Ancient gene clusters initiate monoterpenoid indole alkaloid biosynthesis and C3 stereochemistry inversion

Jaewook Hwang[1§], Jonathan Kirshner[2§], Daniel André Ramey Deschênes[1], Matthew Bailey Richardson[1], Steven J. Fleck[2], Jun Guo[1], Jacob Owen Perley[1], Mohammadamin Shahsavarani[1], Jorge Jonathan Oswaldo Garza-Garcia[1], Alyssa Dawn Seveck[1], Savannah Sadie Doiron[1], Zhan Mai[1], Stephen Nelson Silliphant[1], Larry Calhoun[1], Di Gao[3], Jiazhang Lian[3,4,5], Ghislain Deslongchamps[1*], Victor A. Albert[2*], and Yang Qu[1*]

[1] Department of Chemistry, University of New Brunswick, Fredericton, E3B 5A3, Canada
[2] Department of Biological Sciences, University at Buffalo, Buffalo NY, 14260, USA
[3] Key Laboratory of Biomass Chemical Engineering of Ministry of Education, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou, China
[4] ZJU-Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou, China
[5] Zhejiang Key Laboratory of Smart Biomaterials, Zhejiang University, Hangzhou, China

[§]These authors contributed equally

[*]Corresponding authors:

Ghislain Deslongchamps ghislain@unb.ca
Victor A. Albert vaalbert@buffalo.edu
Yang Qu yang.qu@unb.ca

Author contribution:

JH, JK, GD, VAA, and YQ conceived the research and oversaw overall direction and planning. JH performed all in vitro and in vivo experiments. JH and JK annotated the *Rauvolfia tetraphylla* genome. JK implemented phylogenetic analyses. JK and VAA performed genome structural evolutionary analyses. DARD and MBR conducted homology modeling and docking experiments. SJF performed genomic analyses using Ksrates software. JG, JOP, and MS performed cloning and yeast strain construction. JJOGG, DG, and JL conducted de novo yeast strain construction and analysis. JH, AS, SSD, ZM, SNS, and LC conducted compound purification and NMR analyses. JH, VAA, SJF and JK performed additional plant genome analyses. JH, JK, VAA and YQ wrote the paper with the input from all authors.

Abstract
The inversion of C3 stereochemistry in monoterpenoid indole alkaloids (MIAs), derived from the central precursor strictosidine (3*S*), is essential for synthesizing numerous 3*R* MIAs and oxindoles, including the antihypertensive drug reserpine found in *Rauvolfia serpentina* (Indian snakeroot) and *Rauvolfia tetraphylla* (devil pepper) of the plant family Apocynaceae. MIA biosynthesis begins with the reduction of strictosidine aglycone by various reductases, preserving the initial 3*S* stereochemistry. In this study, we identify and biochemically characterize a conserved oxidase-reductase pair from the Apocynaceae, Rubiaceae, and Gelsemiaceae families of the order Gentianales: the heteroyohimbine/yohimbine/corynanthe C3-oxidase (HYC3O) and C3-reductase (HYC3R). These enzymes collaboratively invert the 3*S* stereochemistry to 3*R* across a range of substrates, resolving the long-standing question about the origin of 3*R* MIAs and oxindole derivatives, and facilitation of reserpine biosynthesis. Notably, *HYC3O* and *HYC3R* are located within gene clusters in both the *R. tetraphylla* and *Catharanthus roseus* (Madagascar periwinkle) genomes, which are partially homologous to an elusive geissoschizine synthase (GS) gene cluster we also identified in these species. In *R. tetraphylla*, these clusters occur closely in tandem on a single chromosome, likely stemming from a single segmental duplication event, while in *C. roseus*, a closely related member of rauvolfioid Apocynaceae, they were later separated by a chromosomal translocation. The ancestral genomic context for both clusters can be traced all the way back to common ancestry with grapevine. Given the presence of syntenic GS homologs in *Mitragyna speciosa* (Rubiaceae), the GS cluster, at least in part, probably evolved at the base of the Gentianales, which split from other core eudicots up to 135 million years ago. We also show that the strictosidine biosynthetic gene cluster, required to initiate the MIA pathway, plausibly evolved concurrently. The reserpine biosynthetic cluster likely arose much later in the rauvolfioid lineage of Apocynaceae. Collectively, our work uncovers the genomic and biochemical basis for key events in MIA evolution and diversification, providing insights beyond the well-characterized vinblastine and ajmaline biosynthetic pathways.

Introduction

Biosynthetic gene clusters (BGCs) are commonly found in bacteria and fungi, where they efficiently regulate the expression of genes involved in defense and communication. With the advent of plant whole-genome sequencing, more plant BGCs have been identified, particularly those associated with specialized metabolism. These clusters highlight the evolutionary drive to coordinate gene expression to produce beneficial metabolites. Large BGCs have been discovered in plants for the biosynthesis of compounds such as triterpenoids and benzylisoquinoline alkaloids, including the 9-gene cluster for saponin QS-21 and the 17-gene cluster for noscapine and morphine biosynthesis [1–3]. Additionally, smaller clusters with fewer biosynthetic genes are scattered across the plant kingdom. For example, a three-gene BGC in *Catharanthus roseus* (Madagascar periwinkle) comprises part of the biosynthetic pathway of the anticancer monoterpenoid indole alkaloid (MIA) vinblastine [4], despite a total involvement of over 30 enzymes [5–9].

MIA biosynthesis begins with the STR BGC, wherein strictosidine is synthesized through coordinated activities of closely linked tryptophan decarboxylase (*TDC*), strictosidine synthase (*STR*), and Multidrug and Toxic Compound Extrusion (MATE) transporter (*MATE1*) genes. The first two genes encode enzymes that condense secologanin and tryptamine to strictosidine, linking the secoiridoid and shikimate pathways. MATE1 facilitates the condensation reaction by importing secologanin into the vacuole, where STR resides [10,11]. Strictosidine in turn serves as the sole precursor for over 3,000 MIAs (Fig. 1). The strictosidine BGC is conserved in MIA-producing species within the Gentianales order, such as *Gelsemium sempervirens* in the Carolina jasmine family, Gelsemiaceae, *C. roseus* in the milkweed family, Apocynaceae, and *Mitragyna speciosa* (kratom) and *Ophiorrhiza pumila* in the coffee family, Rubiaceae. This cluster likely emerged alongside the emergence and diversification of MIAs [12,13] temporally close to the split between Gentianales and other core eudicots as early as 135 million years ago (Mya) [14] .

