


# Elucidating the enzyme network driving *Amaryllidaceae* alkaloids biosynthesis in *Leucojum aestivum*

Basanta Lamichhane<sup>1</sup>, Sarah-Eve Gélinas<sup>1</sup>, Natacha Merindol<sup>1</sup>, Manoj Koirala<sup>1</sup>, Karen Cristine Gonçalves dos Santos<sup>1</sup>, Hugo Germain<sup>1,2</sup> and Isabel Desgagné-Penix<sup>1,2,\*</sup> 

<sup>1</sup>Department of Chemistry, Biochemistry and Physics, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada

<sup>2</sup>Plant Biology Research Group, Trois-Rivières, QC, Canada

Received 12 September 2024;

revised 28 January 2025;

accepted 14 February 2025.

\*Correspondence (Tel 1-819-376-5011; fax 1-819 376-5012; email [isabel.desgagne-penix@uqtr.ca](mailto:isabel.desgagne-penix@uqtr.ca))

**Keywords:** specialized metabolism, cytochrome P450, phenol coupling, biosynthetic pathway, pharmaceuticals, metabolic engineering.

## Summary

*Amaryllidaceae* alkaloids (AAs) are diverse bioactive metabolites with significant pharmaceutical potential, derived from 4'-O-methylnorbelladine (4'OM). The biosynthesis of these compounds involves the condensation of tyramine and 3,4-dihydroxybenzaldehyde by norbelladine synthase (NBS) and/or noroxomaritidine/norcrugsodine reductase (NR), followed by O-methylation. Cytochrome P450 enzymes, particularly the CYP96T family, introduce further structural diversity through C–C couplings, resulting in lycorine, galanthamine and crinine cores. Despite their importance, the exact biosynthetic pathways remain poorly defined. In this study, we describe key enzymes from *Leucojum aestivum* (La), providing crucial insight into AA biosynthesis. Transient expression in *Nicotiana benthamiana* demonstrated that LaNBS and LaNRII catalyse the conversion of tyramine and 3,4-dihydroxybenzaldehyde to norbelladine, which is subsequently O-methylated by a norbelladine-4'-O-methyltransferase (LaN4'OMT) *in planta*. Co-agroinfiltration of LaNBS, LaNRII, LaN4'OMT and LaCYP96T1 resulted in the production of various phenol-coupled products, with lycorine as the predominant compound, alongside haemanthamine, crinine/vittatine and norgalanthamine. This study identifies LaCYP96T1 and LaCYP96T2 as the first monocot enzymes capable of catalysing all three regioselective C–C phenol couplings and also highlights the substrate promiscuity of LaNRII. The findings not only elucidate critical steps in AA biosynthesis but also open new avenues for biotechnological application in producing valuable alkaloids, offering potential for novel drug development.

## Introduction

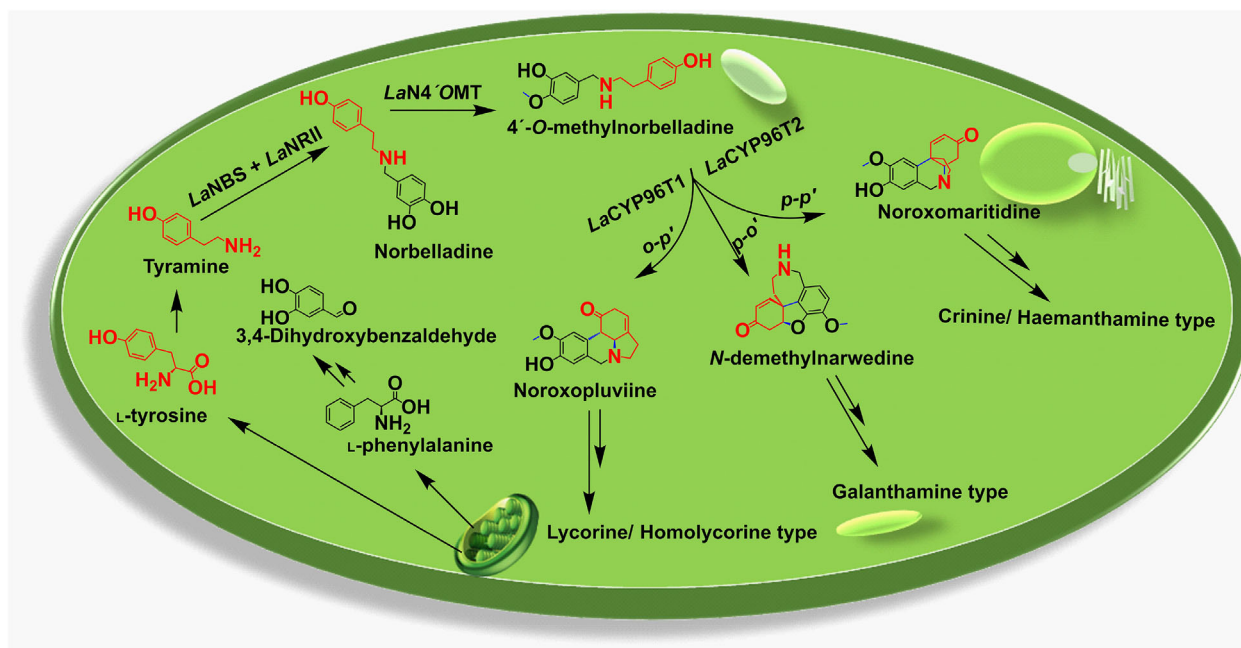
*Amaryllidoideae* from the *Amaryllidaceae* family produce a class of diverse specialized metabolites (SMs) with potent biological and therapeutic effects, known as the *Amaryllidaceae* alkaloids (AAs). Although galanthamine is the only AA commercialized as a treatment of mild symptoms of Alzheimer's disease owing to its acetylcholinesterase inhibitory properties, many other AAs have potential pharmacological applications. For instance, lycorine is extensively studied for its strong antiviral activities, and haemanthamine is known for its anticancer properties (Cahlikova *et al.*, 2020; Girard *et al.*, 2022).

However, the potential exploitation of AAs is hindered by their limited commercial availability (Takos and Rook, 2013). On the one hand, the current supply relies on AAs extracted from wild plants, such as *Leucojum aestivum* (summer snowflake), with varying yields (Georgievaa *et al.*, 2007). On the other hand, total chemical synthesis is challenging due to the intricate carbon ring structures of these metabolites, compounded by environmentally unsustainable purification processes. *In vitro* culture of galanthamine-producing species, primarily *Narcissus confusus* and *L. aestivum*, have gained attention as an alternative strategy (Berkov *et al.*, 2009; Kohut *et al.*, 2007; Koirala *et al.*, 2022; Pavlov *et al.*, 2007) but are not widely used to date due to the resulting limited accumulation of AAs.

Metabolic engineering presents a promising alternative avenue for future AA production, contingent upon a thorough

understanding of the biosynthetic pathway. Yet, while our comprehension of the chemistry and pharmacological properties of AAs is rapidly advancing, there remains a scarcity of knowledge regarding their synthesis (Desgagné-Penix, 2021; Kilgore and Kutchan, 2016; Liyanage *et al.*, 2024; Takos and Rook, 2013).

Roughly, the biosynthesis of AAs is governed by an extensive network of genes that code for enzymes, orchestrating a cascade of biochemical transformations starting from basic amino acid precursors to yield the structurally complex alkaloids characteristic of *Amaryllidoideae* (Figure 1). Specifically, from the plastidic aromatic amino acids pathway, L-tyrosine undergoes decarboxylation to yield tyramine (Xu *et al.*, 2020), while L-phenylalanine, is converted to 3,4-dihydroxybenzaldehyde (3,4-DHBA) most likely through the phenylpropanoid pathway (Desgagné-Penix, 2021; Koirala *et al.*, 2022). Previous *in vitro* studies using norbelladine/norcrugsodine synthase (NBS) and noroxomaritidine/norcrugsodine reductase (NR) from *Narcissus pseudonarcissus*, *N. papyraceus* and *L. aestivum* suggested that the condensation of tyramine and 3,4-DHBA yields norcrugsodine, further reduced to form norbelladine (Figure 1) (Hnin *et al.*, 2024b; Liu *et al.*, 2023; Majhi *et al.*, 2023; Singh *et al.*, 2018; Tousignant *et al.*, 2022). A 4'-O-methyltransferase, identified in *Narcissus* sp. aff. *pseudonarcissus*, *N. papyraceus*, *Lycoris aurea* and *Lycoris radiata*, catalyses the methylation of norbelladine *in vitro* into the metabolic intermediate of AAs, 4'-O-methylnorbelladine (4'OM) (Kilgore *et al.*, 2014; Li *et al.*, 2019;



**Figure 1** Proposed alkaloid biosynthesis pathway in *Leucojum aestivum*. The diagram illustrates the hypothesized sequence of biochemical reactions leading to the synthesis of galanthamine-, lycorine/homolycorine- and crinine/haemanthamine-types of alkaloids in *Leucojum aestivum*, highlighting key enzymatic steps and intermediate compounds. *o-p'*, *ortho-para'* coupling; *p-o'*, *para-ortho'* coupling; *p-p'*, *para-para'* coupling.

Sun *et al.*, 2018). Isotope labelling studies in *L. aestivum* revealed that deuterium-labelled 4'OM was further incorporated into galanthamine and lycorine through a C–C coupling cyclization step, as observed in the biosynthesis pathway of berberine-type alkaloids (Cui *et al.*, 2007; El Tahchy *et al.*, 2010; Iwasa and Kim, 1997). Later studies uncovered that a new class of cytochrome P450, CYP96T1 catalyses mainly *para-para'* (*p-p'*) C–C oxidative coupling of 4'OM in *N. sp. aff. pseudonarcissus* but also yields the *para-ortho'* (*p-o'*) product *N*-demethylnarwedine as a minor product (Kilgore *et al.*, 2016a). Recently, three homologues identified from *Narcissus* cv. Tête-à-Tête were shown to exhibit specific catalytic activity towards *o-p'*, *p-o'* or *p-p'* coupling reactions (Mehta *et al.*, 2024), and four CYP96T1 homologues from *L. aurea* catalysed exclusive *p-p'* or *p-o'* scaffold formation (Liu *et al.*, 2024). These recent discoveries have led to the reconstitution of a partial galanthamine pathway starting from the C–C oxidative coupling of 4'OM in *Nicotiana benthamiana*. While they provided meaningful insight into galanthamine and haemanthamine scaffold formation, the functionality of orthologous *Amaryllidoideae* CYP96T, especially towards the *o-p'* coupling remains unclear. Moreover, the reconstitution of a complete pathway, that is, starting from precursors, rather than *Amaryllidoideae*-restricted 4'OM is crucial both for elucidation purposes and for synthetic biology applications.

*Nicotiana benthamiana* contains all necessary cofactors, intracellular compartments along with protein-folding mechanisms essential for the efficient synthesis of plant proteins (Reed and Osbourn, 2018). It has been extensively used to investigate plant metabolic pathways, from the characterization of enzymes to the elucidation of pathway, such as the monoterpene indole alkaloids (Boccia *et al.*, 2022; Miettinen *et al.*, 2014), strictosidine (Dudley *et al.*, 2022), or colchicine (Nett *et al.*, 2020), to name but a few.

In this study, we reconstituted the biosynthetic pathways of AAs through heterologous expression in *N. benthamiana* using *Agrobacterium*-mediated transient T-DNA expression. By leveraging the genetic tractability of *N. benthamiana*, we hypothesize that it is possible to produce therapeutically valuable AAs. This research provides insights into the complex biosynthetic pathways of these alkaloids and facilitates the development of a metabolic engineering platform.

## Results

### In silico characteristics of pathway candidates

We initially examined the sequences of the proposed enzymes involved in the pathway of AAs (Figure 1), focusing on the candidates identified in the *L. aestivum* transcriptome (Tousignant *et al.*, 2022). Prospects for norbelladine synthesis included LaNBS and LaNR11. LaN4'OMT was proposed to catalyse the formation of 4'-O-methylnorbelladine, and CYP96Ts were candidates for further diversification of SMs through oxidative phenol couplings.

LaNBS belonged to the PR10/Bet v1-like protein superfamily, similar to previously characterized norcoclaurine synthase (NCS) (Figure S1a). One candidate transcript of NBS was identified from the transcriptome and sequencing of the amplified candidate revealed a single alteration c.23A > G (Glu8Gly) (Figure S1b). LaNBS subclustered with orthologous *Narcissus papyraceus* NBS (NpNBS) sequences within a larger clade containing characterized NCS enzymes, distinct from the non-NCS PR10 enzyme clade. The 159 amino acids long candidate shared almost 100% identity with the previously described LaNBS by Tousignant *et al.*, Majhi *et al.*, along with 87% and 86% similarity with orthologs from *Narcissus pseudonarcissus*, *N. papyraceus*, respectively, and 38%, 36% and 31% with NCS from *Coptis japonica*, *Thalictrum flavum* subsp. *glaucom* and *Papaver somniferum*, respectively

(Table S6a). Its sequence and structure included a conserved glycine-rich P-loop, proximal to the entrance of the ligand-binding pocket, a characteristic feature of the Bet v1 protein (Figure S1b, Figure 1c). Eight ligand-binding residues were conserved in all Bet v1 transcripts from various species (Figure S1b), including hydrophobic residues Tyr68, Tyr100 for wider opening of the substrate binding tunnel and Lys83 for intercepting the carbonyl group of the substrate. Modelling of LaNBS revealed a seven-stranded antiparallel  $\beta$ -sheets with two long N- and C-terminal helices surrounding a cleft with a core of hydrophobic residues, confirming its Bet v1-like structure and resemblance with T7NCS (Figure S1c). Previously proposed interacting residues Ser31 (Leu31 in homologous LaNBS), Leu34, Tyr59, Tyr68, Glu71, Phe73, Lys83, Val85, Leu104, Pro139, and Ile143 formed the binding pocket of the enzyme (Figure S1b, Table S7) (Hnin et al., 2024b; Majhi et al., 2023). The similarity in sequence and structures with functional NBSs and characterized NCSs, the latter catalysing the condensation of dopamine and 4-hydroxyphenylacetaldehyde (4-HPAA), supports the hypothesis that the putative LaNBS homologue performs the condensation of 3,4-DHBA and tyramine (Figure 1).

Eleven sequences and two previously reported transcripts were identified as candidates for the reduction of norcraugsodine and other AAs, from the transcriptome and from the NCBI database, respectively. They clustered with the tropinone reductase (TR) subfamily of the short-chain dehydrogenase/reductase (SDR) family (Figure S2a). Transcripts 'NR1', 'NR2' and 'Red.8' clustered with *Narcissus* cv *Tête-à-Tête* NtSDR1, *N. aff. pseudonarcissus* (NpsNR), *N. pseudonarcissus* and *N. papyraceus* (NpNR) clade. NpNR and NpsNR catalyse norbelladine synthesis and/or norcraugsodine reduction, while NtSDR1 accepts noroxopluvine as substrate. Amplification led to the identification of LaNRII matching uncharacterized NR2 in NCBI (UIP35235.1). LaNRII 269 aa sequence corresponded to a classical SDR having TGXXXGXXG cofactor binding motif in position 29–33 and YXXXK active site motif in 173–177 position, as well as a conserved NNAG domain required for stabilization of central  $\beta$ -sheets and reductase activity (Filling et al., 2002), and PGxxxT that is suggested to play a structural role (Figure S2b,c). LaNRII predicted folding was arranged as a seven-stranded parallel  $\beta$ -sheet embedded between a pair of three  $\alpha$ -helices (Figure S2c).

