBN) grant and an NSERC Discovery grant. M.E.P. was supported by an NSERC Postdoctoral Fellowship. V.J.J.M. is supported by a Concordia University Research Chair.

Author contributions

M.E.P. designed the research and performed the experiments with supervision from V.J.J.M. M.E.P. and V.J.J.M. wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the online version of the paper. Correspondence and requests for materials should be addressed to M.E.P. and V.J.J.M.

Data availability

Data supporting the findings of this work are available within the paper and its SI Appendix. Source data for the figures in this paper and the SI Appendix are provided in Dataset S1.

METHODS

Chemicals and reagents. General chemicals and reagents were purchased commercially, including dopamine (Sigma-Aldrich), (*S*)-norcoclaurine (Toronto Research Chemicals), 4-hydroxyphenylacetaldehyde (Toronto Research Chemicals), liensinine (MedChemExpress), and (*R*,*S*)-armepavine (MuseChem), unless stated otherwise. Lotus leaf extract was purchased from Amazon (Plum Flower Brand He Ye) or eBay (Hawaii Pharm Pink Lotus herbal extract).

Plasmids, strains, and growth media. The quadruple auxotrophic *S. cerevisiae* strain BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) is the origin of all engineered strains in this study. Strain LP478 was utilized as the basis for this work(4). Cultures were grown in YPD (10 g/L Bacto Yeast Extract, 20 g/L Bacto peptone, 20 g/L glucose) or synthetic complete (SC) medium [6.7 g/L Difco Yeast Nitrogen Base (YNB) without amino acids, 1.62-1.92 g/L Drop-out Medium Supplements (Millipore-Sigma) minus appropriate amino acids, 20 g/L glucose). Recombinant cells were selected on YPD agar containing 400 μg/mL G418, 200 μg/mL hygromycin B, or 200 μg/mL of each antibiotic. Selection using auxotrophic markers was performed in SC medium lacking histidine. Oligonucleotides utilized in this study are listed in Dataset S2. Plasmids and strains employed in this work are provided in **Supplementary Tables 2** and **3**, respectively.

Yeast strain construction. Genetic modifications to yeast were made via CRISPR-Cas9-mediated genomic integration (27, 28) and *in vivo* DNA assembly (29). Cas9 and gRNA were delivered to yeast using pCas-G418 (ref. (28)) or pCas-Hyg (ref. (30)). Linear gRNA cassettes

were retargeted by PCR and assembled using *in vivo* gap repair with a BsaI- and NotI-digested pCas backbone (30). Roughly 150 ng of digested pCas was combined with 300 ng of linear gRNA cassette and 500-1,000 ng of total repair DNA in a standard 50 μL lithium acetate transformation. Cells were heat-shocked at 42 °C for 30 minutes, recovered for 16 hours, and plated onto yeast agar plates containing appropriate antibiotics. Both linear pCas-G418 and pCas-Hyg plasmids were used for complex multi-part DNA assemblies. The CYP80-CPR-CYB5 combinatorial library was assembled using the Echo 550 acoustic liquid handler. Design of gRNAs was performed by selecting top hits common to both CCTop (31) and a yeast CRISPRi web tool (32). Chromosomal loci for DNA integration were selected from previous studies (11, 33-35). Gene expression cassettes and genomic integration sites utilized in this work are listed in **Supplementary Tables 4** and **5**, respectively. Synthetic DNAs employed in this study are provided in **Dataset S3**. The yeast *GRE2* gene was replaced with a synthetic DNA landing pad (LP5.T3) possessing a unique Cas9 target site (T3).

Microtiter plate assay for BIA quantification. Colonies were picked in triplicate into 0.5 mL of 2× SCS medium (13.4 g/L Difco Yeast Nitrogen Base (YNB) without amino acids, 2× Drop-out Medium Supplements (Millipore-Sigma), 40 g/L sucrose) within 96-well deep well plates. Following 24 h of growth, cultures were back-diluted 50× into 0.5 mL of fresh 2× SCS medium in 96-well deep well plates. Cultures were grown at 30 °C with shaking at 350 rpm. Following 48-72 h of growth, OD_{600} measurements were taken and culture broth was stored at -20 °C for subsequent analysis by LC-MS.

LC-MS and chiral analysis of BIA metabolites. BIA metabolites were extracted from culture broth containing cells and growth medium. Seven μL of culture broth was combined with 27 μL of 100% acetonitrile (ACN). Following five minutes of shaking at room temperature, 146 μL of

0.12% formic acid was added to give a final concentration of 15% ACN and 0.1% formic acid. Samples were diluted an additional 10-fold in 15% ACN and 0.1% formic acid. Samples were centrifuged at 4,000 RCF and 10 μL of extracted culture supernatant was separated on a 1290 Infinity II LC system (Agilent Technologies). Metabolites were separated using the following gradient: 2% B to 10% B from 0-4 min (0.3 mL/min), 10% B to 85% B from 4-6 min (0.3 mL/min), held at 85% B from 6-7 min (0.3 mL/min), 85% B to 2% B from 7-7.1 min (0.3 mL/min), and held at 2% B from 7.1-9 min (0.45 mL/min). Solvent A was 0.1% formic acid in water and solvent B was 0.1% formic acid in 100% ACN. Following separation, eluent was injected into an Agilent 6545 quadrupole time-of-flight MS (QTOF-MS; Agilent Technologies) equipped with a Zorbax Eclipse Plus C18 column (50 × 2.1 mm, 1.8 μm; Agilent Technologies). The sample tray and column compartment were set to 4 °C and 30 °C, respectively. The sheath gas flow rate and temperature were adjusted to 10 L/min and 350 °C, respectively, while drying and nebulizing gases were set to 12 L/min and 55 psig, respectively. The drying gas temperature was set to 325 °C. QTOF data was processed and manipulated using Agilent MassHunter Qualitative Analysis software. Norcoclaurine ($[M + H]^+$ = 272.1287 Da), coclaurine ($[M + H]^+$ = 286.1443 Da), *N*-methylcoclaurine ($[M + H]^+$ = 300.1600 Da), armepavine ($[M + H]^+$ = 314.1756 Da), 1,2-dehydro-*N*-methylcoclaurine ($[M + H]^+$ = 298.1443 Da), 1,2-dehyroarmepavine ($[M + H]^+$ = 312.1600 Da) were identified by exact mass. Norcoclaurine and armepavine were quantified using authentic standards. Pronuciferine ($[M + H]^+$ = 312.1600 Da) and nelumboferine ($[M + H]^+$ = 597.2965 Da) were identified by retention time, exact mass, and MS/MS fragmentation by comparison to extracts from leaves of *N. nucifera*. Guattegaumerine ($[M + H]^+$ = 597.2965 Da) was identified by exact mass and MS/MS fragmentation. Liensinine ($[M + H]^+$ = 611.3121 Da) was identified by retention time, exact mass, and MS/MS

fragmentation by comparison to an authentic standard. MS/MS fragmentation of pronuciferine and guattegaumerine were compared against published fragmentation spectra.

Chiral analysis of norcoclaurine and armepavine was performed by modifying an existing method. Culture supernatant from yeast strains grown in deep-well plates were centrifuged to remove debris and diluted 1:50 or 1:100 in H_2O. Ten μL was loaded onto a SHODEX ORPak CDBS-453 column and separated chromatographically using a 1290 Infinity II HPLC (Agilent Technologies) and the following gradient: 0–18 min, 5% B; 18–19 min, 95% B; 19–24 min, 95% B; 24-25 min, 5% B; 25-40 min, 5% B where Buffer A was 0.1% formic acid and Buffer B was 0.1% formic acid in ACN. Flow rate was 0.18 mL min^{-1} and temperature was held constant at 25 °C. Compounds eluting from the column were analyzed with a 6560 Ion Mobility QTOF mass spectrometer (Agilent Technologies) with the following settings: gas temp, 325 °C; drying gas, 12 L min^{-1}; nebulizer 55 psig; sheath gas temp 350 °C; sheath gas flow 10 L min^{-1}; capillary voltage 4000 V; nozzle voltage 500 V; fragmentor 400 V. Retention times of (*S*)- and (*R*)-norcoclaurine were established by comparison to an (*S*)-norcoclaurine authentic standard (Toronto Research Chemicals Inc.) and *rac*-norcoclaurine, which was spontaneously condensed from dopamine (Sigma) and 4-HPAA (Toronto Research Chemicals) in a KH_2PO_4/ACN buffer (7). Retention times of (*S*)- and (*R*)-armepavine were established by comparison to authentic (*R*,*S*)-armepavine standard and (*S*)-armepavine synthesized by a BY4741 yeast strain harboring *Ps*6OMT, *Ps*CNMT, and *Nn*OMT5 (strain LP626) fed with (*S*)-norcoclaurine.