The recently sequenced genome of *Rauvolfia tetraphylla* (devil-pepper) of the Apocynaceae provided several genomic loci encoding enzymes rich in cinnamyl alcohol dehydrogenase (CAD)-like reductase and uncharacterized cytochrome P450 monooxygenase (CYP) activities [15,16]. Studies on the biosynthesis of vinblastine and ajmaline, an antiarrhythmic drug from *Rauvolfia serpentina* (Indian snakeroot), have revealed numerous CAD-like reductases and CYPs for MIA biosynthesis, scattered across their 20-30 biosynthetic steps [5–7,17–22] . Homologous *R. tetraphylla* loci include genes encoding CAD-like reductases, such as yohimban synthase (*RtYOS*), geissoschizine synthase (*RtGS*), tetrahydroalstonine (THA) synthase (*RtTHAS*), and orthologs of *R. serpentina* (Indian snakeroot) vomilenine 1,2-reductase (*RsVR*) and dihydrovomilenine 19,20-reductase (*RsDHVR*) protein products, which we recently discovered and characterized (Fig. 2a) [19]. While *GS* is not physically clustered with *VR* and *DHVR* in the *R. tetraphylla* genome, their gene products work together in the biosynthesis of ajmaline. *YOS*, however, is tightly linked with *VR* and *DHVR*, and its gene product serves as the gateway enzyme for the yohimbine type MIAs, such as the yohimbine and rauwolscine epimers, and potentially for the antihypertensive drug reserpine that is found in both *Rauvolfia* species (Fig. 1) [15,23].

Despite bearing the yohimbine skeleton, reserpine and its analog rescinnamine exhibit a 3*R* stereochemistry, resulting in a perpendicular indole ring orientation relative to the rest of the

molecule (Fig. 1). The 3*S* stereochemistry originates from the central precursor strictosidine. It has been well established that STRs from various species exclusively produce strictosidine (3*S*) rather than its 3*R* epimer, vincoside, with strictosidine serving as the sole precursor to all downstream MIAs [24,25]. Subsequently, strictosidine β-glucosidase (SGD) removes the glucose moiety from strictosidine, forming the highly labile dialdehyde intermediate strictosidine aglycone. All characterized strictosidine aglycone reductases, such as GS, THAS, and YOS from multiple species, exclusively produce 3*S* products [15,18,21,22] (Fig. 1). Besides reserpine, other notable 3*R* MIAs include 3-epi-THA (akuammigine) from *C. roseus* and *Picralima nitida* (akuamma; Apocynaceae) [26,27], reserpilline from *Ochrosia elliptica* (bloodhorn; Apocynaceae) and *Rauvolfia spp*. [28,29], 3-epi-mitragynine (speciociliatine) from *M. speciosa* (kratom; Rubiaceae), and the oxindoles mitraphylline, speciophylline, and corynoxine, which are biosynthesized from 3*R* intermediates in *M. speciosa*, *Mitragyna parvifolia*, *Cephalanthus occidentalis* (button bush; Rubiaceae), and *Hamelia patens* (firebush; Rubiaceae) [22,30–33] (Fig. 1). The widespread distribution of 3*R* series MIAs in nature is intriguing, as it is not well understood how these stereoisomers emerge.

In this study, we investigate genomic structural evolution between *R. tetraphylla* and *C. roseus* in particular, and more broadly among other Gentianales (*Asclepias syriaca*, *Gelsemium elegans, M. speciosa*, and *O. pumila*) and the early-diverging core eudicot grapevine (*Vitis vinifera*). We identify conserved BGCs encoding enzymes involved in reserpine biosynthesis and MIA C3 stereochemistry inversion. Our findings establish that a CAD-rich BGC ancestral to the modern reserpine and GS BGCs evolved in the stem lineage of Gentianales, within a genomic context already present by the order's ancient divergence with *Vitis*, and likely time-coincident with the evolution of the STR BGC itself. Later, a tandem duplication of this ancestral BGC likely occurred to yield two paralogous BGCs that functionally diverged to lead the biosynthesis of yohimbine-type and corynanthe-type MIAs at least by the time rauvolfioid Apocynaceae evolved. Given the broad GS activity within Gentianales already established by the time Rubiaceae evolved (as early as 118 Mya) [14], we infer that GS activity may have been the ancestral BGC function. The GS BGC paralog, the ca. 200-kb reserpine BGC, includes genes encoding a strictosidine aglycone reductase and the catalytic duo heteroyohimbine/yohimbine/corynanthe C3-oxidase (HYC3O) and C3-reductase (HYC3R). The flavoprotein HYC3O abstracts the H3 hydride from 3*S* MIAs produced by various aglycone reductases, yielding an iminium intermediate that is reduced by the CAD-like enzyme HYC3R to form inverted 3*R* products. Analysis of HYC3O/HYC3R enzymes from five additional Gentianales species revealed a conserved biosynthetic mechanism for 3*R* MIA formation. Homology modeling and substrate docking demonstrated the substrate promiscuity of HYC3O and the dual catalytic activity of HYC3R on both strictosidine aglycone and 3-dehydro substrates. Collectively, our work not only resolves a long-standing question on the biosynthesis 3*R* MIAs from the 3*S* strictosidine precursor, but also uncovers the fundamental role that segmental genomic block duplication likely played in the evolution of complex MIA biosynthesis.