The candidate included conserved Asn132 and Ser160, necessary for the activity of other reductase, such as SDR hsdA from *Comamonas testosteroni* (corresponding residues Asn111 and Ser138) (Filling et al., 2002). LaNRII also displayed Tyr173, aligning with Tyr209 from tropinone reductase SDR of *Cochlearia officinalis* (CoTR), essential for the reduction of tropinone and nortropinone (Brock et al., 2008). LaNRII shared 98%, 93%, 91% and 77% sequence similarity with previously characterized LaNRI, NpNR, NtSDR1, and NpsNR sequences, respectively (Table S6b). Comparison with these reductases revealed the conservation of a total of 49 active site residues with NtSDR1 and NpNR1 (Tables S7 and S8), 15 of which differed with crystalized NpsNR. Conserved residues included Lys33, Ile35, Gly36, Cys53, Arg55, Cys79, Asn108, Ala109, Gly110, Thr111, Leu114, His170, Tyr173, Lys177, Ile205, Thr206, Thr208, Val209, Gln210, Phe214, Leu215, Glu224, and Arg265, previously predicted to interact with ligands. The only active site residue that was distinct in LaNRII compared to NtSDR1 was Ile vs Met239. Additional mutations could be noted in surrounding residues and the N-terminal tail. Together, these results suggest that LaNRII

contains the required features to carry out the reduction catalysis of imine and/or alkaloid substrates.

Six candidate OMTs were identified in the *L. aestivum* transcriptome (Figure S2c). On the basis of maximum likelihood and consensus (not shown) trees, a putative O-methyltransferase (LaN4'OMT) was selected for primer design and amplification. There were 2 amino acid differences compared to the transcriptome sequence, that is, c.52G > A (p.Gly18Ser) and c.299T > C (p.Met100Thr). The identified transcript clustered with class I OMT encompassing other previously characterized enzymes from different species like *C. japonica*, *P. somniferum*, *Lycoris* and *Narcissus* species (Figure S3a). The distance matrix showed >95% similarity with six orthologous enzymes, but only 12% with NtOMT1 (Table S6c). Except for NtOMT1, which catalysed O-methylation of 11-hydroxyvitattine, all reference *Amaryllidaceae* OMT were shown to catalyse O-methylation of norbelladine and phenylpropanoid precursors. Class I OMTs are metal-dependent enzymes characterized by a distinctive feature: a Rossmann fold-binding site for S-adenosyl-L-methionine (SAM), as the methyl donor (Figure S3b,c). Two hundred and thirty-nine amino acid-long LaN4'OMT displayed conserved residues involved in interactions with SAM such as Val55, Glu77, Gly79, Val80, Tyr84, Ala132A, and Asp157 (Li et al., 2019). Conserved Asp155, Asp181, and Asn182 are involved in metal ion binding in orthologous enzymes. Asn182 is also implicated in the methyl transfer catalysis together with Lys158 and Asp230 (Li et al., 2019) (Figure S3b). The predicted LaN4'OMT candidate folded similarly to a catechol OMT, with 7  $\beta$ -sheets encompassed between  $\alpha$ -helices and loops (Figure S3c).

Previous studies have suggested the involvement of 19 amino acids in the substrate specificity of *Amaryllidaceae* OMT (Hnin et al., 2024a; Koirala et al., 2024). LaN4'OMT displayed two residues that were proposed to be associated with a preference towards 3'-O-methylation catalysis of norbelladine (Val30, Gln141), but most were suggestive of 4'-O-methylation (Ala179, Tyr186, Ser188, Tyr191, Leu206, Thr213, Val232) (Table S9). Other variations in residues (N-terminal loop) are surface-exposed and unlikely to interact directly with ligands but could be involved in a catalytic network to support substrate interaction. Overall, this analysis supports that the LaN4'OMT candidate could catalyse the 4'-O-methylation of norbelladine.

Two transcripts were previously reported for CYP96T1 and CYP96T2 (GenBank: UIP35232.1 and UIP35233.1) and in *L. aestivum* transcriptome nine additional candidates were identified (Figure S4a). LaCYP96T candidates clustered with orthologous *p-p'*, *p-o'*, *o-p'* couplings catalysing enzymes, and formed a clade distinct from the CYP80, CYP71, and CYP76 subfamilies (Figure S4a,b). Following amplification, the 512 amino acid-long candidates included a signal peptide (residue 7–29 predicted by IntelPro) to embed the protein in the membrane at the cytosolic surface of the ER. LaCYP96T1 displayed 17 bases alterations resulting in 4 amino acid differences compared to the GeneBank sequence, while LaCYP96T2 had 36 base differences resulting in a change of 12 amino acids. They both subclustered with the recently uncovered *o-p'* coupling catalysing NtCYP96T5 (Mehta et al., 2024). LaCYP96T1 and LaCYP96T2 proteins shared 98% identity and exhibited 92% similarity with LauCYP96T1-like-2 from *Lycoris aurea*, 90% with both LauCYP96T1-like-3 and NtCYP96T5, and 72% with NpsCYP96T1 (Table S6d). The predicted  $\alpha$ -helix-rich

shape of LaCYP96T1 and LaCYP96T2 was characteristic of CYPs (Figure S4c,d), which bind their diverse ligands within a buried, hydrophobic active site accessed through a substrate channel formed by two flexible helices and their connecting loop (Figure S4c). The N-terminal ER-membrane-anchoring  $\alpha$ -helix domain (at the left in Figure S4c,d), and the heme pocket buried inside the enzyme large cavity were apparent in the folded proteins (at the centre of each prediction).

The oxygen binding and activation motif (A/G)GX(D/E)TT was substantially different between LaCYP96T1 and its variants when compared to CYPs that catalyse hydroxylation reactions. LaCYP96T1 and LaCYP96T2 also contained the heme-binding domain (FXXGXRXCXG) that secures the heme in the active site by a strongly preserved thiolate residue (Cys457) which interacts with two molecular oxygens, one oxygen molecule oxidizing the substrate and the other being reduced to form water (Porter and Coon, 1991) (Figure S4b). Ala219, Leu220, Thr303, Leu306, Val307, Thr311, and Pro382-Ile385 are predicted to be involved in substrate binding pocket along with 25 other residues conserved throughout various species (Figure S4b,c, Table S1). Out of 47 active site residues predicted by MOE (Table S7), 11 differed from *NpsCYP96T1*, while they matched with *LauCYP96T1-like2* sequence catalysing *p-p'* coupling and with *NtCYP96T5* (*o-p'* coupling) (Table S10) (Kilgore et al., 2016a; Liu et al., 2024; Mehta et al., 2024). Furthermore, LaCYP96T1 and T2 encoded for Leu123 and Phe124, a combination that was specifically detected in enzymes performing *o-p'* or *p-p'* couplings, but not *p-o'*. This suggested that the LaCYP96T candidates could preferentially catalyse *o-p'* and/or *p-p'* reactions.

Taken together, phylogenetic analysis, multiple sequence alignment, and folding prediction of the five candidates provide evidence to support their capability in executing the proposed biosynthetic pathway (Figure 1).

### Localization assay

We previously showed that LaNBS and LaNR both localize in the cytoplasm (Majhi et al., 2023; Tousignant et al., 2022). To further investigate the spatial organization of the AA pathway and its potential impact on metabolite flux, we studied the subcellular localization of LaN4'OMT and LaCYP96T2 following

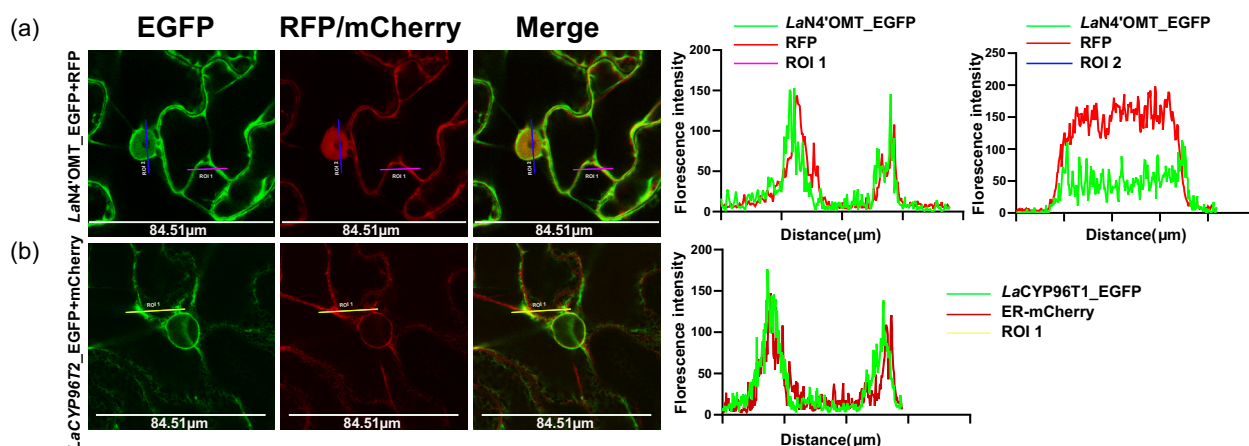
agroinfiltration in *N. benthamiana* (Figure 2). The red fluorescent protein (RFP), which localizes to the cytoplasm and nucleus, and ER\_M-cherry which is restricted to the endoplasmic reticulum (ER), were used as controls.

Both N- and C-terminal LaN4'OMT-EGFP accumulated in the cytoplasm (Figure 2a, Figure S5), while LaCYP96T2 localized in the ER (Figure 2b, Figure S5). Expression of the fusion proteins was verified by western blot. Both amino and carboxy-terminal-tagged EGFP fusion LaN4'OMT proteins were observed at around 54 kDa in blot, and C-terminal-EGFP fusion LaCYP96T2 at around 87 kDa consistent with a GFP fusion (Figure S6a,b). These results suggest that LaN4'OMT resides in the cytosol, the same compartment as LaNBS and LaNR, while LaCYP96T localizes to the ER, confirming the expected subcellular localization of the candidates.

### Reconstitution of the foundation steps of the AA pathway to produce 4'O-methylnorbelladine *in vivo*

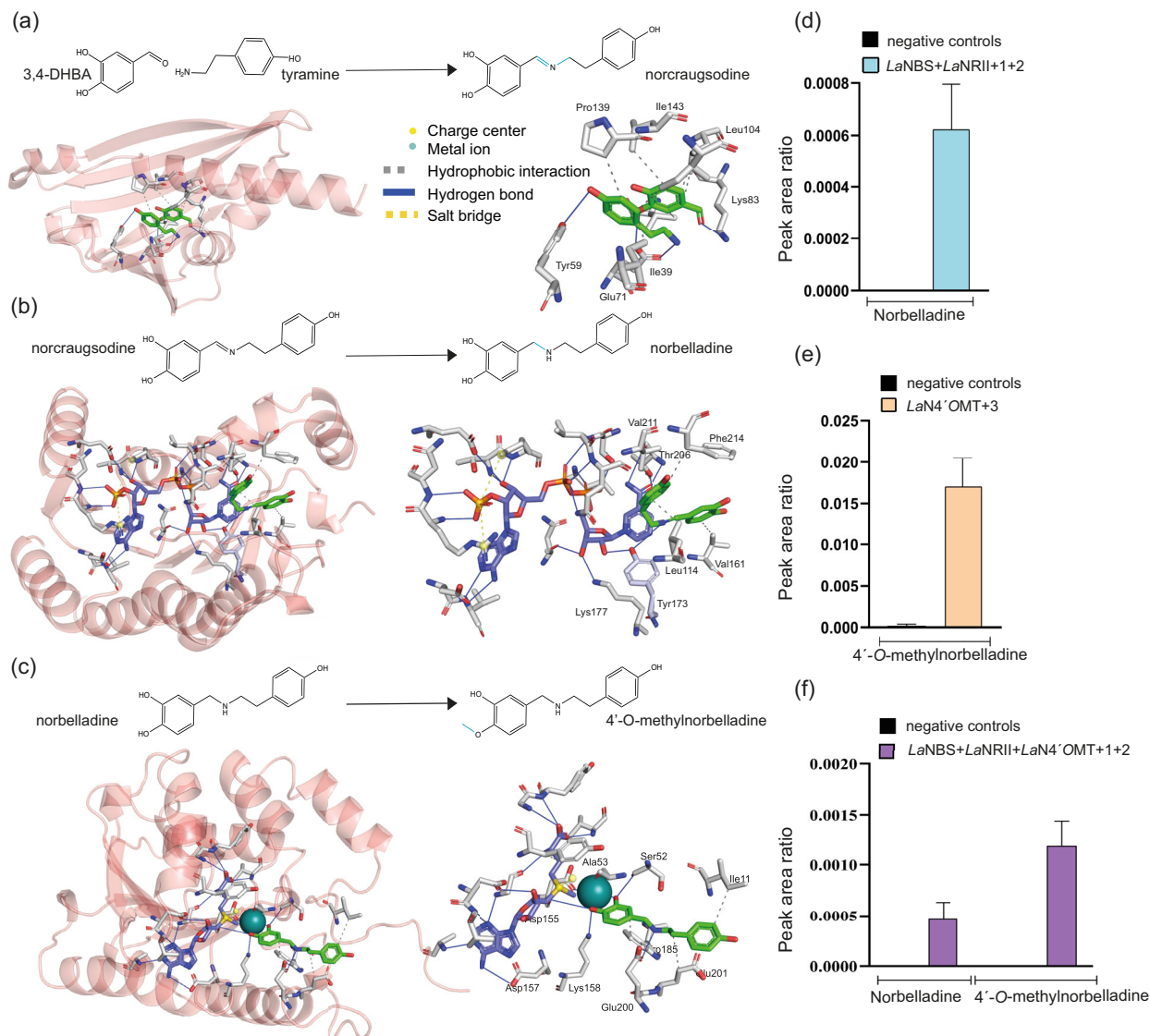
The proposed pathway entails the condensation of 3,4-DHBA and tyramine by LaNBS to yield norcraugsodine, which is subsequently reduced by LaNR11 and O-methylated by LaN4'OMT, as crucial initial steps (Figure 1). Having established the candidates' phylogeny and localization, we proceeded to characterize the interaction of enzymes with their suggested ligands, both *in silico* and *in vivo* (Figure 3).

Tyramine and 3,4-DHBA were docked into the active site of LaNBS with binding scores of  $-5.1$  and  $-5.5$  kcal/mol, respectively. In their predicted calculated position, the ligands faced each other, 3,4-DHBA forming hydrogen bonds with catalytic residues Glu71 and Lys83, and tyramine with Ile39, Tyr59, and Glu71. Additionally, both engaged in hydrophobic interactions with surrounding residues: tyramine interacted with Ile39, Pro139, Ile143, and Val85, while 3,4-DHBA interacted with Leu104 and Leu140 (Figure 3a, right panel, Table S1). The near-parallel orientation of the aromatic rings of the two ligands, and the carbonyl group of 3,4-DHBA facing tyramine's amine group are reminiscent of dopamine and hydroxybenzaldehyde positioning in *TfNCS* (Samanani and Facchini, 2002). Hence, the docking analysis was consistent with Lys83- and Glu71-assisted condensation to form norcraugsodine. The latter is proposed to be



**Figure 2** Subcellular localization of LaN4'OMT and LaCYP96T1 proteins in *Nicotiana benthamiana*. (a) panel showing co-agroinfiltrated C-terminal-EGFP-fusion LaN4'OMT with RFP marker. (b) panel showing co-agroinfiltrated C-terminal-EGFP-fusion-LaCYP96T1 with ER-mcherry marker. The histogram of the right in each panel shows the fluorescence intensity profile of respective panel. Scale bars resemble 84.51 μm. Image processed in Inkscape software.