Statistical analyses. All numerical values are depicted as means ± s.d. Statistical differences between control and engineered strains were assessed via two-tailed Student's *t*-tests assuming equal variances using Excel (Microsoft) or Prism (GraphPad Software Inc.). In all cases, *P*-values < 0.05 were considered significant.

FIGURE LEGENDS

Figure 1. *Nn*CYP80G is a stereospecific proaporphine synthase. a, Proposed natural and artificial routes to armepavine in sacred lotus and engineered yeast, respectively. Lotus possesses five NCS isoforms for synthesis of (*R*)-norcoclaurine, which is subsequently methylated to (*R*)-armepavine. Yeast strains utilized in this study employ a truncated NCS variant from *C. japonica* (*Cj*NCSΔN), which yields exclusively (*S*)-norcoclaurine. *Ps*6OMT, *Ps*CNMT, and *Nn*OMT5 (3× MTase) involved in the conversion of (*S*)-norcoclaurine to (S)-armepavine are boxed. **b,** Analysis of BIA production by engineered strains. Introduction of *Nn*OMT5 to strain LP603 enables high titer production of armepavine. Peak area is shown for all pathway intermediates and absolute titers are depicted in mg/L for (*S*)-norcoclaurine and (*S*)-armepavine. Asterisk (*) denotes a significant increase or decrease ($P < 0.05$) in ion abundance relative to the parent strain. Error bars represent the mean ± s.d. of $n = 3$ independent biological samples. Statistical differences between control and derivative strains were tested using two-tailed Student's *t*-test. **c,** Chiral analysis of norcoclaurine and armepavine synthesized *de novo* or from norcoclaurine-supplemented strains of yeast. Yeast-derived norcoclaurine was compared against authentic (*S*)-norcoclaurine standard and (*R*,*S*)-norcoclaurine generated by the phosphate-catalyzed condensation of dopamine and 4-HPAA. Commercial (*S*)-norcoclaurine standard contains a small proportion of (*R*)-norcoclaurine, as described previously (4). Yeast-derived armepavine was compared against authentic (*R*,*S*)-armepavine standard, sacred lotus extract, and (*S*)-armepavine generated from feeding (*S*)-norcoclaurine to a BY4741 strain harboring *Ps*6OMT, *Ps*CNMT, and *Nn*OMT5 (3× MTase). (*R*)-Conformers of norcoclaurine and armepavine are shaded. **d,** Ion-extracted LC–MS chromatograms of engineered yeast strains harboring a partial pronuciferine pathway (*Ps*6OMT-*Ps*CNMT-*Nn*OMT5-*Nn*CYP80G-*At*ATR2; LP626) fed with

(*R*,*S*)-norcoclaurine or (*S*)-norcoclaurine. LP626 is unable to synthesize BIAs *de novo* due to lack of an NCS enzyme and a dopamine biosynthesis pathway. Retention time of yeast-derived pronuciferine (dark blue) was compared against sacred lotus extract (light blue). **e,** MS/MS fragmentation of pronuciferine from sacred lotus extract and strain LP626 fed with (*R*,*S*)-norcoclaurine. Pronuciferine fragmentation is compared against two prior reports using authentic pronuciferine standard (8, 9).

Figure 2. Screening poppy dehydroreticuline synthase (DRS) and dehydroreticuline reductase (DRR) orthologs for *de novo* synthesis of (*R*)-armepavine and pronuciferine. a, Proposed *de novo* route to (*R*)-armepavine and pronuciferine via stereochemical inversion of (*S*)-*N*-methylcoclaurine and (*S*)-armepavine. (*S*)-Enantiomers are converted to corresponding 1,2-dehydro intermediates using DRS, which are subsequently reduced to (*R*)-conformers using DRR. **b,** Protein phylogenetic tree of DRS orthologs from various poppy species. Abbreviations: *Pr*, *Papaver rhoeas* (common poppy); *Pb*, *Papaver bracteatum* (Iranian poppy); *Ps*, *Papaver somniferum* (opium poppy); *Cm*, *Chelidonium majus* (greater celandine). **c,** Synthesis of 1,2-dehydro-*N*-methylcoclaurine and 1,2-dehydroarmepavine by poppy DRS variants. DRS-encoding genes were integrated into an (*S*)-armepavine-producing strain (LP609) and formation of 1,2-dehydro intermediates was assayed by LC-MS. Asterisk (*) denotes a significant increase or decrease ($P < 0.05$) in ion abundance relative to the precursor strain (LP609). **d,** Reduction of 1,2-dehydro-*N*-methylcoclaurine and 1,2-dehydroarmepavine by poppy DRR orthologs. Plasmid-encoded DRR genes were introduced to LP609 harboring *Pr*DRS, *Nn*CYP80G, and *At*ATR2, and consumption of 1,2-dehydro intermediates was assayed by LC-MS. *N*-Methylcoclaurine and armepavine are reported as combined (*R*)- and (*S*)-conformers. Error bars represent the

mean ± s.d. of $n = 3$ independent biological samples. Asterisk (*) denotes a significant increase or decrease ($P < 0.05$) in ion abundance relative to the precursor strain harboring *Pr*DRS, *Nn*CYP80G, and *At*ATR2. Statistical differences between control and derivative strains were tested using two-tailed Student's *t*-test. **e,** Chiral analysis of armepavine synthesized by strains harboring *Pr*DRS or *Pr*DRS+*Ps*DRR. Implementation of *Pr*DRS+*Ps*DRR in conjunction with *Nn*CYP80G+*At*ATR2 enables stereochemical inversion of (*S*)-armepavine and concurrent synthesis of pronuciferine. Pronuciferine was separated on a standard (non-chiral) C18 column.

Figure 3. Combinatorial CYP80-CPR-CYB5 library for improved pronuciferine synthesis and characterization of *Nn*CYP80A a, Design of a partial 24-member CYP80-CPR-CYB5 combinatorial strain library. Two CYP80G variants (*Nn*CYP80G and *Ls*CYP80G) and CYP80A from *N. nucifera* (*Nn*CYP80A) were paired with different combinations of plant cytochrome P450 reductases (CPRs) and cytochrome *b5* (CYB5s) orthologs. The 24-strain library was assayed with and without pHUM-PrDRS-PsDRR, which expresses *PrDRS* and *PsDRR* genes for stereochemical inversion of BIA metabolites. Abbreviations: At, *Arabidopsis thaliana*; Cc2, *Corydalis cheilanthifolia* CPR2; Cr, *Catharanthus roseus*; Ec, *Eschscholzia californica*; Hc, *Hypericum calycinum*; *Pb*, *Papaver bracteatum*; *Ps*, *P. somniferum*. **b,** Pronuciferine production by strains harboring *Nn*CYP80G or *Ls*CYP80G, pHUM-PrDRS-PsDRR, and different combinations of CPR and CYB5. Asterisk (*) denotes a significant increase or decrease ($P < 0.05$) in pronuciferine ion abundance between the designated strains. All strains exhibited a significant increase ($P < 0.05$) in pronuciferine ion abundance relative to the parent strain lacking a CYP80G, CPR, and CYB5. **c,** bis-BIA synthesis by strains harboring *Nn*CYP80A, pHUM-PrDRS-PsDRR, and different combinations of CPR and CYB5. Asterisk (*) denotes a significant

increase or decrease ($P < 0.05$) in bis-BIA ion abundance between the designated strains. All strains exhibited a significant increase ($P < 0.05$) in bis-BIA ion abundance relative to the parent strain lacking *Nn*CYP80A, CPR, and CYB5. **d,** Proposed pathways to lotus bis-BIA alkaloids in engineered yeast. *Nn*CYP80A catalyzes the stereospecific formation of nelumborine (C8-C3′) or nelumboferine (C7-*O*-C3′) from *N*-methylcoclaurine via distinct phenol coupling routes. Guattegaumerine biosynthesis in *Berberis stolonifera* is shown (boxed), which involves C3′-*O*-C4′ coupling of two units of (*R*)-*N*-methylcoclaurine by *Bs*CYP80A1. **e,** Ion-extracted LC–MS chromatograms of engineered yeast strains harboring *Nn*CYP80A or *Bs*CYP80A1, pHUM-PrDRS-PsDRR, and a CPR. The bis-BIA product from *Nn*CYP80A-producing strains was compared against commercial lotus extract. **f,** MS/MS fragmentation of a bis-BIA product (RT ~5.2 min) from sacred lotus extract and strain LP625 containing *Nn*CYP80A and *At*ATR2. **g,** MS/MS fragmentation of bis-BIA products (RT ~5.3 min) from strain LP625 (*Nn*CYP80A+*At*ATR2) and LP637 (*Bs*CYP80A1+*Hc*CPR). The 597 → 192 transition is the major product observed upon fragmentation of authentic guattegaumerine standard (17).