Results

The *R. tetraphylla* reserpine gene cluster encodes rauwolscine synthase, a flavoprotein, and a CAD-like reductase

To investigate *Rauvolfia* MIA biosynthesis and its genomic context, we began by carefully mining the published *R. tetraphylla* genome for genes homologous to known, functionally relevant loci [15]. This search revealed a highly repetitive, ca. 200 kbp genomic locus rich in genes encoding CAD-like reductases, including the first committed enzyme for reserpine biosynthesis, yohimbane synthase (RtYOS) (Fig. 2a) [15]. Downstream of *RtYOS* were orthologs of two genes encoding *R. serpentina* reductases for ajmaline biosynthesis, vomilenine 1,2-reductase (RsVR) and dihydrovomilenine 19,20-reductase (RsDHVR)[19], arranged as five tandem paralogs (3 *VR* and 2 *DHVR* paralogs). Further downstream was an unknown E3-ubiquitin ligase encoding gene (*E3UL*) and tandem duplicates of an uncharacterized reductase gene orthologous to a highly expressed reductase termed *R. serpentina RsCAD7* in our previous research [19]. After meticulous inspection of this region, we identified a gene encoding a FAD-linked oxidase, belonging to the berberine-bridge enzyme (BBE) type, just 8.4 kbp upstream of *RtYOS* (Fig. 2a). The flavoprotein and CAD-like reductase gene cluster closely resembled a gene cluster in *C. roseus* that contains a dihydrocorynantheine synthase gene (*CrDCS* or *THAS1*) and a gene encoding a flavoprotein of unknown function (Fig. 2a) [4]. The *R. tetraphylla* and *C. roseus* flavoproteins shared 75% amino acid identity, and both also shared 57% amino acid identity with a known BBE-like oxidase in *C. roseus* MIA biosynthesis, *O*-acetylstemmadenine oxidase (CrASO) (Fig. 2b) [5]. The evidence for conserved BGCs between *C. roseus* and *Rauvolfia* warranted further investigation on the *flavoprotein* and *RsCAD7* orthologs in the compact *R. tetraphylla* genomic region we identified.

First, for experimental validation of enzyme function, we synthesized *R. tetraphylla* flavopotein and expressed it in the yeast *Saccharomyces cerevisiae*. Liquid chromatography tandem mass spectrometry (LC-MS/MS) showed that yeast lysate containing this enzyme could oxidize rauwolscine (*m/z* 355) to a new MIA (*m/z* 353) (Fig. 3a). The *R. serpentina* ortholog (90% amino acid identify) exhibited identical activity. The change in the UV absorption maxima of a typical indole (280 nm) as substrate to that with longer wavelengths (350 nm) as product indicated a double bond formation between C3-*N*4, leading to extended conjugation (Fig. 3b, Supplementary fig. 1a). The formation of a 3-dehydro iminium intermediate by hydride abstraction was also consistent with the flavin chemistry in other BBE-like oxidases, such as CrASO and the California poppy BBE [5,34]. $NaBH_4$ treatment selectively reduced the iminium back to 3*S*-rauwolscine, as the geometry of the tertiary nitrogen strongly favours hydride addition from the *re*-face rather than the *si*-face (Supplementary fig.1b). It then became clear that a reductase might reduce the C3-*N*4 double bond and invert the C3 stereochemistry from *S* to *R*, as required for reserpine biosynthesis. We therefore screened the in vitro activities of *R. serpentina* RsCAD7, RsVR, and RsDHVR, orthologous to the reductases encoded by the *R. tetraphylla* gene cluster (>92% amino acid identity, Fig. 2a), together with five other highly expressed *R. serpentina* CADs from our previous study [19]. Only RsCAD7 could reduce the iminium to 3-epi-rauwolscine (3*R*, *m/z* 355, Fig. 3a), bearing identical MS/MS fragmentation and UV absorption maxima to rauwolscine (Fig. 3b and 3c). Therefore, the ca. 200 kbp *R. tetraphylla* BGC contained the rauwolscine synthase *RtYOS*, an oxidase, and a reductase gene required for the first three steps in reserpine biosynthesis.

3-dehydro-rauwolscine was stable in aqueous solution but degraded repeatedly during purification. However, we could confirm the 3-epi-rauwolscine (3*R*) structure by Nuclear Magnetic Resonance (NMR) (Supplementary Fig. 2-7, Table 1 and 2) [35], after purifying it from a large in vitro reaction. The C3 stereochemistry inversion dramatically altered the orientation of

the indole ring to be perpendicular to the rest of the molecule (Fig. 1). The Nuclear Overhauser Effect Spectroscopy (NOESY) correlations between H3 and H19 were a result of the inverted stereochemistry (Supplementary Fig. 6). Instead, H3 and H15 in rauwolscine (3*S*) are α to the plane and correlated to each other in NOESY (Supplementary Fig. 8-13).

The heteroyohimbine/yohimbine/corynanthe C3-oxidase (HYC3O) and C3-reductase (HYC3R) have broad substrate acceptance

The *Rauvolfia* oxidase and reductase duo described above exhibited remarkably large substrate acceptance. *R. tetraphylla* oxidase oxidized rauwolscine, yohimbine, and alloyohimbine (yohimbine type), ajmalicine and tetrahydroalstonine (THA) (heteroyohimbine type), and corynantheidine and mitragynine (corynanthe type) to the corresponding 3-dehydro MIAs (Fig. 3a). Out of all authentic substrates, alloyohimbine was isolated from *Corynanthe johimbe* (Yohimbe; Rubiaceae) bark in this study (Supplementary Fig. 14-18, Table 1 and 2). In all oxidation reactions, the C3-dehydrogenation was evident from the 2 amu *m/z* loss and product UV absorption maxima at 350 nm (Fig. 3a, Supplementary fig. 1). RsCAD7 could reduce the yohimbine and heteroyohimbine types of 3-dehydro intermediates to corresponding 3*R* products, but was inactive with corynanthe intermediates (Fig. 3a, Supplementary fig. 1). Based on these results, we named the oxidase and reductase pair the heteroyohimbine/yohimbine/corynanthe C3-oxidase (HYC3O) and C3-reductase (HYC3R), respectively. Albeit being promiscuous, RtHYC3O showed negligible activities with geissoschizine methyl ether (Fig. 3a), rauwolscine's 20*R*-epimer corynanthine, corynantheidine's 20*S*-epimer dihydrocorynantheine, and mitragynine's 20*S*-epimer speciogynine. RtHYC3O also did not act on 3-epi-ajmalicine, 3-epi-THA, or 3-epi-mitragynine (speciociliatine), indicating that 3*S*-stereochemistry is needed for these substrates. The structures can be found in Figure 1.