**Figure 3** Reconstitution of the foundation steps of the Amaryllidaceae alkaloid pathway to produce 4'-O-methylnorbelladine *in vivo*. (a) Catalysed condensation of 3,4-DHBA and tyramine. (b) *LaNR11*-catalysed reduction of norcraugsodine. (c) *LaN4' OMT*-catalysed 4'-O-methylation of norbelladine. Ligands were docked in the candidate enzyme active site and are shown as green sticks. The AlphaFold2 predicted structure are shown as a transparent deep salmon cartoon (left) and active site residues are shown as light grey sticks. NADPH and SAM are shown as violet sticks.  $Mg^{2+}$  is displayed as a blue sphere. A specific view on the active site residues and ligands positioning with interaction is shown (right). The blue bonds in the schemes display the catalysed reactions. (d) Relative quantification of norbelladine production *in planta* following agroinfiltration with *LaNBS* and *LaNR11* and feeding with 3,4-DHBA and tyramine. (e) Relative quantification of 4'-O-methylnorbelladine production upon agroinfiltration of *LaN4' OMT* and feeding with norbelladine. (f) Relative quantification of norbelladine and 4'-O-methylnorbelladine production by agroinfiltrated *LaNBS*, *LaNR11* and *LaN4' OMT* upon feeding with 3,4-DHBA and tyramine. The product profiles obtained in different assays were analysed by HPLC-MS/MS. The area ratio of each detected analyte was obtained by dividing the peak area of the targeted compound by the one of the internal standard papaverine. Relative quantification was performed using the area ratio of each analyte produced in the assay. Data are the means  $\pm$  SE of three biological repeats with background subtracted. 1 is denoted for tyramine, 2 for 3,4-dihydroxybenzaldehyde and 3 is for norbelladine. *N. benthamiana* infiltrated with *Agrobacterium* harbouring EGFP and respective substrate/s are the negative controls in bar plots. Representative detection of at least three independent experiments with three leaves infiltrated in each replicate.

further reduced by the noroxomaritidine/norcraugsodine reductase *LaNR11*.

Norcraugsodine docked with a score of  $-6.3$  kCal/mol in the active site of *LaNR11*, its N-C double bond laid closest to the nicotinamide ring of the electron donor NADPH and to Tyr173, interacting with Leu114, Val161, Val211, Phe214, and Tyr173 of the enzyme active site (Table S1, Figure 3b). This positioning

aligns with the proposed NADPH-dependent reduction of norcraugsodine to produce norbelladine assisted by Tyr173. In the next step yielding to 4'OM, norbelladine docked in *LaN4' OMT* active site with a score of  $-8.8$  kCal/mol, forming interactions with  $Mg^{2+}$ , Lys158, as well as Ile11, Ser52, Asp155, Trp185, Glu200 and Glu201.  $Mg^{2+}$  was positioned between the ligand 3' and 4'OH groups, the 4'OH located proximal to the SAM methyl

group (Table S1, Figure 3c). The docking results are consistent with the catalysis of a 4'-O-methylation of norbelladine, in interaction with Lys158, Asn182, and Mg<sup>2+</sup>.

To achieve the reconstitution of the original steps of the *Amaryllidaceae* pathway, we agroinfiltrated the genes encoding the candidate enzymes into *N. benthamiana*. Forty-eight hours post-agroinfiltration, we injected the substrates and harvested the leaves for metabolite analysis 24 h later. First, we confirmed protein accumulation by western blotting (Figure S6c–f), detecting bands corresponding to the expected sizes (~20 kDa for LaNBS, ~30 kDa for LaNRII and the EGFP control, ~35 kDa for LaN4'OMT, and ~60 kDa for LaCYP96T1, as well as LaCYP96T2). The expression level was further validated by RT-qPCR, which showed the relative expression of each gene while infiltrated alone or co-infiltrated together (Figure S6h–k). Previous *in vitro* studies suggested that norcaugsodine is unstable and that the yield of norbelladine synthesis increases when LaNBS and LaNR interact (Majhi *et al.*, 2023). Therefore, we co-agroinfiltrated LaNBS and LaNRII with tyramine and 3,4-DHBA (Figure 3d, Figure S7a). Under these conditions, norbelladine was detected in the metabolite extracts. In contrast, feeding EGFP-infiltrated leaves with the substrate, feeding substrates to leaves alone, or non-fed LaNBS and LaNRII-infiltrated leaves did not produce detectable levels of norbelladine.

Subsequently, agroinfiltration of LaN4'OMT with norbelladine confirmed the production of 4'-O-methylnorbelladine *in planta* (Figure 3e, Figure S7b). Finally, co-infiltration of the genes encoding the three enzymes LaNBS, LaNRII and LaN4'OMT with tyramine and 3,4-DHBA as substrates led to the production of both norbelladine and 4'-O-methylnorbelladine, which were undetected in the negative controls (Figure 3f).

To validate the functionality of the candidates in this premise pathway, the activity of LaNBS, LaNRII and LaN4'OMT was verified *in vitro* by using crude enzymes from *N. benthamiana* or purified N-terminal MBP-tagged proteins (Figure S6g). First, crude extracts from *N. benthamiana* leaves agroinfiltrated with LaNBS and LaNRII led to the detection of the condensed product norbelladine upon feeding with tyramine and 3,4-DHBA (Figure S7c). *E. coli* harvested purified MBP-tagged enzymes also led to the synthesis of norbelladine *in vitro*, as detected in two out of three samples. Of note, the amount yielded from purified enzymes expressed in bacteria was almost three times lower compared to crude enzymes from plant leaves (Figure S7c). To get additional insight into the interaction of the substrates with the two purified enzymes, we compared the thermal unfolding profiles of pure LaNBS and LaNRII with and without ligands using a Prometheus. We observed differences in the melting temperature (T<sub>m</sub>) of both enzymes when bound to their ligand, whereas ligand addition to pure MBP did not impact its T<sub>m</sub>. The thermal shift resulted in a decrease of the F350/F330 ratio above the threshold limit (>0.2), confirming significant changes in LaNBS and LaNRII folding upon interaction with 3,4-DHBA and tyramine (Figure S7d–f). Furthermore, we observed that 4'OM is produced from norbelladine *in vitro* upon incubation with purified LaN4'OMT and Mg<sup>2+</sup>. 4'OM levels increased constantly during the 60 min of the reaction, reaching up to ~28 µg/mL (Figure S7g) and no dimethylnorbelladine was detected. As previous studies showed that metal ion can impact OMTs regioselectivity (Su *et al.*, 2022), we tested the reaction in the presence of Ni<sup>2+</sup> instead of Mg<sup>2+</sup>. This led to the detection of 4'OM as a major product, along with 3',4'-O-dimethylnorbelladine (Figure S7h). LaN4'OMT also showed

promiscuity by methylating 3,4-DHBA yielding 7602 ng/mL vanillin as a major product along with 526 ng/mL isovanillin (Figure S7i).

Overall, these results corroborate the proposed grounds of the biosynthetic pathway *in vivo* and *in vitro*.

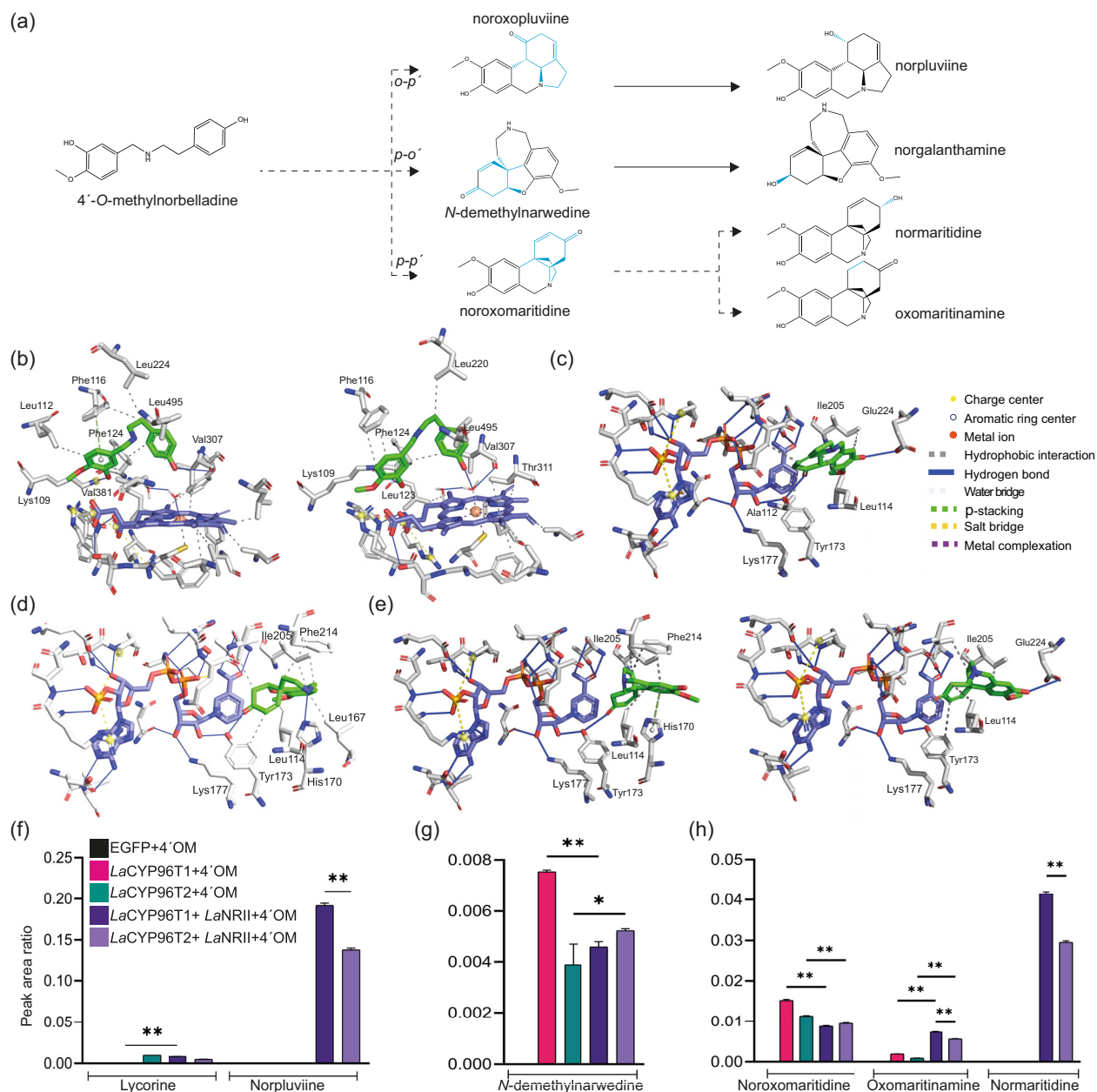
### Multiple C–C phenol couplings by a single CYP96T

In the subsequent steps of the proposed biosynthetic pathway (Figure 1), 4'OM undergoes significant transformations (Figure 4), leading to the generation of a diverse array of alkaloid structures catalysed by CYP96T isoforms (Figure 4a). The active site of these enzymes includes a heme group and two water molecules, essential for correct ligand positioning and catalysis. 4'OM docked with its 4-OH in the vicinity of the iron-bonded water molecule (Figure 4b). The ligand interacted with Lys109, Leu112, Phe116, Phe124, Leu224, Val307, Val381, and Leu495, in CYP96T1, and with Lys109, Phe116, Phe124, Leu220, Val307, and Leu495 in CYP96T2, with the O-methylated ring stabilized by aromatic residues (Table S1, Figure 4b, Figure S8a,b). Phe124 was suggested to be involved in *o*-*p'* and *p*-*p'* coupling reactions, while Val307 was detected in all *o*-*p'*, *p*-*o'*, and *p*-*p'* catalysing enzymes. This positioning supports the catalysis of a C–C-phenol coupling reaction, leading to the synthesis of noroxopluiine, noroxomaritidine, and possibly *N*-demethylnarwedine, in both enzymes active site (Figure 4b).

To validate enzyme activity *in vivo*, we expressed LaCYP96T1 or LaCYP96T2 and fed 4'OM as the substrate. Although we did not detect the *o*-*p'* phenol-coupled product noroxopluiine, lycorine was intriguingly detected with LaCYP96T2 and 4'OM (Figure 4f). In leaves agroinfiltrated with LaCYP96T1 or LaCYP96T2 and 4'OM as substrate, we also observed the *p*-*o'* coupled product *N*-demethylnarwedine (Figure 4g, Figure S9a,b), which was absent in all five negative controls. The accumulation of *N*-demethylnarwedine was nearly twice as high with LaCYP96T1 compared to LaCYP96T2 (*P* < 0.0001, Tukey's multiple comparison post hoc test, one-way ANOVA). Additionally, we detected the *p*-*p'* phenol-coupled product noroxomaritidine and its reduced form oxomaritinamine in leaves agroinfiltrated with LaCYP96T1 or LaCYP96T2 gene and 4'OM (Figure 4h, Figures S10a,b and S11a,b), whereas these products were undetectable in leaves agroinfiltrated with EGFP and 4'OM. Therefore, LaCYP96T1 and LaCYP96T2 catalysed the synthesis of the three types of phenol-coupled products, namely, *p*-*o'* (galanthamine-type), *p*-*p'* (crinine-type) and *o*-*p'* (lycorine-type) involved in AAs biosynthesis pathway *in planta*. *p*-*p'* coupling activity of LaCYP96T1 was verified *in vitro*, using microsomes isolated from agroinfiltrated *N. benthamiana* leaves which led to the formation of noroxomaritidine from 4'OM (Figure S10d).

### Multiple reduction steps by LaNRII

The reduction step is key for stabilizing metabolites and producing end-of-pathway AAs. We hypothesized that LaNRII could accept *N*-demethylnarwedine, noroxopluiine and noroxomaritidine as substrates. Docking results supported this hypothesis, showing noroxopluiine and *N*-demethylnarwedine positioned to reduce their C–O double bonds through interaction with catalytic Tyr173 near the nicotinamide ring of NADPH at the enzyme's active site (Table S1, Figure 4c,d, Figure S8c,d). Noroxomaritidine docked in two distinct orientations, favouring the reduction of the C–O or the C–C double bond, in interaction with Tyr73 (Table S1, Figure 4e, Figure S8e,f).



**Figure 4** Reconstitution of the branching pathway steps to produce Amaryllidaceae alkaloids *in vivo*. (a) Scheme of the proposed *LaCYP96T*-catalysed phenol couplings of 4'-O-methylnorbelladine yielding noroxopluiine, *N*-demethylnarwedine, noroxomaritidine, further reduced by *LaNR11* into norpluviine, norgalanthamine, normaritidine and oxomaritinamine. Blue bonds are the results of the proposed catalysed reactions. (b) Docked 4'-O-methylnorbelladine in the active site of CYP96T1 (left) and CYP96T2 (right). (c) Docked noroxopluiine in the active site of *LaNR11* to yield norpluviine. (d) Docked *N*-demethylnarwedine in the active site of *LaNR11* to yield norgalanthamine. (e) Docked noroxomaritidine in the active site of *LaNR11* to yield normaritidine (left) or oxomaritinamine (right). Candidate active site residues are shown as grey sticks and ligands as green sticks. Heme and NADPH are shown as purple sticks. Iron as a red sphere. Interactions are shown as lines as displayed in the legend. (f) Relative quantification of *ortho-para'* products norpluviine and lycorine. (g) Relative quantification of *para-ortho'* product *N*-demethylnarwedine. (h) Relative quantification of *para-para'* products noroxomaritidine, oxomaritinamine and normaritidine. The area ratio of each detected analyte was obtained by dividing the peak area of the targeted compound by the internal standard papaverine. Relative quantification was performed using the area ratio of each analyte produced in the assay. Data are the means  $\pm$  SE of three biological repeats with background subtracted. *N. benthamiana* infiltrated with *Agrobacterium* harbouring EGFP and respective substrate/s is the negative control in each bar plot. \*\* and \* represent the statistically significant differences among means with  $P < 0.0001$  and  $P < 0.05$ , respectively.