Figure 4. Screening diverse plant *O*-methyltransferases for activity on nelumboferine. a, Proposed neferine biosynthesis pathways in *N. nucifera*. Neferine is synthesized via convergent routes from liensinine or isoliensinine, both of which are presumed to arise from nelumboferine. **b,** Protein phylogenetic tree of 66 *O*-methyltransferase orthologs from various plant species. Abbreviations: *Am*, *Argemone mexicana*; *Cc*, *Corydalis cheilanthifolia*; *Cj*, *Coptis japonica*; *Cm*, *Chelidonium majus*; *Cs*, *Corydalis saxicola*; *Ec*, *Eschscholzia californica*; *Gf*, *Glaucium flavum*; *Hc*, *Hypericum calycinum*; *Nd*, *Nandina domestica*; *Nn*, *Nelumbo nucifera*; *Ns*, *Nigella sativa*; *Pb*, *Papaver bracteatum*; *Ps*, *Papaver somniferum*; *Sc*, *Sanguinaria canadensis*; *Sd*, *Stylophorum*

diphyllum; *Tc*, *Tinospora cordifolia*; *Tf*, *Thalictrum flavum*; *Vh*, *Vancouveria hexandra*; Xs, *Xanthoriza simplicissima.* **c,** Ion-extracted LC–MS chromatograms of engineered yeast strains harboring *Nn*CYP80A, *At*ATR2, *Pr*DRS-*Ps*DRR, and *Ec*OMT1 from *Eschscholzia californica*. Retention time of yeast-derived liensinine (dark blue) was compared against authentic liensinine standard (black) and sacred lotus extract (green). **d,** MS/MS fragmentation of authentic liensinine standard and a yeast strain engineered for liensinine synthesis (*Nn*CYP80A, ATR2, *Pr*DRS, *Ps*DRR, and *Ec*OMT1). **e,** MS/MS fragmentation of tentative liensinine (RT = 2.16 min; dark green) and isoliensinine (RT = 2.25 min; light green) products from commercial lotus extract. Liensinine (611 → 206) and isoliensinine (611 → 192) yield different major fragment ions based on a prior study (21). **Figure 1**

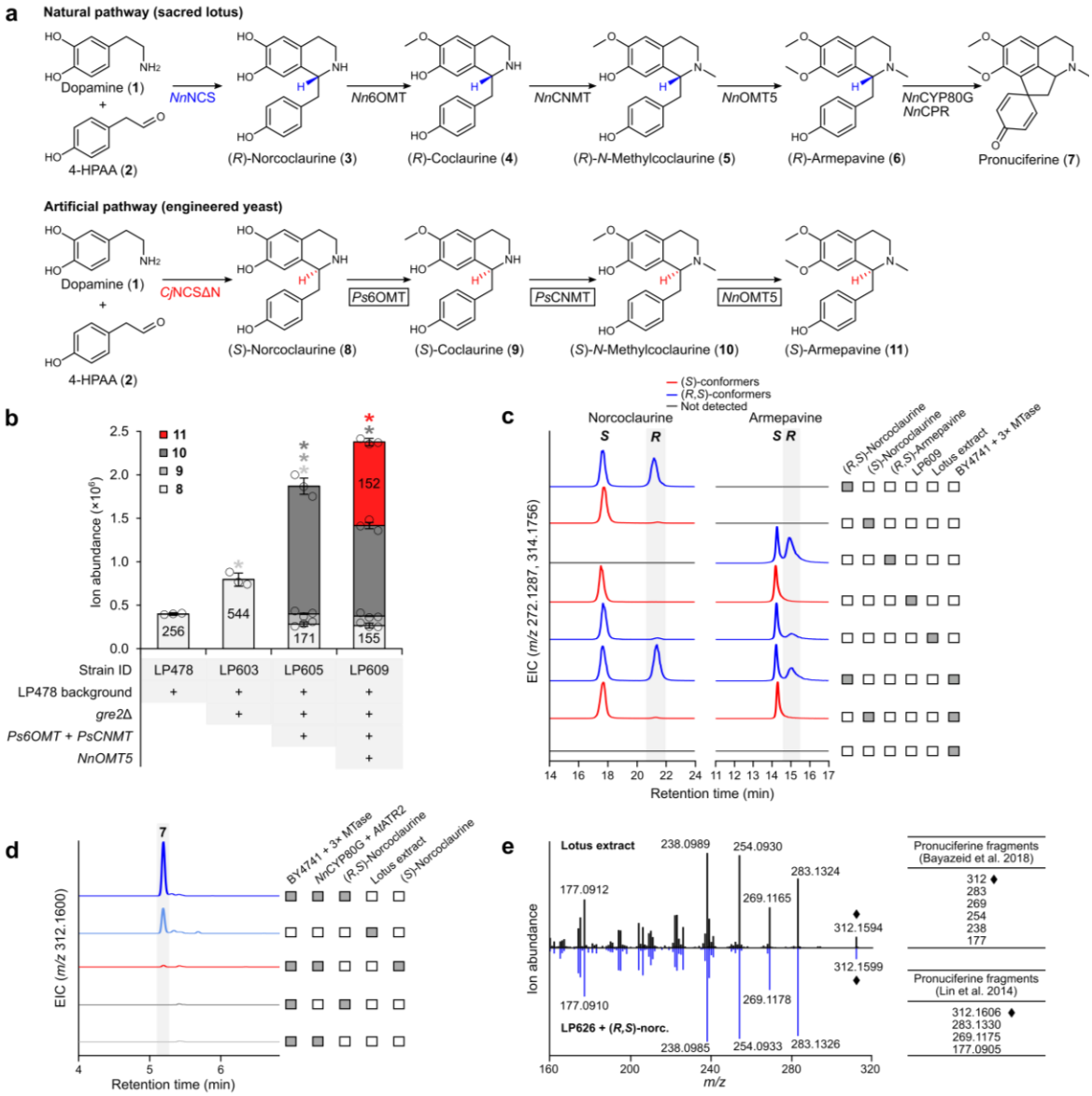

a
Natural pathway (sacred lotus)
Dopamine (1)
4-HPAA (2)
NnNCS
(R)-Norcoclaurine (3)
Nn6OMT
(R)-Coclaurine (4)
NnCNMT
(R)-N-Methylcoclaurine (5)
NnOMT5
(R)-Armepavine (6)
NnCYP80G
NnCPR
Pronuciferine (7)
Artificial pathway (engineered yeast)
CjNCSΔN
(S)-Norcoclaurine (8)
Ps6OMT
(S)-Coclaurine (9)
PsCNMT
(S)-N-Methylcoclaurine (10)
NnOMT5
(S)-Armepavine (11)
b
Ion abundance (×10^6)
11
10
9
8
256
544
171
155
152
Strain ID
LP478
LP603
LP605
LP609
LP478 background
gre2Δ
Ps6OMT + PsCNMT
NnOMT5
c
(S)-conformers
(R,S)-conformers
Not detected
Norcoclaurine
Armepavine
EIC (m/z 272.1287, 314.1756)
Retention time (min)
(R,S)-Norcoclaurine
(S)-Norcoclaurine
(R,S)-Armepavine
LP609
Lotus extract
BY4741 + 3× MTase
d
EIC (m/z 312.1600)
Retention time (min)
BY4741 + 3× MTase
NnCYP80G + AtATR2
(R,S)-Norcoclaurine
Lotus extract
(S)-Norcoclaurine
e
Lotus extract
LP626 + (R,S)-norc.
Ion abundance
m/z
238.0989
254.0930
283.1324
177.0912
269.1165
312.1594
312.1599
177.0910
269.1178
238.0985
254.0933
283.1326
Pronuciferine fragments (Bayazeid et al. 2018)
312
283
269
254
238
177
Pronuciferine fragments (Lin et al. 2014)
312.1606
283.1330
269.1175
177.0905