We confirmed the structures of 3-epi-ajmalicine, 3-epi-THA (akuammigine), and 3-epi-yohimbine (pseudoyohimbine) by 1D/2D-NMR (Supplementary fig. 19-36, Table 1 and 2). Specifically, we isolated 3-epi-ajmalicine from *C. occidentalis*, and 3-epi-THA/3-epi-yohimbine from large in vitro reactions with corresponding 3*S* substrates. The NMR spectra were consistent with their proposed structures [26,35–38]. Notably, heating 3-epi-THA was necessary to reduce the conformer exchange in NMR [37].

Despite degradation issues with 3-dehydro forms of yohimbine/rauwolscine/ajmalicine during purification, we successfully isolated 3-dehydro-THA, providing a strong support for HYC3O's dehydrogenation activity. This intermediate showed improved stability due to the rearrangement from an iminium to an enamine, resulting in 3,14-dehydro-THA, facilitating purification (Supplementary fig. 37-42, Table 1 and 2). Dehydrogenation was confirmed by the disappearance of the H3 resonance and the appearance of H14 alkene resonances (δ 4.98, δ 96.4), replacing the H14 methylene signals in the THA substrate (Supplementary fig. 43-48). The very broad peak observed for 3-dehydro-THA in LC-MS/MS chromatogram (Fig. 3a) may result from iminium/enamine equilibrium and/or conformer exchange. Additionally, we identified 3-epi-mitragynine (speciociliatine) with a commercial authentic standard. With these results, the identity of 3-epi-corynantheidine and 3-epi-alloyohimbine were easily inferred from their LC-MS/MS and UV absorption profiles, which are nearly identical to their 3*S*-epimers and consistent with all other 3*R*-MIAs in our study (Supplementary fig. 1).

Diverse HYC3O and HYC3R enzymes are responsible for MIA C3 stereochemistry inversion in three plant families

Reserpine and other 3*R* MIAs are widely documented in MIA producing species. HYC3Os and HYC3Rs likely play roles in their biosynthesis across different lineages. Sequence alignments identified numerous homologs in three plant families (Supplementary data). In the *C. roseus* genome [39], inferred protein sequences of the physically clustered orthologs (Fig. 2a) shared 75% and 82% amino acid identity with RtHYC3O and RsHYC3R, respectively. A phylogenetic analysis revealed common ancestry among HYC3Os in the Rubiaceae, Gelsemiaceae, and Apocynaceae families (Fig. 2b; see a larger phylogenetic analysis including diverse protein family members that also includes relatives in Gentianaceae [*Eustoma grandiflorum*] and the outgroup species grapevine [*V. vinifera*] in Supplementary fig. 49). This BBE-like oxidase lineage also includes the acetylstemmadenine oxidases (ASOs) found only in Apocynaceae, which participate in late-stage MIA oxidation for tabersonine and vinblastine biosynthesis. The HYC3R clade (Fig. 2c) distinguished itself from other CAD-like, strictosidine aglycone reductases, such as geissoschizine synthase (GS; which we return to below) [18] and DCS. The HYC3R clade also included the Wieland–Gumlich aldehyde synthase (SnvWS) from *Strychnos nux-vomica* (Loganiaceae: Gentianales) [20] , CrTHAS2, and RsVR, which all reduce a double bond geometrically analogous or similar to the C3-*N*4 double bond in their own substrates: 18-hydroxy-norfluorocurarine, 3',4'-anhydrovinblastine iminium, and vomilenine.

To verify their biochemical roles, we expressed the HYC3O/HYC3R orthologs from *C. roseus* and *O. elliptica* (Apocynaceae), *H. patens*, *M. speciosa*, and *M. parvifolia* (Rubiaceae), and *G. sempervirens* (Gelsemiaceae) in yeast. These enzymes showed great sequence diversity, sharing 57-76% amino acid identity among each other.

The HYC3Os demonstrated remarkably large substrate spectra, yet with distinct substrate preferences (Fig. 3d). RtHYC3O showed the highest dehydrogenation activity with rauwolscine, consistent with its role in reserpine biosynthesis. Interestingly, it also exhibited strong activity with mitragynine, an unnatural substrate for a *Rauvolfia* enzyme. Four other HYC3Os also showed high activity with rauwolscine. A notable exception was MsHYC3O, which was inactive with rauwolscine, but highly active with yohimbine, ajmalicine, and THA. CrHYC3O, inactive with ajmalicine and yohimbine, showed activity with THA and yohimbine. HpHYC3O displayed balanced activity among the seven tested substrates, a feature also shared by GsHYC3O2. In contrast, GsHYC3O1 preferred heteroyohimbine and yohimbine substrates, with negligible activity with the corynanthe types. Strictosidine, with its bulky glucose, was not a HYC3O substrate.

For the reduction reaction, the HYC3Rs from *C. roseus*, *O. elliptica*, and *H. patens* effectively reduced many heteroyohimbine and yohimbine intermediates (Fig. 3e). Notably, the fully aromatized beta-carbolines serpentine and alstonine (Fig. 1) were not HYC3R substrates. We additionally confirmed that SnvWS, CrTHAS2, and RsVR did not bear HYC3R activity, despite their close phylogenetic ancestry with HYC3Rs (Fig. 2c).

Interestingly, searches in the transcriptomes/genomes of *M. speciosa, M. parvifolia* (Rubiaceae) and *G. sempervirens* [12,22,40] did not yield any putative genes within the HYC3R phylogenetic clade. Close relatives of unknown function were, however, discovered in *O. pumila* (Rubiaceae)

and *Alstonia scholaris* (Apocynaceae) in a large-scale phylogenetic analysis (Supplementary fig. 50). Nonetheless, total proteins from *M. parvifolia* and *M. speciosa* leaves were able to reduce most 3-dehydro MIAs (Fig. 3e). We further screened highly expressed CAD-like genes from *M. parvifolia* and *G. sempervirens* and identified two putative *HYC3R*s of distant evolutionary ancestry (Fig. 2c). MpHYC3R, previously identified as a THAS in *M. speciosa* [41], showed closer evolutionary ties with CrRedox1, an enzyme required for stemmadenine biosynthesis. On the other hand, GsHYC3R was placed phylogenetically closer to the strictosidine aglycone reductases, yet it showed poor aglycone reducing activity. Both Mp/GsHYC3Rs could successfully reduce the heteroyohimbine/yohimbine intermediates but not corynanthe types (Fig. 3e). In fact, all tested HYC3Rs showed negligible or no activity with corynanthe intermediates. The 3-dehydro corynanthe reducing activity found in *Mitragyna* leaves suggests that additional HYC3Rs may not share high sequence similarity with CAD-like reductases, which would be responsible for the biosynthesis of speciociliatine and other 3*R*-corynanthe MIAs in this genus.