To further investigate the pathway reconstitution, we co-expressed *LaCYP96T1* or *LaCYP96T2* with *LaNR11* and fed 4'OM as substrate. This resulted in the detection of the reduced *o-p'* phenol-coupled product norpluviine (Figure 4f). The norpluviine

signal was significantly higher in leaves infiltrated with *LaCYP96T1* and *LaNR11* (~1.4 times,  $P < 0.0001$ ) compared to *LaCYP96T2* and *LaNR11* (Figure 4f, right panel, Figure S12a,b). Norpluviine was not detected in any negative controls or samples

lacking the reductase enzyme. Lycorine was also detected in samples infiltrated with either of the *LaCYP96T* and the *reductase* (Figure 4f, Figure S13). Interestingly, co-infiltrating *LaCYP96T1* and *LaNR11* resulted in a decreased accumulation of *N*-demethylnarwedine to the level of *LaCYP96T2* ( $P < 0.0001$ , Tukey's multiple comparison post hoc test, one-way ANOVA, Figure 4g). On the other hand, the yield of *N*-demethylnarwedine rose slightly (by 1.3-fold), but significantly ( $P = 0.0105$ ) when *LaNR11* was co-infiltrated with *LaCYP96T2*.

Co-infiltrating *LaCYP96T1* or *LaCYP96T2* and *LaNR11* also led to an increase in oxomaritinamine by nearly 4- and 6-fold respectively, compared to when infiltrated without *LaNR11* ( $P < 0.0001$ , Tukey's multiple comparison test by ordinary two-way ANOVA, Figure 4h). Strikingly, agroinfiltration of *LaCYP96T1* or *LaCYP96T2* with *LaNR11* resulted in a strong signal of normaritinidine, which had not been detected in samples with *LaCYP96T1* or *LaCYP96T2* alone (Figure 4h, Figure S14). Normaritinidine accumulated significantly more following the infiltration of *LaCYP96T1* compared to *LaCYP96T2* (1.4-fold,  $P < 0.0001$ ). Noroxomaritinidine, oxomaritinamine, and normaritinidine were not detected in any negative controls. Thus, in addition to norbelladine biosynthesis, *LaNR11* catalyses the reduction of all phenol-coupled products *in planta*.

To further substantiate the enzyme activity, we conducted *in vitro* assays using microsomes isolated from *Nicotiana benthamiana* leaves expressing *LaCYP96T1*, in conjunction with the purified reductase enzyme (*LaNR11*) expressed in *E. coli*. These assays with 4'OM as substrate, revealed norpluviine as the predominant product, accompanied by the formation of normaritinidine and noroxomaritinidine (Figure S12d,e). These results confirmed the ability of *LaCYP96T1* and *LaNR11* to yield *o*-*p'* and *p*-*p'* reduced products *in vitro*.

### Full pathway reconstitution with four enzymes

To confirm that the four candidates could reconstitute the complete pathway, from aromatic amino acid and phenylpropanoid precursors, we infiltrated *LaNBS*, *LaNR11*, *LaN4'OMT*, and *LaCYP96T1* or *LaCYP96T2*, and supplemented with tyramine and 3,4-DHBA. This resulted in the detection of norbelladine, 4'OM, the *o*-*p'* coupled products norpluviine and lycorine, the *p*-*p'* products noroxomaritinidine, normaritinidine, haemanthamine and vittatine/crine, and the *p*-*o'* products *N*-demethylnarwedine and norgalanthamine (Table S11). Lycorine was the major product formed by both *LaCYP96T* enzymes, with production nearly three times higher for *LaCYP96T1* compared to *LaCYP96T2* ( $1.16 \pm 0.077$  ng/mg and  $0.41 \pm 0.016$  ng/mg of biomass, respectively).

While noroxomaritinidine and normaritinidine were consistently detected in all samples, vittatine and haemanthamine were only confirmed when *LaCYP96T1* was included, with a production of 4.85 ng/mL. Norgalanthamine was detected, but only confirmed in 1 out of 3 samples for *LaCYP96T1*, and 2 out of 3 samples for *LaCYP96T2* (Table S11), indicating that *LaCYP96T2* more efficiently yielded norgalanthamine compared to *LaCYP96T1*. In summary, the results support the previous stepwise findings and validate the reconstructed pathway with the four enzymes, as illustrated in Figure S15.

## Discussion

In this study, we identified and characterized the functionality of the key enzymes involved in the biosynthesis of the

pharmaceutically potent AA using a multifaceted approach that included homology-based sequence comparison, phylogenetic reconstruction, subcellular localization studies, docking, and *in planta* assays. The synthesis of norbelladine, the first AA of the pathway, is a critical biosynthetic step. The identified *LaNBS* candidate belongs to the pathogenesis-related 10 (PR10)/Bet v1 protein superfamily, which has gained significant attention for its role in condensation reactions in the initial steps of benzyloquinoline alkaloids (BIA) and AA biosynthesis pathways. A prominent member of this superfamily, (*S*)-norcoclaurine synthase (NCS) catalyses the condensation of dopamine and 4-HPAA to produce (*S*)-norcoclaurine, the first committed step in the biosynthesis of BIA such as morphine and codeine (Samanani *et al.*, 2004; Samanani and Facchini, 2002). A truncated NCS gene from *C. japonica* condensed 4-hydroxydihydrocinnamaldehyde (4-HDCA) and dopamine, derived from L-phenylalanine and L-tyrosine respectively, to yield 1-phenethylisoquinoline, a precursor in the colchicine alkaloid biosynthetic pathway (Nett *et al.*, 2020; Qiao *et al.*, 2023). Similarly, *LaNBS* encodes an enzyme catalysing the condensation of tyramine and 3,4-DHBA into norbelladine (Singh *et al.*, 2018; Tousignant *et al.*, 2022). Although a norcraugsodine/noroxomaritinidine reductase has also been shown to catalyse this step (Kilgore *et al.*, 2016b), the condensation of tyramine and 3,4-DHBA to yield norbelladine in yeast requires both NBS and NR enzymes (Majhi *et al.*, 2023), suggesting that NBS catalyses the condensation to form norcraugsodine, which is then reduced to norbelladine by NR. The thermal unfolding analysis further strengthened the evidence of interaction of the substrates 3,4-DHBA+tyramine with *LaNBS* and *LaNR11*. Docking analysis indicates that the condensation of the amine and the aldehyde is facilitated by catalytic residues Tyr59, Glu71 and Lys83, and stabilized by hydrophobic interactions, consistently with previous studies on *NpNBS*, *LaNBS* and *NpsNBS* (Hnin *et al.*, 2024b; Majhi *et al.*, 2023). Norcraugsodine was positioned in a bent position in the active site of NR to expose its C–N double bond to Tyr173 and NADPH. Similarly to tyramine and piperonal in *NpsNR* crystals, the ligand was further stabilized by interactions with Ile205 (Ala207 in *NpsNR*) and His170 (His172) (Kilgore *et al.*, 2016b). While mutagenesis studies are required to validate the role of these residues, the results emphasize the critical importance of these key amino acids in NBS and NR biosynthesis of norbelladine. Consistent with previous *in vitro* findings and with the current docking analysis, we confirmed the catalysis of norbelladine formation in *in vivo* as well as *in vitro* reactions, as detected by LC–MS/MS. However, the low quantities produced suggest that this step in the AA pathway may be rate-limiting, warranting further optimization to enhance its efficiency.

Norbelladine then undergoes 3' or 4'-O-methylation as a critical step for diversification into various AAs (Figure 1). Some plant OMTs, particularly those involved in phenylpropanoid and flavonoid biosynthesis, often exhibit substrate versatility (Liscombe *et al.*, 2012). *In vitro* studies in *N. sp. aff. pseudonarcissus* have shown that OMT regiospecifically methylates norbelladine at the 4'O position (Kilgore *et al.*, 2014). However, the OMT from *N. papyraceus*, *Galanthus elwesii* and *Lycoris* species are promiscuous, methylating at the 3' or 4' position depending on the ligand, and on the metal ion (Koirala *et al.*, 2024; Li *et al.*, 2019; Su *et al.*, 2022; Sun *et al.*, 2018). Sequence comparison and docking analysis suggest that the *LaN4'OMT* candidate identified in this study likely yields 4'-O-methylnorbelladine (4'OM), with norbelladine 4'O to face the SAM methyl group in interaction with  $Mg^{2+}$ , Asp155 and Lys158. *In planta* agroinfiltration



confirmed the regiospecificity, yielding 4'OM as the sole product and similar results were obtained *in vitro* while using  $Mg^{2+}$  as ion. Intriguingly, the *LaN4'OMT* enzyme also contained specific residues that were previously associated with the 3'-O-methylation of norbelladine (Val30, Gln141). As  $Ni^{2+}$  was shown to reverse the regioselectivity of *G. elwesii* class I OMT triple mutant (Su et al., 2022), this ion was selected for further *in vitro* experiments. Its addition instead of  $Mg^{2+}$ ,  $Ni^{2+}$  resulted in the formation of 3',4'-O-dimethylnorbelladine from norbelladine catalysed by *LaN4'OMT*, although its level was 250-fold lower than 4'OM. Additionally, *para*- and *meta*-methylated derivatives of 3,4-DHBA, yielding vanillin and isovanillin, were observed. Mutagenesis studies revealed that Ser52 and Tyr186 play pivotal roles in determining *para* versus *meta*-O-methylation preferences (Hnin et al., 2024a). Further exploration of combinations of mutations and divalent metal ions is needed to gain insight into regioselectivity to increase dimethylated product yields. For the first time, co-infiltration of *LaNBS*, *LaNRll* and *LaN4'OMT*, along with supplementation of tyramine and 3,4-DHBA, led to the detection of norbelladine and 4'OM, reconstituting the foundational steps of the pathway *in vivo*.

At this critical branching point in the biosynthetic pathway, 4'OM undergoes three potential phenol couplings to generate skeletal frameworks that give rise to over 650 known AAs. Phenol couplings catalysed by CYPs in the BIA pathway of several plants, such as *Gloriosa superba*, *Berberis stolonifera*, *C. japonica*, *P. somniferum*, and *Argemone mexicana* lead to the production of pharmaceutically potent compounds. The discovery of *NpsCYP96T1* in *N. sp. aff. pseudonarcissus* marked a significant advancement in AA biosynthesis, as it represented the first enzyme of its kind identified in a monocot. *NpsCYP96T1* catalyses the ring closure 4'OM, strongly favouring the formation of the *p-p'* coupled product, noroxomaritidine, over the *p-o'* product, *N*-demethylnarwedine (Kilgore et al., 2016a). Further studies uncovered other *Amaryllidaceae* CYP96Ts with specific coupling preferences (Liu et al., 2024; Mehta et al., 2024). In this study, we identified and examined orthologs *LaCYP96T1* and *LaCYP96T2*, which were localized to the ER membrane. Membrane-bound proteins, including cytochromes P450, are proposed to act as nucleation factors that orchestrate the assembly of metabolons at the endoplasmic reticulum membrane interface (Bassard et al., 2012). Consistently, our localization results suggest that *LaCYP96T1* and *LaCYP96T2* may recruit additional cytosolic metabolic enzymes, such as *LaNBS*, *LaNRll* and *LaN4'OMT*, to form a metabolon, although further studies are required to confirm this hypothesis.

Phylogenetic analysis indicated that the candidates clustered with the *o-p'* coupling enzyme *NtCYP96T5* (Mehta et al., 2024). Docking studies revealed that heme interacted with Leu123 and 4'OM with Phe124, and Val307 among other key residues in the enzyme catalytic pocket. Leu123 and Phe124 have been associated with either *p-p'* or *o-p'* coupling (Liu et al., 2024). Mutations of residues at both positions were shown to impact coupling specificity, with Ile123 and Leu124 specific to *p-o'* and *p-p'* coupling catalysis, but absent from *o-p'* coupling enzymes, while Leu123 and Phe124 could not yield *p-o'* coupled products (Liu et al., 2024). By contrast, Val307, predicted to interact with water or with the oxygen of the substrate phenol ring, seems to be compatible with the three types of couplings. The analyses of the active site sequence and docking of the candidates suggested that the enzymes could catalyse *o-p'* and *p-p'* reactions. Intriguingly, like *NtCYP96T5* from Mehta et al., *LaCYP96T1* and

*LaCYP96T2* were not active *in vitro* when expressed in yeast (data not shown). While *NtCYP96T5* activity has not been verified *in vitro*, the isolation of microsomes from *N. benthamiana* leaves agroinfiltrated with *LaCYP96T1* was successful and resulted in the formation of minor *p-p'* and major *o-p'* products from 4'OM, when co-incubated with the reductase (*LaNRll*) purified from bacteria. This *in vitro* assay confirmed the *in silico* predictions, demonstrating that *LaCYP96T1* and *LaCYP96T2* can yield various coupled products, although *p-o'* derived alkaloids were not detected.

*In planta*, upon feeding 4'OM, *LaCYP96T1* and *LaCYP96T2* catalysed the formation of noroxomaritidine and *N*-demethylnarwedine. *LaCYP96T2* yielded a detectable amount of lycorine as well. The addition of reduction steps showed that both enzymes could yield the three coupled products, suggesting that they are promiscuous with regard to the product they can form. There is growing evidence that CYPs exhibit significant substrate and product promiscuity (Hansen et al., 2021). However, this notion of multifunctionality appears somewhat contradictory to recent studies by Mehta et al. (2024) and Liu et al. (2024), where specific amino acid sequences were reported to be associated with exclusive coupling patterns. In these studies, the residues involved in substrate binding were consistent with those found in *LauCYP96T1-like2* and *NtCYP96T5* transcripts, which yielded *p-p'* coupled product (haemanthamine pathway) and *o-p'* coupled products (lycorine pathway), respectively. Despite these similarities, *LaCYP96T1* and *LaCYP96T2* catalysed the formation of all three couplings, with *o-p'* as a major reaction product. Mehta et al. suggested Ala308 in *NtCYP96T1* was required for *p-p'* coupling and Leu308 for the *p-o'* coupling by *NtCYP96T6* by performing a mutagenesis study whereas *o-p'* coupling *NtCYP96T5* contained Val308. This requirement, however, may be context-dependent and not applicable to sequences from other species as demonstrated by *LauCYP96T1-like2*, *LauCYP96T1-like3* and their mutants, which catalysed *p-p'* and *p-o'* products formation, despite displaying a valine (Val307) at the corresponding position (Table S10) (Liu et al., 2024). Docking studies further highlighted the relevance of conserved Lys109, Leu112, Phe116, Leu220, Leu224, Thr311, and Leu495 in substrate stabilization. However, complementary residues that enable promiscuity in product formation for *LaCYP96T1* and *T2* remain unknown at this stage. Notably, the cloned sequences displayed a unique glutamine at position 251 in the vicinity of the catalytic site compared to all other characterized CYPs, which could play a role (Figure S4b). Further mutagenesis studies that explore single and combinations of amino acids will be necessary to reveal the precise requirements for substrate positioning to yield the different couplings.

The current results also highlight the promiscuity of *LaNRll*. Docking studies suggested that *LaNRll*, which previously reduced norcraugosidine to norbelladine accepted noroxopluiene, *N*-demethylnarwedine, and noroxomaritidine (in two positions) with similar scores in its active sites, conserving the key interaction with Tyr173. Co-infiltration of *LaCYP96T1* or *LaCYP96T2* with *LaNRll* supplemented with 4'OM yielded six detectable products from the three coupling branches, namely *p'-o* *N*-demethylnarwedine, the *p-p'* coupled product, noroxomaritidine, oxomaritidine and normaritidine, together with the *o-p'* phenol coupling norpluiene and lycorine. These results, confirmed by *in vitro* assays, revealed that *LaNRll* is involved in the synthesis of all three types of phenol-coupled products in the AA biosynthetic pathway. The ability of *LaNRll* to reduce all three

phenol-coupled products at different positions is novel, but consistent with findings for *NpsNR* from *N. sp. aff. pseudonarcissus*, and *ZtSDR* from *Zephyranthes treatiae*, which reduced imines and ketones (Kilgore *et al.*, 2016b; Roth *et al.*, 2018). Similarly, in tropane alkaloid biosynthesis, SDR presented substrate flexibility, as seen in *Cochlearia officinalis* SDR (CoTR) and *ThTRL* from *Tarenaya hassleriana* (Li *et al.*, 2022; Reinhardt *et al.*, 2014).