Figure 2

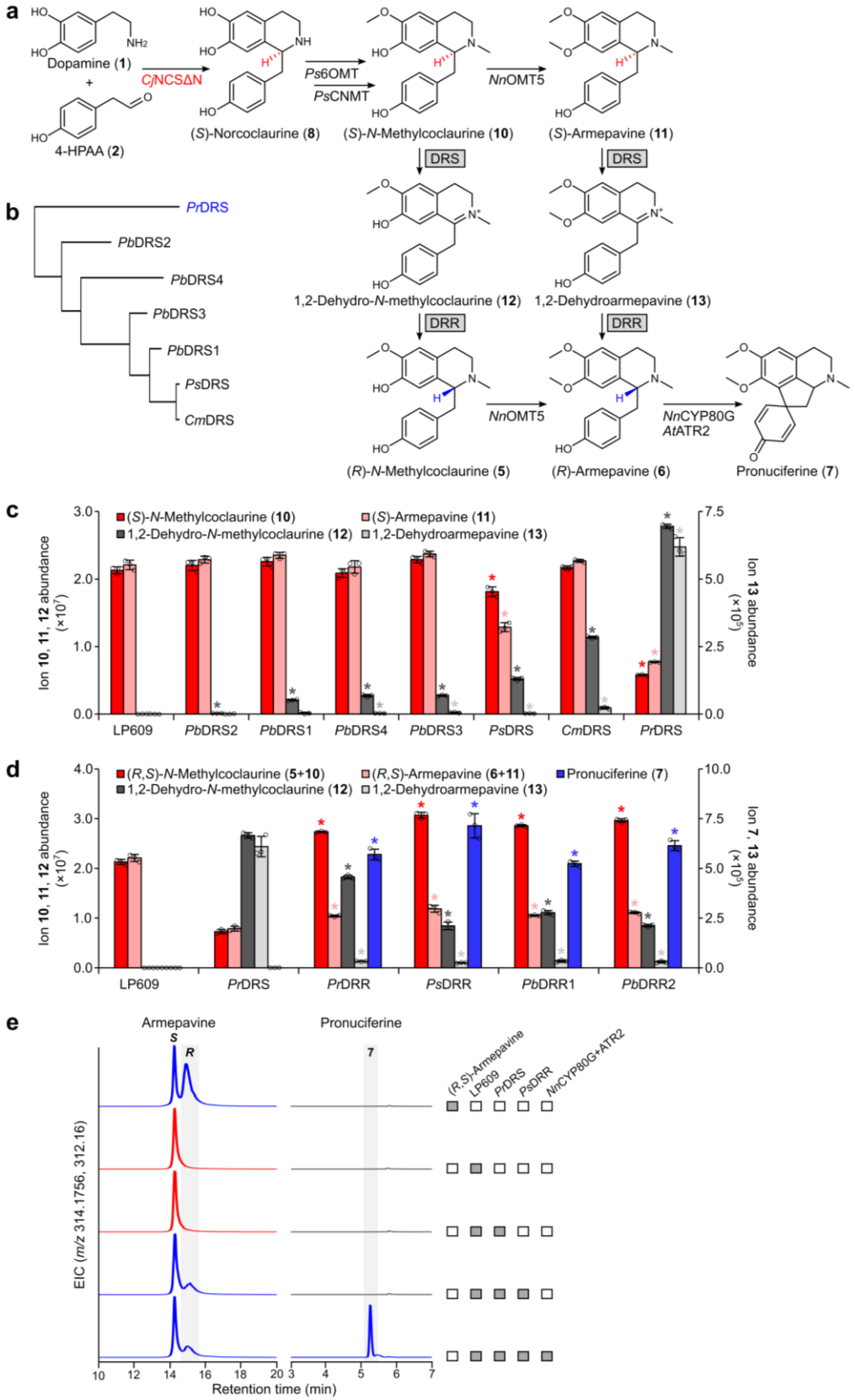

a
Dopamine (1)
4-HPAA (2)
CjNCSΔN
(S)-Norcoclaurine (8)
Ps6OMT
PsCNMT
(S)-N-Methylcoclaurine (10)
NnOMT5
(S)-Armepavine (11)
DRS
1,2-Dehydro-N-methylcoclaurine (12)
1,2-Dehydroarmepavine (13)
DRR
(R)-N-Methylcoclaurine (5)
(R)-Armepavine (6)
NnCYP80G
AtATR2
Pronuciferine (7)
b
PrDRS
PbDRS2
PbDRS4
PbDRS3
PbDRS1
PsDRS
CmDRS
c
(S)-N-Methylcoclaurine (10)
(S)-Armepavine (11)
1,2-Dehydro-N-methylcoclaurine (12)
1,2-Dehydroarmepavine (13)
Ion 10, 11, 12 abundance (×10^7)
Ion 13 abundance (×10^5)
LP609
PbDRS2
PbDRS1
PbDRS4
PbDRS3
PsDRS
CmDRS
PrDRS
d
(R,S)-N-Methylcoclaurine (5+10)
(R,S)-Armepavine (6+11)
Pronuciferine (7)
1,2-Dehydro-N-methylcoclaurine (12)
1,2-Dehydroarmepavine (13)
Ion 10, 11, 12 abundance (×10^7)
Ion 7, 13 abundance (×10^5)
LP609
PrDRS
PrDRR
PsDRR
PbDRR1
PbDRR2
e
Armepavine
Pronuciferine
S
R
7
(R,S)-Armepavine
LP609
PrDRS
PsDRR
NnCYP80G+ATR2
EIC (m/z 314.1756, 312.16)
Retention time (min)