Active site architectures reveal the substrate spectra of HYC3Os and the dual catalytic activities of HYC3Rs

Homology modeling and substrate docking using AlphaFold 3 and the Molecular Operating Environment (MOE) revealed the putative active sites and reaction mechanisms in our HYC3Os and HYC3Rs (Fig. 4a-e, and Supplementary Fig. 51-73). For RtHYC3O, hydrogen bonds between Q432/E434 and rauwolscine's indole and carbomethoxy groups ensured the correct substrate orientation (Fig. 4a and c). FAD then initiated hydride abstraction at the H3 position via its reactive N5, leading to the formation of 3-dehydro-rauwolscine (Fig. 4c). The docking pose of mitragynine was similar, with its 9-methoxyl group fitting into a spacious binding pocket (Fig. 4a). All HYC3Os, except HpHYC3O, contained two conserved Cys/His residues forming bicovalent bonds with FAD, which is a key feature of the BBE-like flavin-containing oxidases [42]. However, the C173A substitution in HpHYC3O resulted in a single His-FAD bond, which in turn created a more spacious active site to accommodate structurally diverse substrates (Fig. 4a). Hydrogen bonds among Q400, R286, N398, and substrate's N4 facilitated proper docking. The E431A substitution in HpHYC3O also created a binding pocket that could accommodate the carbomethoxy group in geissoschizine methyl ether, making it a suitable substrate, unlike other HYC3Os. In both Rt/HpHYC3Os, MIA substrates adopted similar binding postures, with the indole group aligning well with FAD through π-stacking interactions (Supplementary figure 51-66). The distance between FADs' reactive N5 and the H3 position in MIA substrates (under 3 Å), along with docking scores, generally correlated with observed substrate reactivity (Supplementary figure 51-73).

Since HYC3Rs are CAD-like reductases with a common ancestry to strictosidine aglycone reductases such as THAS and DCS, we next tested their activity with the aglycone. We expressed all our HYC3Rs, along with several control strictosidine aglycone reductases, in a de novo strictosidine aglycone-producing yeast strain AJM-dHYS that was previously engineered for the biosynthesis of ajmalicine, ajmaline, catharanthine, and vindoline [19,43,44]. All HYC3Rs, as well as SnvWS and CrTHAS2, could reduce strictosidine aglycone to THA to various degrees (Fig. 4f). Some enzymes additionally produced ajmalicine, mayumbine, and an unknown *m/z* 353 isomer (red dots, Fig. 4f), all of which are THA isomers.

In contrast, YOS, GS, and DCS outside the HYC3R clade lacked HYC3R activity. These enzymes displayed a broad product spectrum when reducing strictosidine aglycone. In addition to THA, RsYOS, RtYOS, and CrDCS (THAS1) also produced four yohimban epimers: rauwolscine, yohimbine, corynanthine, and alloyohimbine in decreasing amounts, as well as heteroyohimbine epimers mayumbine, and ajmalicine (Fig. 4f). CrDCS (THAS1) predominantly produced the corynanthe type demethyldihydrocorynantheine, which was also detected in trace amounts in Rs/RtYOS reactions (Fig. 4f). These activities aligned with the structural diversity of strictosidine aglycones [15,18,21,22], reflecting their orthologous nature.

Strictosidine aglycone naturally exists as a mixture in equilibrium, including two primary forms: cathenamine (20,21-dehydro-THA) and 19-epi-cathenamine. Cathenamine can spontaneously protonate at C20 in solution, forming 4,21-dehydro-THA, which is then reduced by THAS-bound NADPH to produce THA [21] (Fig. 4e). Reducing 3,4-dehydro-THA and 4,21-dehydro-THA requires slight adjustments in positioning these two substrates relative to NADPH for proper double bond reduction. Compared to CrTHAS2, which lacks HYC3R activity, RsHYC3R possesses a more spacious active site, accommodating both substrates (Fig. 4b, 4e, and Supplementary fig. 67-73). The F127 residue in RsHYC3R further stabilized 3-dehydro-THA. In contrast, CrTHAS2's active site was significantly narrower, due to the presence of two bulky residues, V290 and I313, which replace two alanine residues that occupy the same position in RsHYC3R (Fig. 4b). Additionally, the inward conformation of W60 in CrTHAS2 further constricted the active site, preventing it from accommodating 3-dehydro-THA for effective reduction. The spacious active site in RsHYC3R, however, allowed for proper docking of 3-dehydro-rauwolscine and other analogs, facilitating HYC3R's dual catalytic activity.

The reserpine BGC of *R. tetraphylla* and *C. roseus* likely evolved as a segmental duplication of the GS BGC, which itself predates the Gentianales crown group

As noted above, the strong sequence similarity and functional promiscuity among CADs and BBE-like oxidases in Gentianales warranted a rigorous genomic study of these biosynthetic enzymes. As had been considered (but rejected) for the STR BGC [13], it is conceivable that BGCs encoding biosynthetic assemblies for complex MIAs, interrelated through a universal precursor, could have arisen via convergent evolution. This hypothesis is less parsimonious than evolution of a single ancestral BGC containing all relevant gene families, followed by later duplication and functional diversification. While it seems clear that various HYS, THAS, DCS, YOS, GS, and HYC3R enzymes involved in converting strictosidine aglycone into heteroyohimbine-, yohimbine-, and corynanthe-type MIAs resulted from ancient duplications (Fig. 2c), their common ancestry from a single BGC remains unexplored. We therefore extended our genomic search in *R. tetraphylla* and *C. roseus* to locate loci encoding CADs (HYC3R, YOS, GS, HYS, THAS, Redox1, and DPAS) and BBE-like oxidases (HYC3O and ASO; Fig. 2b).