Short-chain dehydrogenase/reductase catalyse the interconversion of hydroxyl and oxo groups in substrates using a conserved Ser-Tyr-Lys catalytic triad (Jornvall *et al.*, 1995). Tyrosine functions as the central acid–base catalyst in its deprotonated tyrosinate state, supported by lysine, which lowers tyrosine's pKa and aligns NADP<sup>+</sup> for hydride transfer (Benach *et al.*, 1999). The serine stabilizes the substrate and intermediates via hydrogen bonding. (Brock *et al.*, 2008; Chen *et al.*, 1993; Gani *et al.*, 2008; Kavanagh *et al.*, 2008). In the *LaNRII* enzyme, Ser159, Tyr173 and Lys177 are predicted to form the catalytic triad and surround the NADP ring. Moreover, Tyr173 directly interacted with the ligands. The role of additional residues, such as His170 and Glu224 in various substrates reduction requires further investigation. Compared to *NpsNR* and *NtSDR1*, *LaNRII* encodes an alanine instead of Ser174, between residues of the catalytic triad, along with a 9% and 29% difference in amino acids, respectively. While the promiscuity of *NtSDR1* has not been assessed, these sequence differences may contribute to variations in substrate specificity. However, further mutagenesis studies are recommended to explore the promiscuity of *LaNRII*.

Finally, whole pathway co-infiltration (*LaNBS*, *LaNRII*, *LaN4'OMT* and *LaCYP96T1* or *LaCYP96T2*), along with precursors tyramine and 3,4-DHBA, yielded lycorine as the major product, with *LaCYP96T1* being almost three times more efficient than *LaCYP96T2*. While norpluviine likely resulted from noroxopluiine reduction by *LaNRII*, the synthesis of lycorine requires, in theory, more enzymatic steps, such as oxygen bridging and hydroxylation (Desgagné-Penix, 2021; Jayawardena *et al.*, 2024), possibly carried out by endogenous enzymes in *N. benthamiana* as observed in other studies (Boccia *et al.*, 2022; Caputi *et al.*, 2018; Dong *et al.*, 2013; van Herpen *et al.*, 2010). In addition to lycorine, Norgalanthamine was observed in leaves co-infiltrated with *LaNBS*, *LaNRII*, *LaN4'OMT*, *LaCYP96T1* or *LaCYP96T2*, supplemented with tyramine and 3,4-DHBA, confirming the possible reduction of *N*-demethylnarwedine.

Infiltration of the entire pathway also resulted in the detection of trace amounts of crinine and haemanthamine when *LaCYP96T1* was included, whereas these products were not detected with *LaCYP96T2*. This finding suggests that *LaCYP96T1* is more efficient not only in producing lycorine but also in driving the biosynthesis of crinine/haemanthamine, whereas *LaCYP96T2* primarily yielded lycorine and norgalanthamine. The proposed biosynthetic pathway for haemanthamine suggests that a CYP71 family enzyme facilitates the methyl bridging, leading to the formation of vittatine/crinine. Subsequently, vittatine/crinine undergoes hydroxylation to produce 11-hydroxyvittatine, followed by *O*-methylation. These transformation steps are likely carried out by endogenous enzymes present in *N. benthamiana*, as has been reported in the reconstitution of the MIA pathway for strictosidine biosynthesis (Miettinen *et al.*, 2014). Although the activity of endogenous enzymes in *N. benthamiana* may complicate the precise elucidation of the pathway, it offers a potential advantage by reducing the reliance on heterologous enzyme expression for biotechnological applications. However, the

competition between these endogenous enzymes and the specific enzymes identified from *L. aestivum* needs to be verified. Future studies should focus on co-infiltrating the genes encoding *N*-methyltransferases, hydroxylases, and enzymes involved in oxygen or methyl bridging, alongside their functional characterization in heterologous hosts such as yeast, to provide deeper insight into these biosynthetic pathways.

## Conclusion

In conclusion, our study elucidates the enzymatic steps involved in the biosynthesis of *Amaryllidaceae* alkaloids, demonstrating that four enzymes, *NBS*, *NR*, *N4'OMT*, and *CYP96T*, are sufficient to yield the three types of phenol coupling reaction in *N. benthamiana*, ultimately producing the potent antiviral and anticancer alkaloid, lycorine. Furthermore, our results highlight the substrate as well as product promiscuity of *LaNRII*, *LaCYP96T1* and *LaCYP96T2*. The investigation of the residues responsible for this promiscuity may pave the way towards understanding enzyme flexibility, developing novel biocatalysts and enhance our knowledge of evolutionary processes in protein function. Overall, we provide robust evidence for the synthesis of *Amaryllidaceae* alkaloids *in planta* and deliver new knowledge for metabolic engineering approaches to achieve sustainable production of SMs through synthetic biology. The elucidation of the biosynthetic pathway of *Amaryllidaceae* alkaloids in *L. aestivum* represents a significant advancement in our understanding of plant-specialized metabolism.

## Materials and methods

### Database of Amaryllidaceae transcriptomes

We constructed a database of *Amaryllidaceae* transcriptomes by downloading either raw read files (fastq) or assembly files (fasta) publicly available. If multiple datasets were available for the same species and cultivar, we selected the one for the assembly file was available. For species with multiple transcriptome assemblies available in NCBI Transcriptome Shotgun Assembly, we used the following criteria to select the assembly to include in the database: most recent, publication information available, best assembly statistics (i.e. lowest contig count, longest N50 length). If only the raw datasets were available, we selected the one associated with a peer review publication, and used all data available if all were peer-reviewed. Species, accession numbers and, when available, links to the original assembly files are provided in Table 1.

The raw reads were first processed with Fastp (v0.23.1) (Chen *et al.*, 2018) software with cut front mean quality 20, cut tail mean quality 20 and length required 50. The resulting clean reads were assembled with Trinity version 2.14.0 (Grabherr *et al.*, 2011) using default parameters. The completeness of the transcriptome assembly was assessed with BUSCO (v5.2.2) (Simao *et al.*, 2015) using liliopsida\_odb10. To limit the technical differences between the assemblies, all annotations were done using the same pipeline, even when predicted protein sequences and annotations were available. We used the Trinotate pipeline version 4.0.0 (<http://trinotate.github.io/>) (Bryant *et al.*, 2017) with: TransDecoder version 5.5.0 (<http://transdecoder.github.io/>) was used for prediction coding sequences; HMMer version 3.2.2 (Eddy, 2008, 2011) with Pfam-A database v3.1b2 was used to find protein family domains; prediction of transmembrane domains was done with TmHMM version 2.0c (Krogh *et al.*, 2001); identification of noncoding RNA was performed with Infernal version 1.1.4

**Table 1** List of species included in the database, with accession numbers for raw data in BioProject or BioSample and, when available, accession numbers or names of assembly files

Species	Reference	Raw data accession	Assembly accession
<i>Amaryllis belladonna</i>	One Thousand Plant Transcriptomes Initiative (2019)	ERS1829785	LDME <sup>†</sup>
<i>Clivia miniata</i>	Li et al. (2022), Wang et al. (2018)	PRJNA480383, PRJNA813401	
<i>Crinum × powellii</i>	Koirala et al. (2023)	PRJNA962562	
<i>Galanthus elwesii</i>	Kilgore et al. (2016)	PRJNA306273	G_elwesii_Trinity.tar.gz <sup>‡</sup>
<i>Galanthus</i> sp. MBK-2015	Kilgore et al. (2016)	PRJNA306697	Galanthus_Trinity.tar.gz <sup>‡</sup>
<i>Hippeastrum vittatum</i>	Zeng et al. (2024)	PRJNA862291	
<i>Hippeastrum × hybridum</i>	Li et al. (2022)	PRJNA608969	
<i>Leucojum aestivum</i>	Tousignant et al. (2022)	PRJNA720900	
<i>Lycoris chinensis</i>	Zhang et al. (2022b)	PRJNA847051	
<i>Lycoris longituba</i>	Li et al. (2020)	PRJNA590043	
<i>Lycoris radiata</i>	Wang et al. (2021)	PRJCA006232	CRA004779 <sup>§</sup>
<i>Lycoris sprengeri</i>	Yang et al. (2021)	PRJNA714286	
<i>Narcissus papyraceus</i>	Hotchandani et al. (2019)	PRJNA407433	
<i>Narcissus pseudonarcissus</i> cv King Alfred	Singh and Desgagné-Penix (2017)	PRJNA392294	
<i>Narcissus</i> aff. <i>pseudonarcissus</i>	Kilgore et al. (2016)	PRJNA301357	narcissus_9-20_101_112.tar <sup>‡</sup>
<i>Narcissus tazetta</i> var. <i>chinensis</i>	He et al. (2020)	PRJNA523125	GSE126727 <sup>§</sup>
<i>Narcissus viridiflorus</i>	One Thousand Plant Transcriptomes Initiative (2019)	ERS1829787, ERS1829786	TRRQ <sup>†</sup>
<i>Phycella</i> aff. <i>cyrtanthoides</i>	One Thousand Plant Transcriptomes Initiative (2019)	ERS3670342	DMIN <sup>†</sup>
<i>Rhodophiala pratensis</i>	One Thousand Plant Transcriptomes Initiative (2019)	ERS1829788	JDTY <sup>†</sup>
<i>Traubia modesta</i>	One Thousand Plant Transcriptomes Initiative (2019)	ERS1829789	ZKPF <sup>†</sup>
<i>Zephyranthes candida</i>	Zhang et al. (2022a)	PRJNA796382	
<i>Zephyranthes treatiae</i>	One Thousand Plant Transcriptomes Initiative (2019)	ERS1829790	DPFW <sup>†</sup>

<sup>†</sup><http://gigadb.org/dataset/100910>.<sup>‡</sup><https://medplantmseq.org/>.<sup>§</sup>Downloaded from the China National Center for Bioinformatics <https://download.cncb.ac.cn/gsa/>.<sup>†</sup>Assembly available for download at NCBI Gene Expression Omnibus.

(Nawrocki et al., 2009); and Blast+ version 2.13 (Camacho et al., 2009) against the UniProt Swiss-Prot database (release 2024\_04). Parameters for each software were set following Trinotate suggestions. For further analysis, only sequences with no hits blast hits (against UniProt Swiss-Prot) or with hits to Viridiplantae proteins were kept in the database, the remaining were considered contaminants.

### Phylogenetic analysis

The candidate transcripts of *L. aestivum* selected in this study include homologous (norbelladine/norcrugsodine synthase [NBS] and noroxomatidine/norcrugsodine reductase [NR], (Majhi et al., 2023)) and orthologous sequences (4'-O-methyltransferase [N4'OMT] and noroxomaritidine synthase CYP96Ts), (Hnin et al., 2024a; Kilgore et al., 2016a; Koirala et al., 2024) of previously characterized enzymes pulled from transcriptome database. To find putative NBS, NR, N4'OMT and CYP96 in *L. aestivum*, we used the sequences of the enzymes characterized previously in *Papaver somniferum*, *N. pseudonarcissus* and *N. aff. pseudonarcissus* (UniProt accessions: NCS\_PAPSO, NR\_NARAP, NOMB\_NARAP and C96T1\_NARAP, respectively) as query in BLAST searches (blast+/2.14.0 (Camacho et al., 2009), e-Value <0.001, identity >30%) of the transcriptome of *L. aestivum*. Predicted sequences contained both start and stop codons. The candidate transcripts of *L. aestivum* were aligned against PR10/BetIV family proteins (in the case of LaNBS), short chain dehydrogenase/reductases (for LaNR), O-methyltransferases (for LaN4'OMT) and cytochrome P450 proteins (for LaCYP96T)

obtained from UniProt, SwissProt, or the National Center for Biotechnology Information (NCBI) databases (accession numbers available in Table S1) using ClustalW (Sievers et al., 2011) (default parameters). Catalytic sites and conserved motifs were identified on the basis of published literature or NCBI conserved domain search tool or InterPro database following the alignment (Paysan-Lafosse et al., 2023). The percentage identity matrix was calculated in UGENE software (Okonechnikov et al., 2012) and the multiple sequence alignment was viewed in Jalview software (Waterhouse et al., 2009). Phylogenetic trees were constructed using the ModelRevelator software tool "Iqtree" with 1000 ultrafast bootstraps, 1000 branch tests, as well as 1000 with a minimum correlation coefficient of 0.99, followed by standard IQ tree search parameters (Hoang et al., 2018; Kalyanammoorthy et al., 2017; von Haeseler et al., 2014). ModelRevelator is a model selection tool underpinned by two deep neural networks. The models resulting from neural network estimates are closer to the ground truth than maximum likelihood trees (Burgstaller-Muehlbacher et al., 2023). iTOL was used for tree visualization and representation (Letunic and Bork, 2021).

### Molecular modelling

Structural prediction of *L. aestivum* enzymes including norbelladine 4'-O-methyltransferase (LaN4'OMT), norbelladine/norcrugsodine synthase (LaNBS), noroxomaritidine/norcrugsodine reductase (LaNR) and cytochrome P450 LaCyp96T1 and cytochrome P450 LaCyp96T2 was performed using CollabFold v1.5.3 (AlphaFold2 via MMseqs2) (Mirdita

et al., 2022). Models were visualized and superimposed using Pymol 3.0.0 (Schrödinger). MOE2022.09 software (Chemical Computing Group) was employed for analysing model conformations and preparing receptors for docking, as outlined in (Majhi et al., 2023). Pocket residues of predicted enzymes were superimposed with reference crystal structures in MOE. Crystal structures of *Rattus norvegicus* and *Homo sapiens* CatOMTs (PDB:1H1D; (Bonifacio et al., 2002) and 3A7E, respectively) were used as references for LaN4'OMT, norcoclaurine synthase from *Thalictrum flavum* for LaNBS (PDB:2VQ5; (Ilari et al., 2009)), *N. pseudonarcissus* King Alfred noroxomaritidine/norcraugsodine reductase for LaNR (PDB:5FF9; (Kilgore et al., 2016b)) and *Priestia megaterium* CYP102A1 (PDB:6JMH; (Stanfield et al., 2020)) and *Mycobacterium tuberculosis* cyclo(l-Tyr-l-Tyr) C-C coupling CYP121 (PDB:3G5H; (Belin et al., 2009; Dumas et al., 2014; Peng et al., 2023)) for LaCYP96T1 and LaCYP96T2. Methyl donor S-adenosylmethionine (SAM for LaN4'OMT), metal ion ( $Mg^{2+}$  for LaN4'OMT), cofactor nicotinamide adenine dinucleotide phosphate (NADPH for LaNR), as well as heme and two vicinal  $H_2O$  (LaCyp96T1 and 2) were included in the active sites at coordinates of orthologous crystal structures. Structures underwent preparation including correcting issues, capping, charging termini, selecting appropriate alternate, protonation, calculating optimal hydrogen position and charges and energy minimization with tethered active site residues and fixed NADPH,  $H_2O$  and heme, SAM and  $Mg^{2+}$ , in corresponding enzymes, before docking. The docking site was predicted based on residues in interaction with ligands of orthologous crystal structures, as described in (Majhi et al., 2023), or using dummy atoms added to the receptor with the MOE Site Finder tool. The docking site included  $Mg^{2+}$  and SAM,  $H_2O$  and heme, or NADPH, where applicable. Ligands (3,4-DHBA, tyramine (LaNBS), norcraugsodine (LaNRll), norbelladine (LaN4'OMT) and 4'-O-methylnorbelladine (LaCYP96T1 and LaCYP96T2)) isomeric smiles codes were retrieved from PubChem and submitted to the ZINC20 database to download 3D data files (Irwin et al., 2020). Every protomer predicted at pH = 7 was included as a possible ligand.