Figure 3

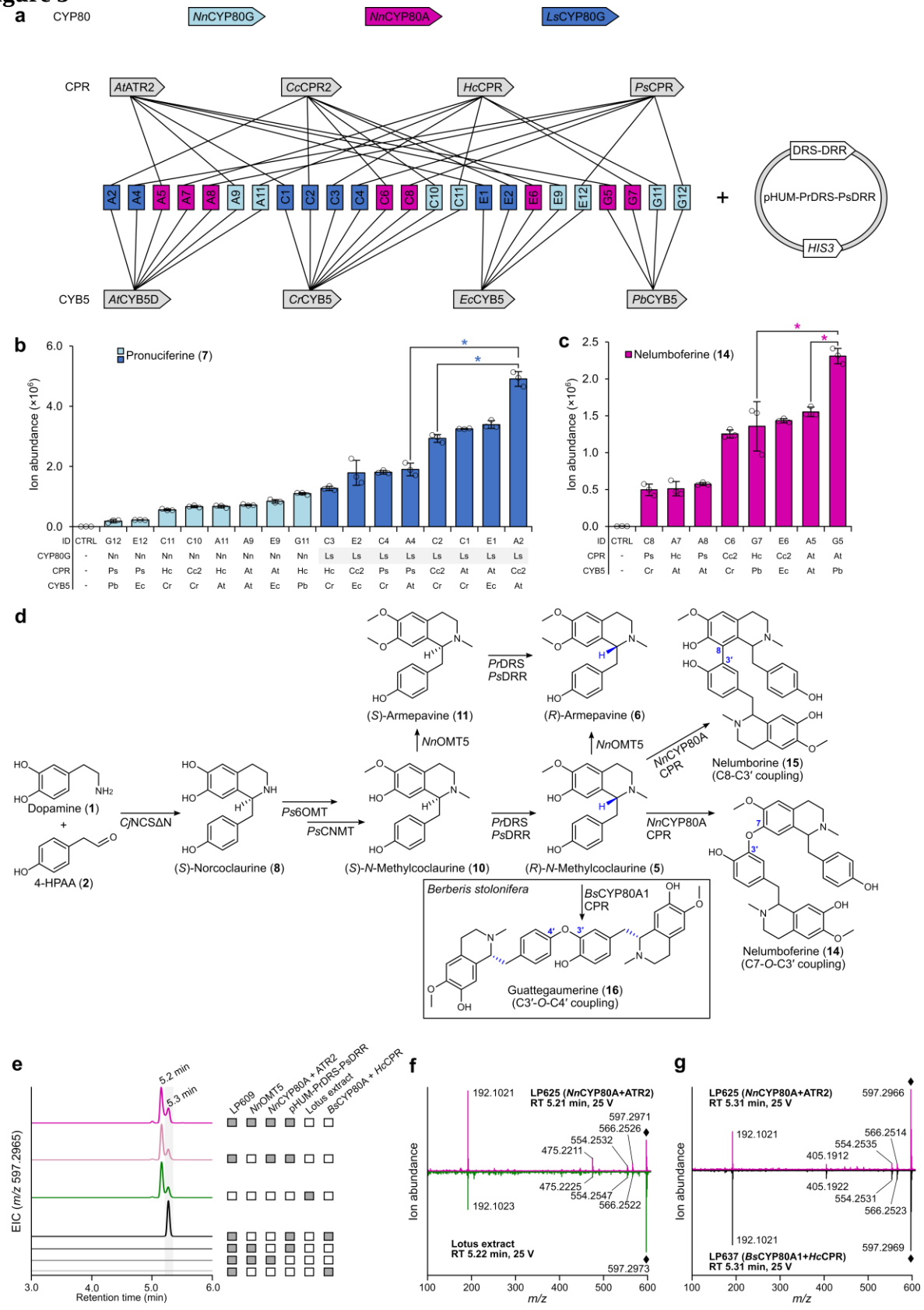
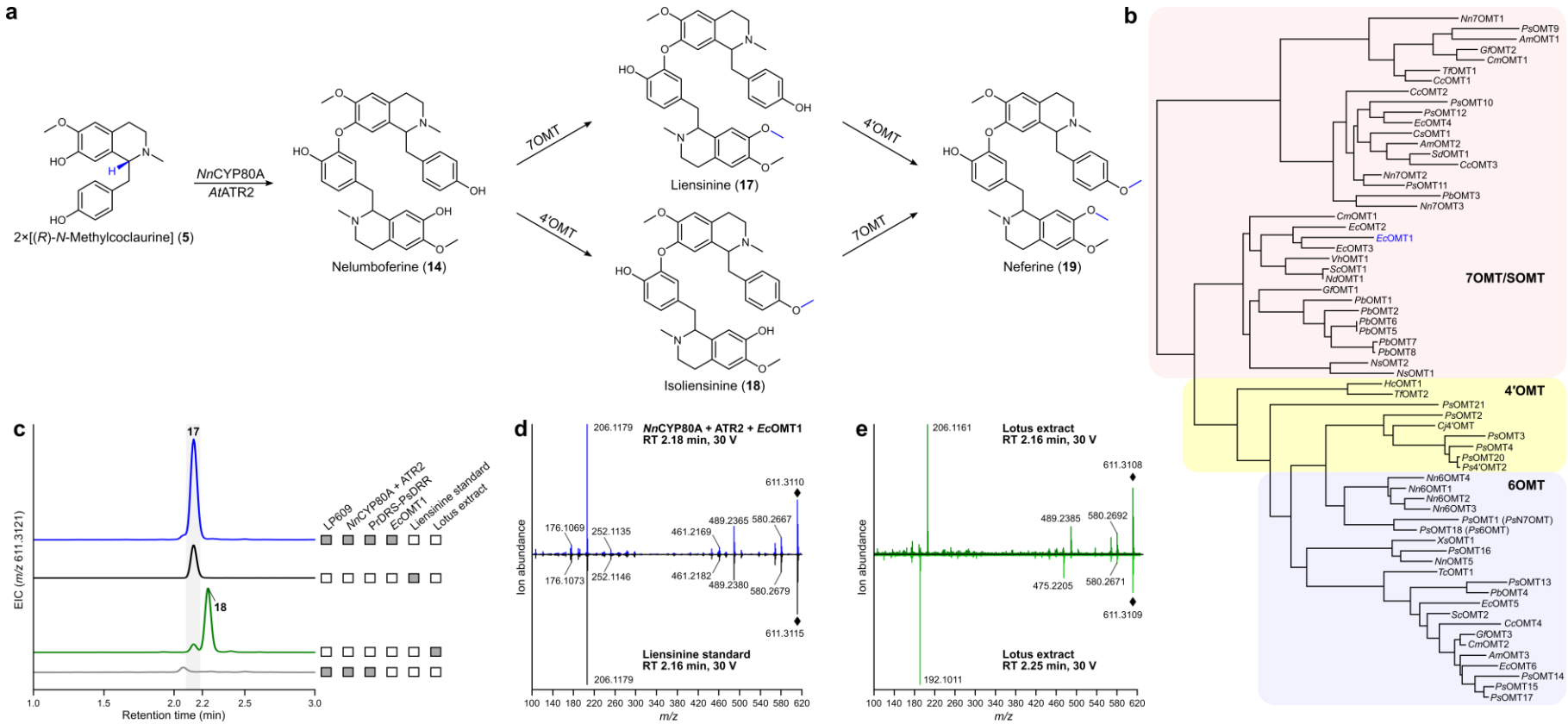

a
CYP80
NnCYP80G
NnCYP80A
LsCYP80G
CPR
AtATR2
CcCPR2
HcCPR
PsCPR
A2 A4 A5 A7 A8 A9 A11 C1 C2 C3 C4 C6 C8 C10 C11 E1 E2 E6 E9 E12 G5 G7 G11 G12
+
DRS-DRR
pHUM-PrDRS-PsDRR
HIS3
CYB5
AtCYB5D
CrCYB5
EcCYB5
PbCYB5
b
Pronuciferine (7)
Ion abundance (×10^6)
ID CTRL G12 E12 C11 C10 A11 A9 E9 G11 C3 E2 C4 A4 C2 C1 E1 A2
CYP80G - Nn Nn Nn Nn Nn Nn Nn Ls Ls Ls Ls Ls Ls Ls Ls
CPR - Ps Ps Hc Cc2 Hc At At Hc Hc Cc2 Ps Ps Cc2 At At Cc2
CYB5 - Pb Ec Cr Cr At At Ec Pb Cr Ec Cr At Cr Cr Ec At
c
Nelumboferine (14)
Ion abundance (×10^6)
ID CTRL C8 A7 A8 C6 G7 E6 A5 G5
CPR - Ps Hc Ps Cc2 Hc Cc2 At At
CYB5 - Cr At At Cr Pb Ec At Pb
d
(S)-Armepavine (11)
PrDRS
PsDRR
(R)-Armepavine (6)
NnOMT5
NnOMT5
NnCYP80A
CPR
Nelumborine (15)
(C8-C3′ coupling)
Dopamine (1)
+
4-HPAA (2)
CjNCSΔN
(S)-Norcoclaurine (8)
Ps6OMT
PsCNMT
(S)-N-Methylcoclaurine (10)
PrDRS
PsDRR
(R)-N-Methylcoclaurine (5)
NnCYP80A
CPR
Nelumboferine (14)
(C7-O-C3′ coupling)
Berberis stolonifera
BsCYP80A1
CPR
Guattegaumerine (16)
(C3′-O-C4′ coupling)
e
5.2 min
5.3 min
EIC (m/z 597.2965)
Retention time (min)
LP609
NnOMT5
NnCYP80A + ATR2
pHUM-PrDRS-PsDRR
Lotus extract
BsCYP80A + HcCPR
f
LP625 (NnCYP80A+ATR2)
RT 5.21 min, 25 V
192.1021
597.2971
566.2526
554.2532
475.2211
475.2225
554.2547
566.2522
192.1023
Lotus extract
RT 5.22 min, 25 V
597.2973
Ion abundance
m/z
g
LP625 (NnCYP80A+ATR2)
RT 5.31 min, 25 V
597.2966
192.1021
566.2514
554.2535
405.1912
405.1922
554.2531
566.2523
192.1021
597.2969
LP637 (BsCYP80A1+HcCPR)
RT 5.31 min, 25 V
Ion abundance
m/z

Figure 4


a
2×[(R)-N-Methylcoclaurine] (5)
NnCYP80A
AtATR2
Nelumboferine (14)
7OMT
4'OMT
Liensinine (17)
Isoliensinine (18)
4'OMT
7OMT
Neferine (19)
b
7OMT/SOMT
4'OMT
6OMT
c
EIC (m/z 611.3121)
Retention time (min)
LP609
NnCYP80A + ATR2
PrDRS,PsDRR
EcOMT1
Liensinine standard
Lotus extract
d
NnCYP80A + ATR2 + EcOMT1
RT 2.18 min, 30 V
Liensinine standard
RT 2.16 min, 30 V
Ion abundance
m/z
e
Lotus extract
RT 2.16 min, 30 V
Lotus extract
RT 2.25 min, 30 V
Ion abundance
m/z