This search uncovered a new *C. roseus* BGC at least partially homologous to the reserpine BGC, which we named the GS BGC (Fig. 2a). The GS BGC encodes ASO and its uncharacterized paralog, THAS2, two GS paralogs, two CAD-like reductases (CrCAD1/2) with unknown functions [5] , an 8-hydroxygeraniol oxidoreductase (8HGO) from the early iridoid pathway [45] , and several other uncharacterized 8HGO paralogs. In *R. tetraphylla*, the GS BGC contains all these components except ASO, consistent with the absence in this lineage of iboga/aspidosperma MIAs requiring ASO. Additionally, in *C. roseus*, genes encoding Redox1 and DPAS, which are

both involved in iboga/aspidosperma MIA biosynthesis, are tightly clustered with three paralogs of *CrCAD5* having unknown function (Fig. 2a) [5]. Moreover, *HYS* and *THAS4* are closely associated with a *HYS* paralog in *C. roseus*. This duplicative homology of BGCs that functionally integrate complex MIA production highlights a likely mechanism - gene or genomic block duplication and neofunctionalization - driving MIA emergence and diversification within Gentianales.

To further investigate this potentially unifying mechanism, we searched for genomic blocks associated with the reserpine BGC (*HYC3O*/*HYC3R*), GS BGC (*ASO*/*GS*), *Redox1*/*DPAS* cluster, and *HYS*/*THAS4* cluster across the Gentianales order. Initial syntenic searches with the SynFind [46] application in CoGe [47], using as query the *R. tetraphylla* scaffold on which the reserpine and GC BGCs were located, uncovered many blocks in various Gentianales species (*A. syriaca*, *O. pumila, M. speciosa, G. elegans, C. roseus*) and *V. vinifera* that showed strong matches, the shared synteny of which was subsequently verified using MCScan in the JCVI package [48] (Fig. 2a, d). *M. speciosa* [49] , with its known lineage-specific whole genome duplication (WGD), displayed two subgenomes mapping to the same *R. tetraphylla* scaffold. Three distinct blocks on different chromosomes of the *C. roseus* genome (NCBI ASM2450571v1, reannotated using GeMoMa [50] as with the assembly of *R. tetraphylla*) [39] mapped to the single *R. tetraphylla* scaffold, and to single scaffolds of all other species except tetraploid *M. speciosa*, for which two scaffolds matched subgenome-wise (Fig. 2d, Supplementary fig. 74). These results indicated that two chromosomal translocations occurred following the split of *C. roseus* and *R. tetraphylla*.

In *C. roseus*, the *THAS4*/*HYS* cluster remained physically linked with its reserpine cluster, while its GS BGC relocated to one of the three genomic blocks via its two species-specific translocations. *CrTHAS4* and *HYS* are tandem duplicates and sister genes in our large, multispecies trees (Supplementary Fig. 50), where they appear with a sister clade that includes *R. tetraphylla* and *A. scholaris* homologs. *R. tetraphylla THAS4* and *HYS* genes were not located in the same genomic block as its reserpine and GS BGCs, preventing confirmation of a *THAS4*/*HYS* cluster in *Rauvolfia*. The phylogenetically uncovered *A. scholaris* genes merit further investigation in future studies. We did not detect a *Redox1*/*DPAS* cluster in *R. tetraphylla* in the same syntenic context, which given Apocynaceae phylogeny [51] and the presence of Redox1/DPAS in *A. scholaris* (an early diverging member of Apocynaceae) and *Vinca minor* and *Tabernaemontana* spp. (close relatives of the *C. roseus* and *R. tetraphylla* sister taxon pair), suggests that the Redox1/DPAS cluster may have been lost in *R. tetraphylla*.

Regarding the reserpine BGC, we found it to be shared exclusively between *R. tetraphylla* and *C. roseus*. Their small genomic segments containing the reserpine BGC appeared altogether absent in other examined species (Fig. 2d), indicating that it likely represents a genomic insertion in the rauvolfioid common ancestor of the two species. In contrast, the genomic context for the GS BGC was conserved all the way back to the Gentianales common ancestor with *V. vinifera* (Fig. 2d). Indeed, CAD homologs exist in the same syntenic context as those active in GS biosynthesis through all Gentianales examined, even being present in the non-MIA-producing species *A. syriaca* and *V. vinifera* (Fig. 2d), those in the latter perhaps representing ancestral genes from which the GS cluster eventually evolved. To test this hypothesis, we cloned two syntenic CAD-like reductases from *V. vinifera* (monolignol aldehyde reductase MLR1/2 thereafter, Fig. 2c) and confirmed their identities as bona-fide CADs for reducing cinnamyl and coniferyl aldehyde to

their alcohols (Supplementary Fig. 75). The newly identified GsHYC3R (Figs. 2c, 3e) was found located in its GS cluster. As such, the ancestral genomic region now containing the reserpine and GS BGCs in *R. tetraphylla* and other taxa may have served as fertile ground for the evolution of multiple biosynthetic pathways and products, many of which remain to be discovered.

The proximity and gene content similarity between the reserpine and GS BGCs in the rauvolfioid ancestor, coupled with the apparent genomic insertion of the reserpine BGC, leads us to propose that the reserpine BGC originated as a segmental (tandem) block duplication of the GS BGC region. Subsequent tandem (within BGC) gene duplications (and/or deletions, such as the disappearance of ASO in *Rauvolfia*) and functional divergence within both BGCs likely shaped their current forms. It seems likely that the *THAS4*/*HYS* and *Redox1*/*DPAS* clusters evolved within Apocynaceae, therefore postdating the evolution of the GS BGC, which is present even in *M. speciosa* of Rubiaceae, the first-diverging lineage of Gentianales. As all genes in the *THAS4*/*HYS* and *Redox1*/*DPAS* clusters encode CADs, it is possible that these clusters evolved in situ and de novo within the large, CAD-rich syntenic block conserved from Gentianales to *V. vinifera*. It is also possible that they derived from partial segmental duplications stemming from the older GS BGC. Further research into these possibilities may shed additional light on the intertwined MIA biosynthetic pathways.