The AMBER10:EHT force field was employed. Triangle Matcher was used as the placement method for 200 poses and flexible docking was performed with tethered induced fit refinement, yielding 10 best poses. Docking poses were analysed by comparison with crystalized protein-ligand complexes, and the first pose (with the best docking score) coherent with crystalized enzyme ligand positioning and with the catalysed reaction was selected for each ligand. PLIP was utilized to analyse the interactions between ligands and receptor residues (Adasme et al., 2021), and the images were further processed using PyMOL along with the prediction of hydrogen bonds between the metal ion or the water molecules and the docked ligand, where applicable.

### RNA extraction and cDNA synthesis

*Leucojum aestivum* bulbs were purchased from The Seed Company, Canada and grown in the laboratory plant room. Four centimetres of the apical part of a 1-month-old newly emerging leaf was cut, immediately ground in liquid nitrogen by mortar and pestle and stored at  $-80^{\circ}C$ . Fifty milligrams of grinded leaf tissue was used for targeted metabolite analysis with a HPLC-MS/MS to identify key metabolites, that is, 4'-O-methylnorbelladine, norgalanthamine and galanthamine. The remaining leaf tissue was used for subsequent RNA extraction. RNA was extracted from 80 mg snap frozen tissue with the RNeasy® Plant Mini Kit

(Qiagen) following the manufacturer's protocol without the optional steps. cDNA synthesis was performed using the SensiFAST cDNA synthesis kit by Meridian Bioscience. Briefly, the SensiFAST cDNA Synthesis Mix containing random hexamers and anchored oligo-dT primers was mixed with RNA, followed by the addition of reverse transcriptase and reaction buffer, to be then incubated at  $42^{\circ}C$ . After the reaction, the enzyme was inactivated by heating at  $85^{\circ}C$  for 5 min. The resulting cDNA was then stored at  $-80^{\circ}C$ .

### PCR amplification of key candidate genes LaNBS, LaNRll, LaN4'OMT, LaCYP96T1 and LaCYP96T2

The coding sequences of norbelladine synthase (LaNBS), noroxomaritidine reductase (LaNRll), 4'-O-methyltransferase (LaN4'OMT) and cytochrome P450 (LaCYP96T1 and LaCYP96T2) of *L. aestivum* were amplified using primers designed with attB1 and attB2 overhangs for Gateway cloning in expression vector pK7WG2 (Karimi et al., 2007). Primers for LaNBS, LaNRll, LaN4'OMT and LaCYP96T1/2 were designed for expression with  $\times 3$  AU1,  $\times 3$  FLAG,  $\times 3$  AU5 and  $\times 3$  Myc tag at their C-terminal, respectively, tag sequences were included in the reverse primers (Table S2). Primers for localization study of the candidate gene LaN4'OMT were designed for cloning into expression vectors pB7WGF2 (N-terminal EGFP fusion) and pB7FWG2 (C-terminal EGFP fusion) vectors, while LaCYP96T2 primers were designed to clone into pB7FWG2 (C-terminal EGFP fusion) (Table S2). The coding sequences of LaNBS, LaNRll, LaN4'OMT were also amplified with primers containing restriction sites BamHI and HindIII to clone into pMAL-c2X plasmid. PCR amplifications were conducted PrimeStar GXL premix (TaKaRa Bio) in T100 Thermal cycler of Bio-Rad. PCR program parameters were 5 min  $98^{\circ}C$  1 cycle, 10 s  $98^{\circ}C$ , 30 s ( $55^{\circ}C$ – $65^{\circ}C$ ; primer specific), 1 min/kb  $72^{\circ}C$  for 34 cycles, 5 min  $72^{\circ}C$  1 cycle and final infinite hold at  $4^{\circ}C$ . Amplification was checked by running 2  $\mu L$  of PCR product in 0.9% agarose gel and images were taken in Bio-Rad ChemiDoc Imaging system. PCR clean-up and gel purification were done according to manufacturer's protocol by GeneAid-PCR & Gel Kit.

### Cloning and transformation

Amplified PCR fragments with attB1 and attB2 were cloned with Gateway cloning according to manufacturer's protocol (Thermo Fisher Scientific). Briefly, the Gateway-adapted attP-flanked pDONR 221 vectors with kanamycin resistance gene were used for BP recombination reaction to generate an attL-flanked entry clone with attB-flanked-LaNBS, -LaNRll, -LaN4'OMT, -LaCYP96T1, -LaCYP96T2, -LaN4'OMT (C-terminal EGFP fusion), -LaN4'OMT (N-terminal EGFP fusion) and -LaCYP96T2 (C-terminal EGFP fusion) DNA fragment. These entry vectors were transformed into *Escherichia coli* DH5 $\alpha$  chemicompetent cells, subsequently grown on a kanamycin (50  $\mu g/mL$ ) LB agar plate. Following colony PCR, Miniprep of positive clones was performed according to the manufacturer's protocol of Presto Mini Plasmid Kit from Geneaid. A positive entry clone of each construct was sent for Sanger sequencing (Genome Quebec). Once integrity was confirmed, the selected plasmids were incubated overnight with LR reagents to perform a LR recombination reaction between an attL-containing entry clone and attR-containing destination vectors (pK7WG2, pB7FWG2 and pB7WGF2 vectors). The LR reaction mix was transformed into *E. coli* DH5 $\alpha$  chemicompetent cells and grown in spectinomycin (50  $\mu g/mL$ ) agar plates. The positive clones were confirmed by colony PCR, and plasmid



extraction was done as mentioned above, growing positive colonies in 5 mL LB broth with spectinomycin 50 µg/mL. These expression clones were transformed in *Agrobacterium* GV3101. After 48 h of incubation in LB plates at 28 °C, the presence of insert in the construct was once again confirmed with colony PCR.

Amplified PCR fragments of *LaNBS*, *-LaNRll*, *-LaN4'OMT* along with the vector pMALc2X were digested by BamHI and HindIII. Fragments were ligated using T4-DNA ligase (New England Biolabs) and transformed in chemically competent *E. coli* DH5α cells leaving them to grow in ampicillin (100 µg/mL) agar plates for each gene. Colony PCR and sequencing were carried out for verification of the integrity of each gene in the plasmid. Plasmids were further transformed in *E. coli* Rosetta (BL21) and grown on ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL) agar plate.

### Transient expression of candidate genes in *N. benthamiana* (in vivo assay)

Individual *Agrobacterium* GV3101 colonies, carrying the desired gene constructs (*LaNBS*, *LaNRll* and *LaN4'OMT*, *LaCYP96T1* and *LaCYP96T2*), were streaked on LB plates with 50 µg/mL spectinomycin, 50 µg/mL rifampicin and 30 µg/mL gentamicin and grown at 28 °C for 2 days. With the cell scraper, colonies were scraped from the plate, resuspended in *Agrobacterium* induction buffer (10 mM MES, pH 5.6, 10 mM MgCl<sub>2</sub> and 150 µM acetosyringone) and incubated for 2 h at room temperature. 4–5-week-old *N. benthamiana* grown in a growth chamber with 22 °C temperature and 16 h of light/8 h of dark conditions were used for the experiments. *N. benthamiana* leaves were infiltrated or co-infiltrated with *Agrobacterium* suspensions of 0.6 of optical density at 600 nm (OD<sub>600</sub>) on the leaf abaxial side using a needleless syringe. Unless otherwise mentioned, three leaves each of three independently infiltrated plants of each experimental condition were used to test the hypothesis. At first, *Agrobacterium* suspensions harbouring the genes were infiltrated in the plant leaves and incubated for 48 h. In the second step, leaves were infiltrated with 100 µM substrates. Following an incubation of 24 h, the leaves were harvested, snap-frozen and preserved at –80 °C for further analysis. The snap-frozen samples were crushed into powder using 5-mm diameter stainless steel beads, shaking at 25 Hz for 1.5 min in TissueLyser II (Qiagen).

The expression of enzymes after agroinfiltration of genes of interest in *N. benthamiana* leaves was verified by RT-qPCR and western blot. Briefly, 100 mg of tissue was resuspended in 0.5 mL of lysis buffer (100 mM Tris-HCl, pH 7.5, 1 mM phenylmethylsulfonyl fluoride, 10% glycerol, 10 mM β-mercaptoethanol and 2% polyvinylpyrrolidone) and incubated for 30 min on ice with periodic, gentle inversion. Extracts were centrifuged at 12 000 *g* at 4 °C for 30 min to pellet debris, after which the supernatant was collected in 3 tubes and preserved at –80 °C for future experiments. Specific primary antibodies were used to verify expression on PVDF membranes after semi-dry transfer and blocking by 5% milk overnight at 4 °C. For *LaNBS* and *LaN4'OMT*, goat anti-AU1 and anti-AU2 polyclonal antibodies from Thermo Fisher Scientific were employed, respectively. A rabbit anti-FLAG monoclonal antibody from Cell Signalling Technology (NEB) was used for *LaNRll*, while an anti-Myc polyclonal antibody raised in rabbit from Millipore Sigma-Aldrich was applied for *LaCYP96T1* and *T2*. For the *LaN4'OMT*-EGFP fusion protein and *LaCYP96T2*-EGFP fusion protein, a primary anti-GFP/CFP/YFP monoclonal antibody raised in mouse was used. Host-specific secondary antibodies conjugated with Horseradish peroxidase

(HRP) were utilized to detect the bands in the membrane, that is, Immun-Star goat anti-mouse (GAM)-HRP conjugate and Immun-Star goat anti-rabbit (GAR)-HRP conjugate from BioRad, and anti-goat-HRP from Jackson ImmunoResearch Laboratories.

Further validation of each gene expression was conducted by RT-qPCR. RNA was isolated from 200 mg of snap-frozen samples. Seven hundred microlitres of Trizol were added and samples vortexed. Three hundred microlitres chloroform was added, samples were vigorously shaken and incubated on ice for 3 min, then centrifuged at 12 000 × *g* at 4 °C. The mixture separated into a lower red phenol-chloroform and interphase, and a colourless upper aqueous phase. The upper phase was transferred to a new tube containing 35 µL of NaCl (5 M) and 400 µL Isopropanol and incubated at –20 °C for 10 min. The mixture was centrifuged at 12 000 × *g* at 4 °C, and the supernatant was discarded, leaving the total RNA pellet at the bottom. The pellet was washed twice with 1 mL of 75% ethanol, centrifuged at 7500 × *g* at 4 °C and air dried for 10 min. DNase treatment was performed using TurboDNase according to the standard protocol from the supplier (Invitrogen TURBO DNA-free Kit). RT-qPCR was conducted using Luna® Universal One-Step RT-qPCR Kit Protocol (NEB). The threshold cycle (Ct) value of each gene was normalized against the Ct value of the reference gene *histone 3*. The relative gene expression levels were determined using the comparative ΔΔCt method by calculating the average Ct values obtained from technical triplicates of pooled biological triplicate samples. The obtained results were analysed and visualized in Prism.

### Protein expression, and purification

*Escherichia coli* Rosetta (BL21) cells harbouring pMAL-c2X-*LaNBS*, pMAL-c2X-*LaNRll* and pMAL-c2X-*LaN4'OMT* were grown in LB broth with ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL) overnight as seed culture. The next morning, 200 mL of medium was added, and bacteria were grown at 37 °C with 210 rpm till the OD<sub>600</sub> reached 0.5–0.6. For induction of protein, 0.25 mM isopropyl-β-D-thiogalactopyranoside (IPTG), 50 µM zinc chloride (ZnCl<sub>2</sub>) and 50 µM magnesium chloride (MgCl<sub>2</sub>) were added and cultures further incubated at 18 °C, 150 rpm. After 16 h, cells were harvested at 5000 *g* and protein extraction was performed following sonication by adding ice-chilled protein extraction (PE) buffer [30 mM Tris-HCl (pH:8), 150 mM NaCl, 1 mM EDTA, 10% Glycerol, 1 mM PMSF and 0.1% protease inhibitor cocktail (Cell Signalling Technology, NEB)]. Sonication parameters were 40% amplitude for a total of 8 min with 15 s run and 30 s pause. The lysates were centrifuged for 60 min at 16 000 *g* at 4 °C. Supernatant was mixed with 1 mL of three times pre-washed amylose resin, kept on a rocking shaker for 2 h and subsequently passed through a filter column to isolate resin. The resin was washed three times with PE buffer under gravity at 4 °C and the flowthrough was discarded. One millilitre of PE buffer containing 25 mM maltose was added to resin and pure protein was collected. Purified proteins were further washed three times with buffer containing Tris-HCl (pH:8) and 20% glycerol using Amicon® Ultra Centrifugal Filter (Millipore), 10 kDa at 2000 *g* for 10–15 min at 4 °C. Finally, the pure, concentrated and desalted proteins were aliquoted in individual tubes and preserved in –80 °C.

For the *in vitro* reaction of cytochrome enzymes, microsomes were isolated from *N. benthamiana*. As aforementioned in 2.6, CYP96T1/2 were extracted from 9 leaves of 3 biological triplicates of *N. benthamiana*. The 20 mL of supernatant was kept in a

Beckman Coulter polycarbonate centrifuge tube (25 × 89 mm) and ultracentrifugation was performed at 100 000 *g* for 90 min. Supernatants were removed slowly and the microsomes on the bottom of the tubes were washed by softly resuspending in 20 mL ice chill Tris–HCl (pH 8) and 20% glycerol. After repeating the centrifugation and washing twice, the supernatant was discarded and the pelleted microsomes were resuspended in pre weighted tube with 1 mL of Tris–HCl (pH 8) with 20% glycerol, aliquoted in individual tubes and preserved in –80 °C.

### *In vitro* enzymatic assay

*In vitro* reaction for purified LaNBS, LaNR1I, was carried out using 80 µg of purified NBS enzyme, 60 µg of purified NR enzyme, 10 µM tyramine, 300 µM 3,4-DHBA and 1 mM NADPH in 50 mM potassium phosphate buffer (pH 7.8), in a total volume of 50 µL, incubated at 35 °C for 2 h and quenched at 65 °C for 20 min. The reaction was alternatively conducted using 200 µg of crude enzyme extracted from *N. benthamiana* leaves along with 100 µM tyramine, 100 µM 3,4-DHBA and 1 mM NADPH in 50 mM potassium phosphate buffer (pH 7.8). To illustrate the formation of 4'OM, 5 µM of LaN4'OMT enzyme and 2 mM SAM along with 1 mM nickel or magnesium were incubated with 100 µM norbelladine in a total reaction volume of 200 µL of sodium phosphate buffer (pH 7.4) at 45 °C for a half hour or as indicated, and quenched by equal volume of 100% methanol (Koirala *et al.*, 2024). A similar reaction with 100 µM 3,4-DHBA as substrate was also performed to check the promiscuity of LaN4'OMT enzyme in the presence of Mg<sup>2+</sup> ions. Enzymatic assay for cytochrome enzyme was carried out using the microsomal fractions prepared from *N. benthamiana* leaves. The reaction was carried out in a total volume of 150 µL with 50 mM phosphate buffer (pH 7.8), using approximately 100 µg of microsomes, 1 mM NADPH and 50 µM 4'OM, incubated at 30 °C for a time course followed by quenching with an equal volume of 100% acetonitrile. To obtain the reduced product, the same reaction condition was carried out in a total volume of 150 µL in addition to 75 µg of pure reductase enzyme (LaNR1I) obtained from *E. coli*. Papaverine was added after quenching the *in vitro* reactions in such a way to obtain 1 ppm of final concentration while mixing with the mobile phase before injection in LC–MS/MS.