REFERENCES

1. Menéndez-Perdomo IM & Facchini PJ (2018) Benzylisoquinoline alkaloids biosynthesis in sacred lotus. *Molecules* 23(11):2899.
2. Yang G-M, *et al.* (2018) Isolation and identification of a tribenzylisoquinoline alkaloid from *Nelumbo nucifera* Gaertn, a novel potential smooth muscle relaxant. *Fitoterapia* 124:58-65.
3. He C-L, *et al.* (2021) Identification of bis-benzylisoquinoline alkaloids as SARS-CoV-2 entry inhibitors from a library of natural products. *Signal transduction and targeted therapy* 6(1):1-3.
4. Pyne ME, *et al.* (2020) A yeast platform for high-level synthesis of tetrahydroisoquinoline alkaloids. *Nat. Commun.* 11(1):1-10.
5. Kashiwada Y, *et al.* (2005) Anti-HIV benzylisoquinoline alkaloids and flavonoids from the leaves of *Nelumbo nucifera*, and structure–activity correlations with related alkaloids. *Biorg. Med. Chem.* 13(2):443-448.
6. Menéndez-Perdomo IM & Facchini PJ (2020) Isolation and characterization of two *O*-methyltransferases involved in benzylisoquinoline alkaloid biosynthesis in sacred lotus (*Nelumbo nucifera*). *J. Biol. Chem.* 295(6):1598-1612.
7. Pesnot T, Gershater MC, Ward JM, & Hailes HC (2012) The catalytic potential of *Coptis japonica* NCS2 revealed–Development and utilisation of a fluorescamine-based assay. *Adv. Synth. Catal.* 354(16):2997-3008.
8. Bayazeid O, *et al.* (2018) An LC-ESI-MS/MS method for the simultaneous determination of pronuciferine and roemerine in some *Papaver* species. *J. Chromatogr. B* 1096:223-227.
9. Lin Z, Yang R, Guan Z, Chen A, & Li W (2014) Ultra-performance LC separation and quadrupole time-of-flight MS identification of major alkaloids in Plumula Nelumbinis. *Phytochem. Anal.* 25(6):485-494.
10. Yang M, *et al.* (2017) Digital gene expression analysis provides insight into the transcript profile of the genes involved in aporphine alkaloid biosynthesis in lotus (*Nelumbo nucifera*). *Frontiers in plant science* 8:80.
11. Bourgeois L, Pyne ME, & Martin VJ (2018) A highly characterized synthetic landing pad system for precise multicopy gene integration in yeast. *ACS synthetic biology* 7(11):2675-2685.
12. Grewal PS, Samson JA, Baker JJ, Choi B, & Dueber JE (2021) Peroxisome compartmentalization of a toxic enzyme improves alkaloid production. *Nat. Chem. Biol.* 17(1):96-103.
13. Pyne ME & Martin VJ (2022) Microbial synthesis of natural, semisynthetic, and new-to-nature tetrahydroisoquinoline alkaloids. *Current Opinion in Green and Sustainable Chemistry* 33:100561.
14. Farrow SC, Hagel JM, Beaudoin GA, Burns DC, & Facchini PJ (2015) Stereochemical inversion of (*S*)-reticuline by a cytochrome P450 fusion in opium poppy. *Nat. Chem. Biol.* 11(9):728-732.
15. Galanie S, Thodey K, Trenchard IJ, Interrante MF, & Smolke CD (2015) Complete biosynthesis of opioids in yeast. *Sci* 349(6252):1095-1100.
16. Winzer T, *et al.* (2015) Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. *Sci* 349(6245):309-312.

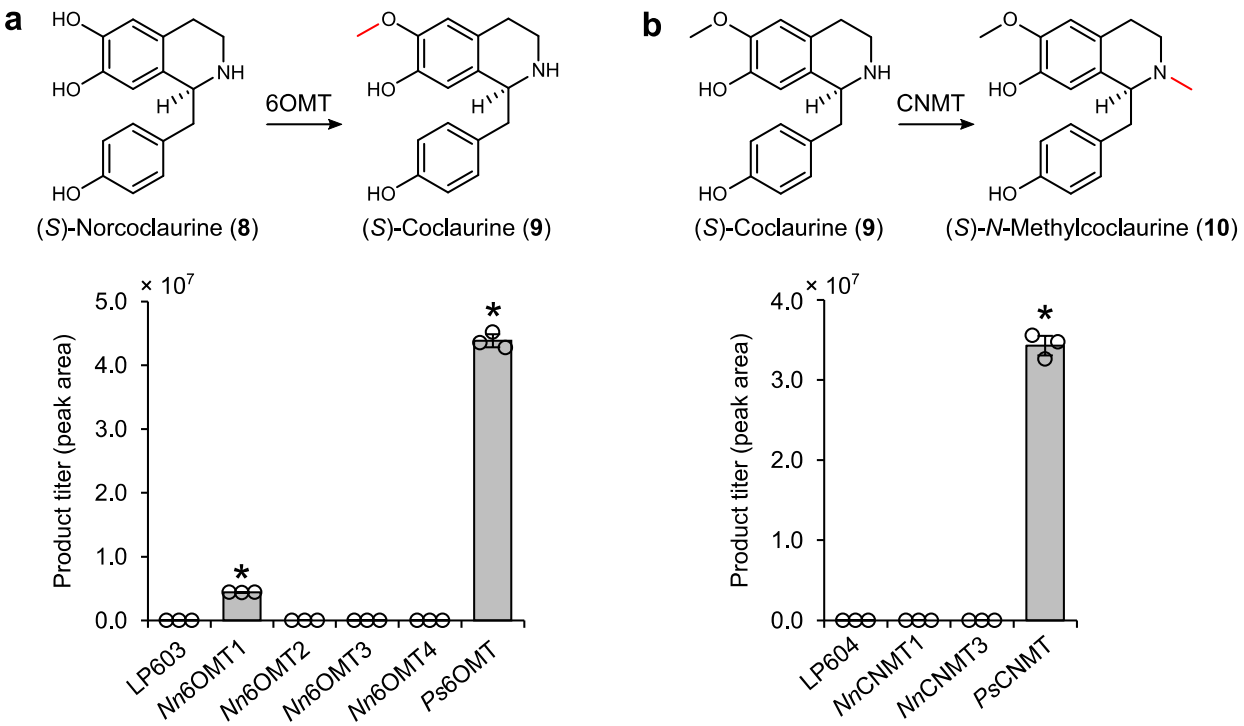
17. Payne JT, Valentic TR, & Smolke CD (2021) Complete biosynthesis of the bisbenzylisoquinoline alkaloids guattegaumerine and berbamunine in yeast. *Proceedings of the National Academy of Sciences* 118(51):e2112520118.
18. Itoh A, *et al.* (2011) Bisbenzylisoquinoline alkaloids from *Nelumbo nucifera*. *Chem. Pharm. Bull.* 59(8):947-951.
19. Fujii N, Inui T, Iwasa K, Morishige T, & Sato F (2007) Knockdown of berberine bridge enzyme by RNAi accumulates (*S*)-reticuline and activates a silent pathway in cultured California poppy cells. *Transgenic Res.* 16(3):363-375.
20. Purwanto R, Hori K, Yamada Y, & Sato F (2017) Unraveling additional *O*-methylation steps in benzylisoquinoline alkaloid biosynthesis in California poppy (*Eschscholzia californica*). *PCPhy* 58(9):1528-1540.
21. Zhou H, Jiang H, Yao T, & Zeng S (2007) Fragmentation study on the phenolic alkaloid neferine and its analogues with anti-HIV activities by electrospray ionization tandem mass spectrometry with hydrogen/deuterium exchange and its application for rapid identification of in vitro microsomal metabolites of neferine. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry* 21(13):2120-2128.
22. Hong B, *et al.* (2022) Biosynthesis of strychnine. *Nature*:1-6.
23. Nett RS, Lau W, & Sattely ES (2020) Discovery and engineering of colchicine alkaloid biosynthesis. *Nature* 584(7819):148-153.
24. Nishimura K, Horii S, Tanahashi T, Sugimoto Y, & Yamada J (2012) Synthesis and pharmacological activity of alkaloids from embryo of lotus, *Nelumbo nucifera*. *Chem. Pharm. Bull.*:c12-00820.
25. Zhou J, *et al.* (2015) A novel autophagy/mitophagy inhibitor liensinine sensitizes breast cancer cells to chemotherapy through DNM1L-mediated mitochondrial fission. *Autophagy* 11(8):1259-1279.
26. Li H, *et al.* (2021) Nelumbo genome database, an integrative resource for gene expression and variants of *Nelumbo nucifera*. *Scientific data* 8(1):1-7.
27. Horwitz AA, *et al.* (2015) Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. *Cell Systems*:88-96.
28. Ryan OW, *et al.* (2014) Selection of chromosomal DNA libraries using a multiplex CRISPR system. *Elife* 3:e03703.
29. Shao Z, Zhao H, & Zhao H (2009) DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* 37(2):e16-e16.
30. Bean BDM, Whiteway M, & Martin VJJ (2022) The MyLO CRISPR-Cas9 toolkit: a markerless yeast localization and overexpression CRISPR-Cas9 toolkit. *G3 Genes|Genomes|Genetics*.
31. Labuhn M, *et al.* (2017) Refined sgRNA efficacy prediction improves large-and small-scale CRISPR–Cas9 applications. *Nucleic Acids Res.* 46(3):1375-1385.
32. Smith JD, *et al.* (2016) Quantitative CRISPR interference screens in yeast identify chemical-genetic interactions and new rules for guide RNA design. *Genome biology* 17(1):45.
33. Bai Flagfeldt D, Siewers V, Huang L, & Nielsen J (2009) Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. *Yeast* 26(10):545-551.