Strictosidine biosynthesis evolved once in the Gentianales stem lineage

Strictosidine biosynthesis provides the precursor immediately upstream of the CAD families involved in strictosidine aglycone reduction. The occurrence and evolution of both the GS and reserpine BGCs, alongside the STR BGC, are fundamental drivers of the emergence and diversification of MIAs. The STR BGC, containing genes encoding STR, TDC, and a MATE transporter, have been confirmed in *C. roseus* [4], *Rhazya stricta* (Apocynaceae) [52], *O. pumila* [13], *M. speciosa* [40,49], and *G. sempervirens* [12]. We similarly found a syntenic block in *R. tetraphylla* that contained a complete STR BGC (Fig. 5). However, in *A. syriaca* (also Apocynaceae), the corresponding syntenic block contains only a downstream *TDC* homolog, while a block in *Gardenia jasminoides* (Rubiaceae) contains homologs for *TDC* and *MATE* but lacks STR (Fig. 5). This reveals a mosaic distribution of a complete STR BGC across Gentianales phylogeny.

With complete STR BGCs present in some Apocynaceae, *Gelsemium*, *M. speciosa* and *O. pumila*, all MIA producing species, one possibility is that these BGCs convergently assembled in these species, in the same syntenic genomic context. Much more parsimonious is that they descend from a single STR BGC origin, a hypothesis also favored by the authors of the *O. pumila* and *Pachypodium lamarei* (Apocynaceae) genomes [53]. Importantly, by including the grapevine genome, which diverged some 148 Mya from other core eudicots [14] , we uncovered a *Vitis* syntenic block only containing *TDC* and *MATE* homologs, similar to that of *G. jasminoides* (Fig. 5). Therefore, we hypothesize that *STR* joined closely proximate *TDC* and *MATE* genes only once, at some point in the Gentianales stem lineage. Given the age of the latter lineage (~135 My) and the Gentianales crown group (~118 My) [14], we propose that the first committed step in complex MIA biosynthesis had already evolved by the early-mid Cretaceous, likely time-coincident with establishment of the ancestor of the GS BGC.

Discussion

CAD-like reductases, BBE-like oxidases, and CYPs are key players in MIA diversification
Our research here began with investigation of reserpine biosynthesis, which features a prevalent yet enigmatic 3*R* stereochemistry. The Pictet-Spengler reaction between tryptamine and secologanin may generate two diastereomeric products: strictosidine (3*S*) and vincoside (3*R*), both considered earlier as central MIAs precursors to downstream MIAs bearing respective C3 stereochemistry. However, the purification and molecular cloning of strictosidine synthase, as well as radiolabel feeding studies, have established strictosidine as the sole precursor for nearly all naturally occurring MIAs [24,25,54,55].

The emergence of strictosidine synthase (STR) activity in the Gentianales stem lineage laid the foundation for MIA biosynthesis, establishing one of the largest classes of alkaloids, with over 3,000 distinct structures. The 3-gene *TDC*, *MATE*, and *STR* cluster, which is necessary for strictosidine formation in the vacuole, emerged as early as 135 Mya in the Gentianales stem lineage [4,13]. Following this biosynthetic step, nucleus-localized strictosidine β-glucosidase (SGD, Fig. 6) catalyzes strictosidine deglycosylation to yield a highly reactive aglycone. This aglycone likely plays a defensive role against herbivores, reacting with protein nucleophiles to induce coagulation and degradation [56]. The recruitment of cinnamyl alcohol dehydrogenases (CADs) from primary metabolism has been fundamental for the complexity of MIA biosynthesis across Gentianales. These reductases, such as heteroyohimbine synthases (HYS), demethylcorynantheidine synthases (DCS), yohimban synthases (YOS), and geissoschizine synthases (GS), reduce the reactive aglycone to form heteroyohimbine, yohimbine, and corynanthe scaffolds for further MIA diversification (Fig. 6) [7,15,18,21,22].

Among these structural modification pathways, geissoschizine cyclization via cytochrome P450 monooxygenases (CYPs) represents a critical diversification point. These homologous CYPs catalyze dehydrogenation adjacent to either the indole nitrogen (*N*1) or piperidine nitrogen (*N*4), enabling C5-C16, C7-C16, or C2-C16 cyclizations that form the sarpagan, akuammiline, and strychnos skeletons, respectively (Fig. 6) [7,57–59]. When substrates lack a free enol, as with heteroyohimbines, these CYPs instead promote piperidine aromatization, yielding serpentine and alstonine [57,58]. Notably, none of these CYPs could perform only C3-*N*4 oxidation, and were therefore not able to participate in MIA C3-stereochemistry inversion.

Our discovery of tandem oxidation and reduction at C3 by HYC3O and HYC3R has resolved the long-standing question for 3*R* MIA emergence. The 3*R* MIAs, as well as the oxindoles derived from 3*R* MIA precursors, are biosynthesized after C3 stereochemistry inversion by HYC3O/HYC3R activities (Fig. 1). In parallel to geissoschizine oxidation, the BBE-like HYC3Os catalyze specific C3-*N*4 MIA dehydrogenation without further aromatizing the piperidine ring. Our biochemical characterization and protein homology modeling revealed the broad substrate range of HYC3Os, encompassing heteroyohimbine, yohimbine, and corynanthe substrates (Fig. 4a). Despite structural diversity, alkaloid substrates consistently form multiple hydrogen bonds with the active site and π-stacking interactions with FAD (Supplementary fig. 50-73). Interestingly, the C173A substitution in HpHYC3O resulted in a single histidine attachment to FAD, distinguishes itself from other BBE-like oxidases with the typical two-bond FAD connection through both cysteine and histidine [42]. This substitution appears to have contributed to broadening active sites of HpHYC3Os and consequently the substrate spectra (Fig. 3d and 4a). Interestingly, the key oxidase ASO for iboga and aspidosperma MIA

biosynthesis from *C. roseus*, *Vinca minor, Tabernaemontana. elegans* and other Apocynaceae species also lacks the critical cysteine residue and likely contains only a single histidine-FAD bond [5,9] , which theoretically leads to reduced FAD's redox potential [42] . The change of HpHYC3O's redox potential and the evolutionary drive for losing this bond in HpHYC3O and ASOs requires further investigation.

Following HYC3O-mediated C3-*N*4 dehydrogenation, the *si*-face reduction by HYC3R inverts C3 stereochemistry to produce 3*R* MIAs, resulting in a dramatic structural shift. The perpendicular indole orientation within 3*R* MIAs alters both chemical reactivity and biological properties (Fig. 1). Consequently, downstream enzymes must possess distinct active sites to selectively accommodate 3*S* or 3*R* substrates. A key example is the oxindole formation via C7-oxidation and subsequent pinacol rearrangement, where the CYP enzyme specifically recognize only 3*R* substrates [33]. This perpendicular indole orientation is likely essential for the antihypertensive activity of reserpine, which also requires downstream enzymes to adapt to the stereochemically inverted, 3-epi-rauwolscine precursor.