### Thermal unfolding analysis using nanoDSF

For melting temperature measurements, 500 µg/mL of purified MBP-tagged LaNBS was mixed with 100 µM of tyramine and 100 µM of 3,4-DHBA in 50 mM HEPES (pH 6) buffer with 20% glycerol. For LaNR1I, 1 mM of NADPH was used in addition to the substrates, and the reaction was carried out in sodium citrate buffer. Similarly, enzymes without substrates, as well as purified MBP alone, and with substrates were used as a negative control for the experiment. The samples were incubated at room temperature for 5 min before being introduced into high-sensitivity capillaries, each containing 10 µL of sample. Excitation light settings were configured to autodetect, and measurements were conducted across a temperature range of 25–95 °C, with a controlled temperature increase of 1 °C/min, using the Prometheus Panta instrument (NanoTemper Technologies GmbH, Munich, Germany). The protein melting temperature (*T*<sub>m</sub>) was determined by plotting the fluorescence intensity ratio at 350 nm and 330 nm (F350/F330), which reflects changes in the local chemical environment of tryptophan and tyrosine residues. The thermal unfolding profile was analysed using Panta analysis software (NanoTemper Technologies GmbH, Munich, Germany),

and the derivatives of the experimental value were plotted to enhance data visualization by Prism.

### Sample preparation for metabolite analysis

Metabolites were extracted from crushed snap-frozen leaf tissues using methanol–water extraction solvent (8:2, v/v) with a constant ratio of 20 µL/mg of the samples. Samples were placed on a shaker at 200 rpm overnight and then heated at 65 °C for 20 min with a closed lid. Supernatants were collected after centrifuging at 16 000 *g* for 10 min and dried using a rotary evaporator. Sample extracts were reconstituted in the mobile phase (Milli-Q water and methanol, both containing formic acid 0.1% (90:10)) at a constant ratio of 20 µL/mg unless stated, vortexed and filtered through 0.22 µm PTFE filter prior analysis conducted by High-performance liquid chromatography coupled with tandem mass spectrometry (HPLC–MS/MS). Papaverine was added as an internal standard prior to extraction, ensuring a final concentration of 1 ppm upon reconstitution in the mobile phase. Relative quantification was subsequently performed by normalizing metabolite signal intensities to the area under the curve of the internal standard.

### Instrumentation and chromatographic conditions

HPLC–MS/MS system equipped with an Agilent Jet Stream ionization source, a binary pump, an autosampler and a column compartment was used for the analysis (Agilent Technologies, Santa Clara, California, USA). Compound separation was achieved using a Kinetex EVO C18 column (150 × 4.6 mm, 5 mm, 100 Å; Phenomenex, Torrance, USA). Five µL of each sample were injected into the column that was set at 30 °C. A gradient method made of (A) formic acid 0.1% v/v in Milli-Q water and (B) formic acid 0.1% v/v in methanol with a flow rate of 0.4 mL/min was used to achieve chromatographic separation. The HPLC elution program was the following: 0 min, 10% B; 10.0 min, 10% B; 20.0 min, 100% B; 25.0 min, 100% B; 26.0 min, 10% B. The total run time was 30 min per sample to allow the reconditioning of the column prior to the next injection. The parameters used in the MS/MS source were set as follows: gas flow rate 10 L.min<sup>–1</sup>, gas temperature 300 °C, nebulizer 45 psi, sheath gas flow 11 L.min<sup>–1</sup>, sheath gas temperature 300 °C, capillary voltage 4000 V in ESI+ and 3500 V in ESI– and nozzle voltage 500 V. Agilent MassHunter Data Acquisition (version 1.2) software was used to control the HPLC–MS/MS. MassHunter Qualitative Analysis (version 10.0) and MassHunter Quantitative QQQ Analysis (version 10.0) softwares were used for data processing. Compound identification was performed using authentic reference standards or by employing collision-induced dissociation (CID) mass spectra published in the literature (Kilgore *et al.*, 2016a,b; Mehta *et al.*, 2024), in cases where reference standards were not available. MRM transitions and instrument parameters used for identifications in N4'OMT and CYP96T experiments are included in Tables S3 and S4, respectively. Instrument parameters used in Product Ion Analysis mode to obtain the full CID mass spectra of signals detected in MRM mode are presented in Table S5.

### Subcellular localization study

Subcellular localization of LaN4'OMT and LaCYP96T2 was determined using confocal microscopy by infiltrating culture of *A. tumefaciens* (GV3101) harbouring pB7FWG2–LaN4'OMT, pB7WGF2–LaN4'OMT, or pB7FWG2–LaCYP96T2 into 3–4 weeks old *N. benthamiana* plants. *Agrobacterium* suspensions were prepared and infiltrated as described above. For co-localization

assay, co-infiltration of the suspensions aforementioned, along with *Agrobacterium* harbouring RFP or ER-mCherry plasmids (Nelson *et al.*, 2007) were mixed to a final OD<sub>600</sub> of 0.6 and infiltrated. After 48 h of infiltration, the abaxial epidermis of leaf discs was placed on a microscopic slide in a water drop, shielded by a cover slip and photographed immediately. A confocal laser scanning microscope (Leica TCS SP8; Leica Microsystems) with a 40X/1.30 oil immersion objective was used to observe protein localization. Las X software was used to create the composite photos (Leica Microsystems). Colocalization graph was created using fluorescence intensity.

## Funding

This research was financially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) award number RGPIN-2021-03218 to I.D-P. This work was also supported by the Canada Research Chair on plant-specialized metabolism Award No CRC-2018-00137 to I.D-P. Thanks are extended to the Canadian taxpayers and the Canadian government for supporting the Canada Research Chairs Program.

## Acknowledgements

We are grateful to Dr Bridget Milorey and Cyril Castel of NanoTember Technologies, Inc. for assistance with differential scanning fluorescence assay. We thank Patrick Lagüe from Université Laval for sharing his expertise on molecular docking. We thank Dr Bharat Bhusan Majhi for providing *LaNBS*, and *LaNRll* gene, Melodie B. Plourde for helping with confocal imaging, and Dr Fatma Meddeb for timely help in obtaining the lab materials along with helpful discussions.

## Author contributions

Basanta Lamichhane: Conceptualization, methodology, investigation, formal analysis, validation, visualization, writing—original draft and reviewing and editing; Sarah-Eve Gélinas: Methodology, resources, and editing, and natasha merindol: Conceptualization, methodology, formal analysis, project administration, resources, writing – reviewing and editing; Manoj Koirala: Methodology, resources; karen cristine gonzalves dos santos: methodology, resources; Hugo Germain: Conceptualization, methodology, resources, co-supervision, reviewing and editing; Isabel Desgagné-Penix: Conceptualization, methodology, resources, funding acquisition, supervision, writing – reviewing and editing.

## Conflict of interest

The authors declare that there is no situation real, potential, or apparent conflict of interest.

## Data availability statement

The data that supports the findings of this study are available in the supplementary material of this article.

## References

Adasme, M.F., Linnemann, K.L., Bolz, S.N., Kaiser, F., Salentin, S., Haupt, V.J. and Schroeder, M. (2021) PLIP 2021: expanding the scope of the protein-ligand interaction profiler to DNA and RNA. *Nucleic Acids Res.* **49**, W530–W534.

- Bassard, J.E., Richert, L., Geerinck, J., Renault, H., Duval, F., Ullmann, P., Schmitt, M. *et al.* (2012) Protein-protein and protein-membrane associations in the lignin pathway. *Plant Cell*, **24**, 4465–4482.
- Belin, P., Le Du, M.H., Fielding, A., Lequin, O., Jacquet, M., Charbonnier, J.B., Lecoq, A. *et al.* (2009) Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA*, **106**, 7426–7431.
- Benach, J., Atrian, S.I., González-Duarte, R. and Ladenstein, R. (1999) The catalytic reaction and inhibition mechanism of *Drosophila* alcohol dehydrogenase: observation of an enzyme-bound NAD-ketone adduct at 1.4 Å resolution by X-ray crystallography. *J. Mol. Biol.* **289**, 335–355.
- Berkov, S., Pavlov, A., Georgiev, V., Bastida, J., Burrus, M., Ilieva, M. and Codina, C. (2009) Alkaloid synthesis and accumulation in *Leucosium aestivum* in vitro cultures. *Nat. Prod. Commun.* **4**, 1934578X0900400328.
- Boccia, M., Grzech, D., Lopes, A.A., O'Connor, S.E. and Caputi, L. (2022) Directed biosynthesis of new to nature alkaloids in a heterologous *Nicotiana benthamiana* expression host. *Front. Plant Sci.* **13**, 919443.
- Bonifacio, M.J., Archer, M., Rodrigues, M.L., Matias, P.M., Learmonth, D.A., Carrondo, M.A. and Soares-Da-Silva, P. (2002) Kinetics and crystal structure of catechol-o-methyltransferase complex with co-substrate and a novel inhibitor with potential therapeutic application. *Mol. Pharmacol.* **62**, 795–805.
- Brock, A., Brandt, W. and Drager, B. (2008) The functional divergence of short-chain dehydrogenases involved in tropinone reduction. *Plant J.* **54**, 388–401.
- Bryant, D.M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M.B., Payzin-Dogru, D., Lee, T.J. *et al.* (2017) A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* **18**(3), 762–776.
- Burgstaller-Muehlbacher, S., Crotty, S.M., Schmidt, H.A., Reden, F., Drucks, T. and von Haeseler, A. (2023) ModelRevelator: fast phylogenetic model estimation via deep learning. *Mol. Phylogenet. Evol.* **188**, 107905.
- Cahlíkova, L., Breiterova, K. and Opletal, L. (2020) Chemistry and biological activity of alkaloids from the genus *Lycoris* (Amaryllidaceae). *Molecules*, **25**, 4797.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
- Caputi, L., Franke, J., Farrow, S.C., Chung, K., Payne, R.M.E., Nguyen, T.D., Dang, T.T. *et al.* (2018) Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science*, **360**, 1235–1239.
- Chen, Z., Jiang, J.C., Lin, Z.G., Lee, W.R., Baker, M.E. and Chang, S.H. (1993) Site-specific mutagenesis of *Drosophila* alcohol dehydrogenase: evidence for involvement of tyrosine-152 and lysine-156 in catalysis. *Biochemistry*, **32**, 3342–3346.
- Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**, i890.
- Cui, W., Iwasa, K., Sugiura, M., Takeuchi, A., Tode, C., Nishiyama, Y., Moriyasu, M. *et al.* (2007) Biotransformation of phenolic 1-benzyl-N-methyltetrahydroisoquinolines in plant cell cultures followed by LC/NMR, LC/MS, and LC/CD. *J. Nat. Prod.* **70**, 1771–1778.
- Desgagné-Penix, I. (2021) Biosynthesis of alkaloids in Amaryllidaceae plants: a review. *Phytochem. Rev.* **20**, 409–431.
- Dong, L., Miettinen, K., Goedbloed, M., Verstappen, F.W., Voster, A., Jongsma, M.A., Memelink, J. *et al.* (2013) Characterization of two geraniol synthases from *Valeriana officinalis* and *Lippia dulcis*: similar activity but difference in subcellular localization. *Metab. Eng.* **20**, 198–211.
- Dudley, Q.M., Jo, S., Guerrero, D.A.S., Chhetry, M., Smedley, M.A., Harwood, W.A., Sherden, N.H. *et al.* (2022) Reconstitution of monoterpene indole alkaloid biosynthesis in genome engineered *Nicotiana benthamiana*. *Commun. Biol.* **5**, 949.
- Dumas, V.G., Defelipe, L.A., Petruk, A.A., Turjanski, A.G. and Marti, M.A. (2014) QM/MM study of the C-C coupling reaction mechanism of CYP121, an essential cytochrome p450 of *Mycobacterium tuberculosis*. *Proteins*, **82**, 1004–1021.
- Eddy, S.R. (2008) A probabilistic model of local sequence alignment that simplifies statistical significance estimation. *PLoS Comput. Biol.* **4**, e1000069.
- Eddy, S.R. (2011) Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195.