34. Mikkelsen MD, *et al.* (2012) Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab. Eng.* 14(2):104-111.
35. Reider Apel A, *et al.* (2016) A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 45(1):496-508.

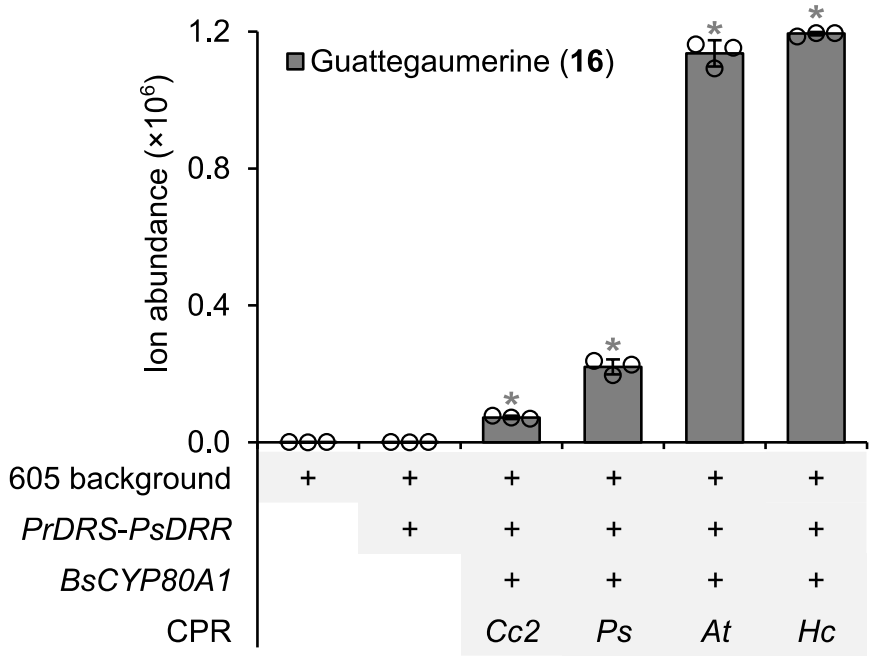
SUPPLEMENTARY INFORMATION

Pyne and Martin

Supplementary Results



Supplementary Figure 1. Comparison of core BIA pathway methyltransferases from opium poppy (*Papaver somniferum*) and sacred lotus (*Nelumbo nucifera*). (**a**) Screening 6OMT variants for methylation of (*S*)-norcoclaurine to (*S*)-coclaurine. 6OMT isoforms from *N. nucifera* (*Nn*6OMT1-*Nn*6OMT4) were compared against *Ps*6OMT from *P. somniferum* for production of (*S*)-coclaurine in strain LP603. (**b**) Screening CNMT variants for methylation of (*S*)-coclaurine to (*S*)-*N*-methylcoclaurine. CNMT isoforms from *N. nucifera* (*Nn*CNMT1 and *Nn*CNMT3) were compared against *Ps*CNMT from *P. somniferum* for production of (*S*)-*N*-methylcoclaurine in strain LP604 expressing *Ps6OMT*. Asterisk (*) denotes a significant increase ($P < 0.05$) in titer relative to the parent strain. Statistical differences between control and derivative strains were tested using two-tailed Student's *t*-test. Error bars represent s.d. of three biological replicates.



Supplementary Figure 2. Comparison of guattegaumerine production by strains harboring *Bs*CYP80A1 and different plant CPR variants. Asterisk (*) denotes a significant increase or decrease ($P < 0.05$) in ion abundance relative to the precursor strain. Error bars represent the mean ± s.d. of $n = 3$ independent biological samples. Statistical differences between control and derivative strains were tested using two-tailed Student's *t*-test. Abbreviations: *At*, *Arabidopsis thaliana*; Cc2, *Corydalis cheilanthifolia* CPR2; Hc, *Hypericum calycinum*; *Ps*, *P. somniferum*.

Supplementary Table 1 – Tentative activity assignments of diverse plant *O*-methyltransferases on 1-benzylisoquinolines

OMT	Source	Tentative activity		
		6OMT	7OMT	4′OMT
*Am*OMT1	*Argemone mexicana*		+	
*Am*OMT2				
*Am*OMT3				
*Cc*OMT1	*Corydalis cheilanthifolia*			
*Cc*OMT2				
*Cc*OMT3				
*Cc*OMT4				
*Cj*4′OMT	*Coptis japonica*			
*Cm*OMT1	*Chelidonium majus*			
*Cm*OMT2				
*Cm*OMT1				
*Cs*OMT1	*Corydalis saxicola*			
*Ec*OMT1	*Eschscholzia californica*		+	
*Ec*OMT2 (7OMT/SOMT)			+	
*Ec*OMT3				
*Ec*OMT4				
*Ec*OMT5			+	
*Ec*OMT6				
*Gf*OMT1	*Glaucium flavum*			
*Gf*OMT2				
*Gf*OMT3				
*Hc*OMT1	*Hypericum calycinum*			
*Nd*OMT1	*Nandina domestica*			
*Nn*6OMT1	*Nelumbo nucifera*			
*Nn*6OMT2				
*Nn*6OMT3				
*Nn*6OMT4				
*Nn*7OMT1				
*Nn*7OMT2				
*Nn*7OMT3				
*Nn*OMT5			+	
*Ns*OMT1	*Nigella sativa*			
*Ns*OMT2				
*Pb*OMT1	*Papaver bracteatum*			
*Pb*OMT2				
*Pb*OMT3				
*Pb*OMT4				
*Ps*OMT1 (*Ps*N7OMT)	*Papaver somniferum*		+	

*Ps*OMT2				+
*Ps*OMT3				
*Ps*OMT4				
*Ps*OMT5				
*Ps*OMT6				
*Ps*OMT7				
*Ps*OMT8				
*Ps*OMT9				
*Ps*OMT10				
*Ps*OMT11				
*Ps*OMT12				
*Ps*OMT13				
*Ps*OMT14				+
*Ps*OMT15				
*Ps*OMT16			+	
*Ps*OMT17		+		
*Ps*OMT18 (*Ps*6OMT)		+		
*Ps*OMT19 (*Ps*4′OMT2)				+
*Ps*OMT20				+
*Ps*OMT21				
*Sc*OMT1	*Sanguinaria canadensis*			
*Sc*OMT2				
*Sd*OMT1	*Stylophorum diphyllum*			
*Tc*OMT1	*Tinospora cordifolia*			
*Tf*OMT1	*Thalictrum flavum*			
*Tf*OMT2				
*Vha*OMT1				
*Xs*OMT1	*Xanthoriza simplicissima*			

Supplementary Table 2 – Plasmids utilized in this study

Plasmid	Description	Source or reference
pBOT-HIS	CEN6/ARS4^{ori}, pMB1^{ori}, Amp^{R} , Kan^{R} , *HIS3*, P_{TEF1}-GFP-T_{PGI1}	[1]
pCAS-G418	P_{RNR2}-*cas9*$_{NLS}$-T_{CYC1}, pUC, 2μ, P_{tRNA_Tyr}-3′HDV-gRNA-Scaffold-T_{SNR52}, P_{TEF1}-*kanMX*-T_{TEF1}	[2]
pCAS-Hyg	P_{RNR2}-*cas9*$_{NLS}$-T_{CYC1}, pUC, 2μ, P_{tRNA_Tyr}-3′HDV-gRNA-Scaffold-T_{SNR52}, P_{TEF1}-*HphNTI*-T_{TEF1}	[3]
pPSG450	CEN6/ARS4^{ori}, ColE1, Kan^{R}, *HIS3*, P_{TEF1}-*EcNMCH*-T_{TDH3}, P_{TDH3}-*Ps6OMT*-T_{ADH1}, P_{PGK1}-*Ps4′OMT2*-T_{ENO1}, P_{TEF2}-*PsCNMT*-T_{SSA1}, P_{HHF1}-*AtATR2*-T_{ENO2},	[3]
pHUM	CEN6/ARS4^{ori}, ColE1, Amp^{R}, *HIS3*, *URA3*, *MET17*	[4]
pHUM-PrDRR	pHUM backbone, P_{TDH3}-*Pr*DRR-T_{EFM1}	This study
pHUM-PsDRR	pHUM backbone, P_{TDH3}-*Ps*DRR-T_{EFM1}	This study
pHUM-PbDRR1	pHUM backbone, P_{TDH3}-*Pb*DRR1-T_{EFM1}	This study
pHUM-PbDRR2	pHUM backbone, P_{TDH3}-*Pb*DRR2-T_{EFM1}	This study
pHUM-PrDRS-PsDRR	pHUM backbone, P_{TDH3}-*Pr*DRS-T_{IDP1}, P_{CCW12}-*Ps*DRR-T_{EFM1}	This study

Supplementary Table 3 – Strains utilized in this study

ID	Description	Parent	Locus+manipulation	Reference
LP478	(*S*)-Norcoclaurine strain	NA	NA	[3]
LP603	(*S*)-Norcoclaurine strain	LP478	*gre2*Δ	This study
LB604	(*S*)-Coclaurine strain	LB603	USERX-3-*Ps6OMT*-USERX-3	This study
LP605	(*S*)-*N*-Methylcoclaurine strain	LP603	USERX-3-*Ps6OMT*-*PsCNMT*-USERX-3	This study
LP609	(*S*)-Armepavine strain	LP605	USERXI-2-*NnOMT5*-USERXI-2	This study
LP624	*De novo* A9 pronuciferine assembly	LP609	USERX-1-*NnCYP80G*-*AtATR2*-*AtCYB5D*-USERX-1	This study
LP626	BY4741 C1 pronuciferine assembly	BY4741	USERX-1-*NnCYP80G*-*AtATR2*-*CrCYB5*-USERX-1	This study
LP625	*De novo* G5 pronuciferine assembly	LP609	USERX-1-*NnCYP80A*-*AtATR2*-*PbCYB5*-USERX-1	This study
LP615	BY4741 3× MTase strain	BY4741	USERX-3-*Ps6OMT*-*PsCNMT*-USERX-3 USERXI-2-*NnOMT5*-USERXI-2	This study
LP623	*De novo* A2 pronuciferine assembly	LP609	USERX-1-*LsCYP80A*-*CcCPR2*-*AtCYB5D*-USERX-1	This study
LP634	Chromosomal *PrDRS*-*PsDRR* in A2 pronuciferine assembly	LP623	416d-*PrDRS*-*PsDRR*-416d	This study
LP631	Deletion of *NnOMT5* from G5 *Nn*CYP80A assembly	LP625	Δ*NnOMT5*	This study
LP636	Chromosomal *PrDRS*-*PsDRR* in G5 *Nn*CYP80A Δ*NnOMT5* assembly	LP631	416d-*PrDRS*-*PsDRR*-416d	This study
LP638	Chromosomal *PrDRS*-*PsDRR* in (*S*)-*N*-Methylcoclaurine strain	LP605	416d-*PrDRS*-*PsDRR*-416d	This study
LP637	*Bs*CYP80A1+*Hc*CPR host	LP638	USERXI-2-*BsCYP80A1*-*HcCPR*-USERXI-2	This study

Supplementary Table 4 – Expression cassettes utilized in this study

Description	Cassette	Locus
Ps6OMT + PsCNMT	P_{TEF1}-*Ps6OMT*-T_{VMA2}-LTP1-P_{CCW12}-*PsCNMT*-T_{EFM1t}	USERX-3
NnOMT5	P_{TDH3}-*NnOMT5*-T_{TDH1}	USERXI-2
Nn6OMT1	P_{TDH3}-*Nn6OMT1*-T_{TDH1}	106a
Nn6OMT2	P_{TDH3}-*Nn6OMT2*-T_{TDH1}	106a
Nn6OMT3	P_{TDH3}-*Nn6OMT3*-T_{TDH1}	106a
Nn6OMT4	P_{TDH3}-*Nn6OMT4*-T_{TDH1}	106a
Ps6OMT	P_{TDH3}-*Ps6OMT*-T_{TDH1}	106a
NnCNMT1	P_{TDH3}-*NnCNMT1*-T_{TDH1}	106a
NnCNMT3	P_{TDH3}-*NnCNMT3*-T_{TDH1}	106a
PsCNMT	P_{TDH3}-*PsCNMT*-T_{TDH1}	106a
CmDRS	P_{TDH3}-*CmDRS*-T_{IDP1}	FgF 21
PbDRS1	P_{TDH3}-*PbDRS1*-T_{IDP1}	FgF 21
PbDRS2	P_{TDH3}-*PbDRS2*-T_{IDP1}	FgF 21
PbDRS3	P_{TDH3}-*PbDRS3*-T_{IDP1}	FgF 21
PbDRS4	P_{TDH3}-*PbDRS4*-T_{IDP1}	FgF 21
PrDRS	P_{TDH3}-*PrDRS*-T_{IDP1}	FgF 21
PsDRS	P_{TDH3}-*PsDRS*-T_{IDP1}	FgF 21
PbDRR1	P_{TDH3}-*PbDRR1*-T_{EFM1}	pHUM plasmid
PbDRR2	P_{TDH3}-*PbDRR2*-T_{EFM1}	pHUM plasmid
PrDRR	P_{TDH3}-*PrDRR*-T_{EFM1}	pHUM plasmid
PsDRR	P_{TDH3}-*PsDRR*-T_{EFM1}	pHUM plasmid
PrDRS-PsDRR	P_{TDH3}-*PrDRS*-T_{IDP1}-LTP1-P_{CCW12}-*PsDRR*-T_{EFM1}	416d
NnCYP80A	P_{TDH3}-*NnCYP80A*-T_{ATG10}	USERX-1
NnCYP80G	P_{TDH3}-*NnCYP80G*-T_{ATG10}	USERX-1
LsCYP80G	P_{TDH3}-*LsCYP80G*-T_{ATG10}	USERX-1
AtATR2	P_{TDH3}-*AtATR2*-T_{ATG10}	USERX-1
CcCPR2	P_{TDH3}-*CcCPR2*-T_{ATG10}	USERX-1
HcCPR	P_{TDH3}-*HcCPR*-T_{ATG10}	USERX-1
PsCPR	P_{TDH3}-*PsCPR*-T_{ATG10}	USERX-1
AtCYB5D	P_{TDH3}-*AtCYB5D*-T_{ATG10}	USERX-1
CrCYB5	P_{TDH3}-*CrCYB5*-T_{ATG10}	USERX-1
EcCYB5	P_{TDH3}-*EcCYB5*-T_{ATG10}	USERX-1
PbCYB5	P_{TDH3}-*PbCYB5*-T_{ATG10}	USERX-1
BsCYP80A1 + HcCPR	P_{TDH3}-*BsCYP80A1*-T_{ATG10}-LTP1-P_{HHF1}-*HcCPR*-T_{PGK1}	USERXI-2

Supplementary Table 5 – *S. cerevisiae* integration sites utilized in this study

Target site ID	Target site sequence[a]	Reference
GRE2	TAGAATACGGAATTTTCTCGCGG	[3]
NnOMT5	TAACAGCTTTGCAGTTTTCGTGG	This study
USERX-1	GTAGCTACAAGAACATATGGTGG	[5]
USERX-3	GATCGCCGAATGGCACGCGAGGG	[5]
USERXI-2	AGTTTAGATGTAGGTTTTAGCGG	[5]
USERXII-1	CTTTGATTTTGGCATCGGTTCGG	[5]
416d	TAGTGCACTTACCCCACGTTCGG	[6]
511b	GGGTTTGGCACAATTTGGCTTGG	[6]
911b	GTAATATTGTCTTGTTTCCCTGG	[6]
106a	ATACGGTCAGGGTAGCGCCCTGG	[6]

[a] PAMs are underlined.

REFERENCES

1 Pyne, M. *et al.* in *Methods Enzymol.* Vol. 575 195-224 (Elsevier, 2016).

2 Ryan, O. W. *et al.* Selection of chromosomal DNA libraries using a multiplex CRISPR system. *Elife* **3**, e03703 (2014).

3 Pyne, M. E. *et al.* A yeast platform for high-level synthesis of tetrahydroisoquinoline alkaloids. *Nat. Commun.* **11**, 1-10 (2020).

4 Mülleder, M., Campbell, K., Matsarskaia, O., Eckerstorfer, F. & Ralser, M. *Saccharomyces cerevisiae* single-copy plasmids for auxotrophy compensation, multiple marker selection, and for designing metabolically cooperating communities. *F1000Research* **5** (2016).

5 Mikkelsen, M. D. *et al.* Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab. Eng.* **14**, 104-111 (2012).

6 Reider Apel, A. *et al.* A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **45**, 496-508 (2017).