- El Tahchy, A., Boisbrun, M., Ptak, A., Dupire, F., Chrétien, F., Henry, M., Chapleur, Y. *et al.* (2010) New method for the study of Amaryllidaceae alkaloid biosynthesis using biotransformation of deuterium-labeled precursor in tissue cultures. *Acta Biochim. Pol.* **57**, 75–82.
- Filling, C., Berndt, K.D., Benach, J., Knapp, S., Prozorovski, T., Nordling, E., Ladenstein, R. *et al.* (2002) Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *J. Biol. Chem.* **277**, 25677–25684.
- Gani, O.A., Adekoya, O.A., Giurato, L., Spyraakis, F., Cozzini, P., Guccione, S., Winberg, J.O. *et al.* (2008) Theoretical calculations of the catalytic triad in short-chain alcohol dehydrogenases/reductases. *Biophys. J.* **94**, 1412–1427.
- Georgievaa, L., Berkov, S., Kondakova, V., Bastida, J., Viladomat, F., Atanassov, A. and Codina, C. (2007) Alkaloid variability in *Leucojum aestivum* from wild populations. *Z. Naturforsch. C*, **62**, 627–635.
- Girard, M.P., Karimzadegan, V., Heneault, M., Cloutier, F., Berube, G., Berthou, L., Merindol, N. *et al.* (2022) Chemical synthesis and biological activities of Amaryllidaceae alkaloid norbelladine derivatives and precursors. *Molecules*, **27**, 5621.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X. *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 652.
- von Haeseler, A., Schmidt, H.A., Bui, M.Q. and Nguyen, L.T. (2014) *IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies*.
- Hansen, C.C., Nelson, D.R., Møller, B.L. and Werck-Reichhart, D. (2021) Plant cytochrome P450 plasticity and evolution. *Mol. Plant*, **14**, 1244–1265.
- He, Y., Xu, M. and Chen, X. (2020) De Novo Transcriptomics Analysis of the Floral Scent of Chinese Narcissus. *Trop. Plant Biol.* **13**(2), 172–188.
- van Herpen, T.W., Cankar, K., Nogueira, M., Bosch, D., Bouwmeester, H.J. and Beekwilder, J. (2010) *Nicotiana benthamiana* as a production platform for artemisinin precursors. *PLoS One*, **5**, e14222.
- Hnin, S.Y.Y., Nakashima, Y., Kodama, T. and Morita, H. (2024a) Structure-based catalytic mechanism of Amaryllidaceae O-methyltransferases. *ACS Catal.* **14**, 11865–11880.
- Hnin, S.Y.Y., Nakashima, Y. and Morita, H. (2024b) Substrate flexibilities of norbelladine synthase and noroxomaritidine/norcrugsodine reductase for hydroxylated and/or methoxylated aldehydes. *Chem. Pharm. Bull.* **72**, 507–511.
- Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q. and Vinh, L.S. (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522.
- Hotchandani, T., de Villiers, J. and Desgagné-Penix, I. (2019) Developmental regulation of the expression of amaryllidaceae alkaloid biosynthetic genes in *Narcissus papyraceus*. *Genes* **10**(8), 594.
- Ilari, A., Franceschini, S., Bonamore, A., Arengi, F., Botta, B., Maccone, A., Pasquo, A. *et al.* (2009) Structural basis of enzymatic (S)-norocclaurine biosynthesis. *J. Biol. Chem.* **284**, 897–904.
- Irwin, J.J., Tang, K.G., Young, J., Dandarchuluun, C., Wong, B.R., Khurelbaatar, M., Moroz, Y.S. *et al.* (2020) ZINC20-A free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* **60**, 6065–6073.
- Iwasa, K. and Kim, C.W. (1997) Biotransformations of protoberberines in cell cultures of *Dicentra spectabilis*. *Phytochemistry*, **46**, 1359–1363.
- Jayawardena, T.U., Merindol, N., Liyanage, N.S. and Desgagné-Penix, I. (2024) Unveiling Amaryllidaceae alkaloids: from biosynthesis to antiviral potential – a review. *Nat. Prod. Rep.* **41**, 721–747.
- Jornvall, H., Persson, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffery, J. and Ghosh, D. (1995) Short-chain dehydrogenases/reductases (SDR). *Biochemistry*, **34**, 6003–6013.
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., Von Haeseler, A. and Jermin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods*, **14**, 587–589.
- Karimi, M., Depicker, A. and Hilson, P. (2007) Recombinational cloning with plant gateway vectors. *Plant Physiol.* **145**, 1144–1154.
- Kavanagh, K., Jörnvall, H., Persson, B. and Oppermann, U. (2008) Medium- and short-chain dehydrogenase/reductase gene and protein families: the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* **65**, 3895–3906.
- Kilgore, M.B. and Kutchan, T.M. (2016) The Amaryllidaceae alkaloids: biosynthesis and methods for enzyme discovery. *Phytochem. Rev.* **15**, 317–337.
- Kilgore, M.B., Augustin, M.M., Starks, C.M., O’Neil-Johnson, M., May, G.D., Crow, J.A. and Kutchan, T.M. (2014) Cloning and characterization of a norbelladine 4’-O-methyltransferase involved in the biosynthesis of the Alzheimer’s drug galanthamine in *Narcissus sp. aff. pseudonarcissus*. *PLoS One*, **9**, e103223.
- Kilgore, M.B., Holland, C.K., Jez, J.M. and Kutchan, T.M. (2016b) Identification of a noroxomaritidine reductase with Amaryllidaceae alkaloid biosynthesis related activities. *J. Biol. Chem.* **291**, 16740–16752.
- Kilgore, M.B., Augustin, M.M., May, G.D., Crow, J.A. and Kutchan, T.M. (2016a) CYP96T1 of *Narcissus sp. aff. pseudonarcissus* catalyzes formation of the Para-Para’CC phenol couple in the Amaryllidaceae alkaloids. *Front. Plant Sci.* **7**, 225.
- Kohut, E., Ördögh, M., Jámor-Benczúr, E. and Máthé, Á. (2007) Results with the establishment of in vitro culture of *Leucojum aestivum*. *Int. J. Hort. Sci.* **13**, 67–71.
- Koirala, M., Goncalves, C., dos Santos, K., Gélinas, S.-E., Ricard, S., Karimzadegan, V., Lamichane, B. *et al.* (2023) Auxin and light-mediated regulation of growth, morphogenesis, and alkaloid biosynthesis in *Crinum x powellii* ‘Album’ callus. *Phytochemistry* **216**, 113883.
- Koirala, M., Karimzadegan, V., Liyanage, N.S., Méridol, N. and Desgagné-Penix, I. (2022) Biotechnological approaches to optimize the production of Amaryllidaceae alkaloids. *Biomolecules*, **12**, 893.
- Koirala, M., Merindol, N., Karimzadegan, V., Gelinas, S.E., Liyanage, N.S., Lamichane, B., Tobon, M.C.G. *et al.* (2024) Kinetic and in silico structural characterization of norbelladine O-methyltransferase of Amaryllidaceae alkaloids biosynthesis. *J. Biol. Chem.* **300**, 107649.
- Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 580.
- Letunic, I. and Bork, P. (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296.
- Li, W., Qiao, C., Pang, J., Zhang, G. and Luo, Y. (2019) The versatile O-methyltransferase LrOMT catalyzes multiple O-methylation reactions in Amaryllidaceae alkaloids biosynthesis. *Int. J. Biol. Macromol.* **141**, 680–692.
- Li, Y., Bai, Y., Fan, T.P., Zheng, X. and Cai, Y. (2022) Characterization of a putative tropinone reductase from *Tarenaya hassleriana* with a broad substrate specificity. *Biotechnol. Appl. Biochem.* **69**, 2530–2539.
- Li, Q., Xu, J., Yang, L., Zhou, X., Cai, Y. and Zhang, Y. (2020) Transcriptome analysis of different tissues reveals key genes associated with galanthamine biosynthesis in *Lycoris longituba*. *Front. Plant Sci.* **11**, 519752.
- Liscombe, D.K., Louie, G.V. and Noel, J.P. (2012) Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Nat. Prod. Rep.* **29**, 1238–1250.
- Liu, J., Han, L., Li, G., Zhang, A., Liu, X. and Zhao, M. (2023) Transcriptome and metabolome profiling of the medicinal plant *Veratrum mengtzeanum* reveal key components of the alkaloid biosynthesis. *Front. Genet.* **14**, 1023433.
- Liu, Z., Sun, B., Li, J., Xiang, Y., Wang, R., Jiang, X., Zhu, X. *et al.* (2024) Functional characterization of CYP96T1-like cytochrome P450 from *Lycoris aurea* catalyzing para-para’ and para-ortho’ oxidative coupling in Amaryllidaceae alkaloids biosynthesis. *Front. Plant Sci.* **15**, 1438102.
- Liyanage, N.S., Awwad, F., Gonçalves dos Santos, K.C., Jayawardena, T.U., Méridol, N. and Desgagné-Penix, I. (2024) Navigating Amaryllidaceae alkaloids: bridging gaps and charting biosynthetic territories – a comprehensive review. *J. Exp. Bot.* **76**, 16–34.
- Majhi, B.B., Gélinas, S.-E., Méridol, N., Ricard, S. and Desgagné-Penix, I. (2023) Characterization of norbelladine synthase and noroxomaritidine/norcrugsodine reductase reveals a novel catalytic route for the biosynthesis of Amaryllidaceae alkaloids including the Alzheimer’s drug galanthamine. *Front. Plant Sci.* **14**, 1231809.
- Mehta, N., Meng, Y., Zare, R., Kamenetsky-Goldstein, R. and Sattely, E. (2024) A developmental gradient reveals biosynthetic pathways to eukaryotic toxins in monocot geophytes. *Cell*, **187**, e5610.
- Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., Woittiez, L. *et al.* (2014) The seco-iridoid pathway from *Catharanthus roseus*. *Nat. Commun.* **5**, 3606.



- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. and Steinegger, M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682.
- Navrocki, E.P., Kolbe, D.L. and Eddy, S.R. (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics*, **25**, 1337.
- Nelson, B.K., Cai, X. and Nebenführ, A. (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. *Plant J.* **51**, 1126–1136.
- Nett, R.S., Lau, W. and Sattely, E.S. (2020) Discovery and engineering of colchicine alkaloid biosynthesis. *Nature*, **584**, 148–153.
- Okonechnikov, K., Golosova, O., Fursov, M. and UGENE Team (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, **28**, 1166–1167.
- One Thousand Plant Transcriptomes Initiative (2019) One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**(7780), 679–685.
- Pavlov, A., Berkov, S., Courtois, E., Gocheva, T., Tuneva, D., Pandova, B., Georgiev, M. et al. (2007) Galanthamine production by *Leucojum aestivum* in vitro systems. *Process Biochem.* **42**, 734–739.
- Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L. et al. (2023) InterPro in 2022. *Nucleic Acids Res.* **51**, D418–D427.
- Peng, W., Li, Z., Wang, S. and Wang, B. (2023) Unravelling the C-C and C-N coupling mechanism for the CYP96T1-catalyzed biosynthesis of Amaryllidaceae alkaloids. *Mol. Catal.* **550**, 113609.
- Porter, T.D. and Coon, M.J. (1991) Cytochrome P-450. Multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. *J. Biol. Chem.* **266**, 13469–13472.
- Qiao, F., He, Y., Zhang, Y., Jiang, X., Cong, H., Wang, Z., Sun, H. et al. (2023) Elucidation of the 1-phenethylisoquinoline pathway from an endemic conifer *Cephalotaxus hainanensis*. *Proc. Natl. Acad. Sci. USA*, **120**, e2209339120.
- Reed, J. and Osbourn, A. (2018) Engineering terpenoid production through transient expression in *Nicotiana benthamiana*. *Plant Cell Rep.* **37**, 1431–1441.
- Reinhardt, N., Fischer, J., Coppi, R., Blum, E., Brandt, W. and Dräger, B. (2014) Substrate flexibility and reaction specificity of tropinone reductase-like short-chain dehydrogenases. *Bioorg. Chem.* **53**, 37–49.
- Roth, S., Kilgore, M.B., Kutchan, T.M. and Muller, M. (2018) Exploiting the catalytic diversity of short-chain dehydrogenases/reductases: versatile enzymes from plants with extended imine substrate scope. *Chembiochem*, **19**, 1849–1852.
- Samanani, N. and Facchini, P.J. (2002) Purification and characterization of norcoclaurine synthase: the first committed enzyme in benzyloisoquinoline alkaloid biosynthesis in plants. *J. Biol. Chem.* **277**, 33878–33883.
- Samanani, N., Liscombe, D.K. and Facchini, P.J. (2004) Molecular cloning and characterization of norcoclaurine synthase, an enzyme catalyzing the first committed step in benzyloisoquinoline alkaloid biosynthesis. *Plant J.* **40**, 302–313.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R. et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539.
- Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3212.
- Singh, A. and Desgagné-Penix, I. (2017) Transcriptome and metabolome profiling of *Narcissus pseudonarcissus* 'King Alfred' reveal components of Amaryllidaceae alkaloid metabolism. *Sci. Rep.* **7**(1). <https://doi.org/10.1038/s41598-017-17724-0>
- Singh, A., Massicotte, M.-A., Garand, A., Tousignant, L., Ouellette, V., Bérubé, G. and Desgagné-Penix, I. (2018) Cloning and characterization of norbelladine synthase catalyzing the first committed reaction in Amaryllidaceae alkaloid biosynthesis. *BMC Plant Biol.* **18**, 1–12.
- Stanfield, J.K., Omura, K., Matsumoto, A., Kasai, C., Sugimoto, H., Shiro, Y., Watanabe, Y. et al. (2020) Crystals in minutes: instant on-site microcrystallisation of various flavours of the CYP102A1 (P450BM3) haem domain. *Angew. Chem. Int. Ed.* **59**, 7611–7618.
- Su, Y., Li, H.-P., Zhang, M., Ding, X.-W., Xu, J.-H., Chen, Q. and Zheng, G.-W. (2022) Regioselectivity inversion of an O-methyltransferase via semi-rational mutagenesis combined with metal ion substitution. *ChemCatChem*, **14**, e202200844.
- Sun, B., Wang, P., Wang, R., Li, Y. and Xu, S. (2018) Molecular cloning and characterization of a meta/para-O-methyltransferase from *Lycoris aurea*. *Int. J. Mol. Sci.* **19**, 1911.
- Takos, A.M. and Rook, F. (2013) Towards a molecular understanding of the biosynthesis of Amaryllidaceae alkaloids in support of their expanding medical use. *Int. J. Mol. Sci.* **14**, 11713–11741.
- Tousignant, L., Diaz-Garza, A.M., Majhi, B.B., Gélinas, S.-E., Singh, A. and Desgagné-Penix, I. (2022) Transcriptome analysis of *Leucojum aestivum* and identification of genes involved in norbelladine biosynthesis. *Planta*, **255**, 30.
- Wang, Q.-M., Cui, J., Dai, H., Zhou, Y., Li, N. and Zhang, Z. (2018) Comparative transcriptome profiling of genes and pathways involved in leaf-patterning of *Clivia miniata* var. *variegata*. *Gene* **677**, 280–288.
- Wang, N., Shu, X., Zhang, F., Zhuang, W., Wang, T. and Wang, Z. (2021) Comparative transcriptome analysis identifies key regulatory genes involved in anthocyanin metabolism during flower development in *Lycoris radiata*. *Front. Plant Sci.* **12**, 761862.
- Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
- Xu, J.J., Fang, X., Li, C.Y., Yang, L. and Chen, X.Y. (2020) General and specialized tyrosine metabolism pathways in plants. *ABIOTECH*, **1**, 97–105.
- Yang, F., Li, C., Das, D., Zheng, Y., Song, T., Wang, L., Chen, M.-X. et al. (2021) Comprehensive transcriptome and metabolic profiling of petal color development in *Lycoris sprengeri*. *Front. Plant Sci.* **12**, 747131.
- Zeng, J., Deng, Y., Iqbal, S., Zhang, J., Wu, K., Ma, G., Li, L. et al. (2024) Direct somatic embryogenesis and related gene expression networks in leaf explants of *Hippeastrum* 'Bangkok Rose'. *Hortic. Plant J.* **10**(2), 556–572.
- Zhang, F., Cheng, G., Shu, X., Wang, N. and Wang, Z. (2022a) Transcriptome analysis of *Lycoris chinensis* bulbs reveals flowering in the age-mediated pathway. *Biomolecules* **12**(7), 899.
- Zhang, C., Shao, X., Wang, X., Shou, L., Liu, Y., Xu, Z. and Guo, Z. (2022b) Development of a general bioluminescent activity assay for peptide ligases. *The FEBS J.* **289**(17), 5241–5258.

## Supporting information

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

**Figure S1** Phylogenetic analysis, multiple sequence alignment and folding of LaNBS.

**Figure S2** Phylogenetic analysis, multiple sequence alignment and folding of LaNR1I.

**Figure S3** Phylogenetic analysis, multiple sequence alignment and folding of LaN4'OMT.

**Figure S4** Phylogenetic analysis, multiple sequence alignment and folding of LaCYP96T1 and LaCYP96T2.

**Figure S5** Cellular localization of LaN4'OMT and LaCYP96T2.

**Figure S6** Western blot, SDS gel and RT-qPCR analysis illustrating the expression of fusion proteins of LaCYP96T1 and LaN4'OMT with EGFP along with LaNBS, LaNR1I, LaN4'OMT, LaCYP96T1, LaCYP96T2 in *N. benthamiana* leaf lysate.

**Figure S7** HPLC-MS/MS chromatograms of norbelladine and 4'-O-methylnorbelladine.

**Figure S8** Docked structures of candidates LaCYP96T and LaNR1I.

**Figure S9** HPLC-MS/MS analysis of N-demethylnorbelladine, produced upon transient expression of LaCYP96T1 with 4'-O-methylnorbelladine as substrate and *in vitro*.

**Figure S10** HPLC-MS/MS analysis of noroxomaritinidine, produced upon transient expression of LaCYP96T1 with 4'-O-methylnorbelladine and *in vitro*.

**Figure S11** HPLC-MS/MS analysis of oxomaritinamine, produced upon transient expression of LaCYP96T1 with 4'-O-methylnorbelladine as substrates in *N. benthamiana*.

**Figure S12** HPLC-MS/MS analysis of norpluviine produced upon transient co-expression of *LaCYP96T1* and *LaNRll* with 4'-O-methylnorbelladine as substrates in *N. benthamiana*, and *in vitro*.

**Figure S13** HPLC-MS/MS chromatograms of the *ortho-para'* product lycorine, that was produced upon transient co-expression of *LaCYP96T1* and *LaNRll* with 4'-O-methylnorbelladine in *N. benthamiana*.

**Figure S14** HPLC-MS/MS chromatograms of the *para-para'* product normaritidine, that was produced upon transient co-expression of *LaCYP96T1* and *LaNRll* with 4'-O-methylnorbelladine in *N. benthamiana*.

**Figure S15** The elucidated alkaloid biosynthesis pathway in *Leucojum aestivum* in this study.

**Table S1** The accession numbers of the sequences used for phylogenetic analysis for respective protein sequence used for phylogenetic analysis in Figures S1–S4.

**Table S2** List of primers used in the experiments.

**Table S3** MRM transitions and instrumental parameters used on HPLC-MS/MS for *LaNBS*, *LaNR* and *LaOMT* experiments.

**Table S4** MRM transitions and instrumental parameters used on HPLC-MS/MS for CYP experiments.

**Table S5** Instrumental parameters used in Product Ion mode on HPLC-MS/MS for CYP experiments.

**Table S6** Multiple Sequence Alignment Distance Matrix (Similarity in percent).

**Table S7** Active Site description and docking results of *LaNBS* (tyramine and 3,4-DHBA), *LaNRll* (norcraugsodine), *LaN4'OMT* (norbelladine) and *LaCYP96T1* and T2 (4'-O-methylnorbelladine).

**Table S8** Differences in residues between SDRs.

**Table S9** Differences in residues between O-methyltransferase enzymes of different substrate specificity.

**Table S10** Differences in residues between CYP enzymes with different product specificity/phenol coupling.

**Table S11** Summary of *in vivo* enzymatic assay.