Recruitment of BBE-like oxidases and CAD-like reductase in MIA biosynthesis

BBE-like oxidases, while widespread across plants, remain largely uncharacterized. Recent studies have uncovered various roles for these enzymes in *Arabidopsis thaliana*. AtBBE13/15 are bona fide cinnamyl/coniferyl alcohol oxidases forming corresponding aldehydes, while AtBBE1/2/19/21 and AtBBE22/23 act as oligogalacturonide oxidases (OGOX) and cellodextrin oxidases (CELLOX), respectively (Fig. 2b) [60–64]. OGOX and CELLOX catalyze anomeric carbon oxidation in diverse oligosaccharides, contributing to cell wall homeostasis and interestingly, plant immune responses, as these oligosaccharrides are damage-associated molecular patterns (DAMPs) derived from the cell wall. Our phylogenetic analyses suggest common ancestry between HYC3Os and these enzymes for cell wall lignification and homeostasis. Systematic investigation on this regard may reveal the evolution origin of HYC3O and ASO from primary metabolism.

Plant CADs and CAD-like reductases constitute a large enzyme family with largely unknown functions. Lineage-specific radiations significantly contributed to their diversity and function [65] . Phylogenetic analysis (Fig. 2c) and the biochemical characterization of two bona fide CADs from the syntenic *Vitis* cluster (Fig. 2c, Supplementary fig. 75) suggest a possible link between MIA-reducing CADs and ancestral CADs for monolignol biosynthesis. Beyond their strictosidine aglycone reducing activity (4,21-reduction), HYC3Rs also acquired MIA 3,4-reduction capacity, a natural extension of the similar iminium reduction reactions observed from these evolved CADs. Additionally, Redox1 catalyzes the 3,4-reduction of a preakuammicine aldehyde-like intermediate for stemmadenine biosynthesis [7]. Dihydroprecondylocarpine synthase (DPAS) reduces the 19,20-double bond in *O*-acetylprecondylocarpine, triggering deacetylation and C16/17 fission to yield dehydrosecodine (Fig. 7) [9]. GS also catalyzes this reaction, demonstrating another example of dual catalytic activity acquired by these reductases [5,66]. The formation of dehydrosecodine by DPAS/GS is a remarkable diversification point in MIA biosynthesis, as this highly reactive intermediate cyclizes into iboga, aspidosperma, and pseudoaspidosperma MIA types, including the famous anticancer drug vinblastine and the opioid agonist ibogaine [5,7,9,67,68].

HYC3O and HYC3R genomic organization and function in Gentianales

The *R. tetraphylla* reserpine BGC contains both *HYC3O* and *HYC3R* for reserpine biosynthesis, alongside many paralogs of CAD-like reductases like vomilenine 1,2-reductase (VR) and 1,2-dihydrovomilenine 19,20-reductase (DHVR), participating in successive reduction for ajmalcine biosynthesis [19]. In contrast, the *C. roseus* reserpine BGC lacks extensive CAD duplications, instead containing single copies of *HYC3O* and *HYC3R* (Fig. 2a), the protein products of which likely facilitate akuammigine (3-epi-THA) production in the roots [26]. Interestingly, a gene encoding a secologanin synthase (*SLS*) paralog, sharing 97% identity at the inferred amino acid level, is also integrated into this cluster (Fig. 2a). The physiological roles of this *SLS* paralog, along with the E3-ubiquitin ligase (*E3UL*) present in both the *R. tetraphylla* and *C. roseus* reserpine clusters, remain to be further explored.

While HYC3O/HYC3R are found across Gentianales, they are not clustered outside rauvolfioid Apocynaceae, and HYC3R appeared multiphyletic distribution (Fig. 2c). *G. sempervirens* (and *G. elegans*) genome encodes HYC3O1 and 2, yet it lacks a HYC3R ortholog closely related to other HYC3Rs. Surprisingly, a different CAD-like reductase encoded by the GS cluster, with closer ancestry with strictosidine aglycone reductases (e.g., HYS and YOS), has independently evolved HYC3R activity (Fig. 3e), despite showing limited strictosidine aglycone reduction activity (Fig. 4e). This is an interesting example of parallel evolution of CAD-like reductases among different plant families. HYC3Os in *Gelsemium* species may play a role in oxidizing intermediates for gelsemine biosynthesis, as this MIA includes a disrupted sarpagan skeleton, potentially derived from oxidative rearrangement. However, the physiological roles of both GsHYC3O and GsHYC3R (and their orthologs in *G. elegans*) remain to be fully elucidated.

In the Rubiaceae family, *H. patens* expresses both a functional HYC3O and a functional HYC3R within the monophyletic HYC3R clade (Fig. 2c). Although the physical existence of an *HYC3O*/*HYC3R* cluster cannot be confirmed due to the absence of a sequenced genome, *H. patens* is known for producing various oxindoles, such as pterodine and speciophylline [32,69], which likely derive from the 3*R* intermediate akuammigine. Thus, a three-step reaction that converts strictosidine aglycone to akuammigine by HpHYC3O/HpHYC3R provides the relevant substrates *in planta*.

M. speciosa and *M. parvifolia* in the Rubiaceae family also produce abundant heteroyohimbine oxindoles (e.g., speciophylline, mitraphylline) and corynanthe oxindoles (e.g., speciociliatine, corynoxine) (Fig. 1) [31,70,71]. Ms/MpHYC3O participate in the 3*S*-intermediate oxidation, but neither species contains a HYC3R ortholog in the monophyletic HYC3R family. Instead, another CAD-like reductase sharing ancestry with Redox1 has acquired both HYC3R and THAS activity (Fig. 2d and Fig. 3e). The tested Ms/Mp HYC3Rs are limited to reducing heteroyohimbine intermediates for oxindole formation, lacking activity toward 3-dehydro corynanthe intermediates. In fact, none of our cloned and expressed HYC3Rs showed this activity, yet the *M. parvifolia* and *M. speciosa* leaf proteins contained HYC3R enzymes for reducing all three types of 3-dehydro intermediates to 3*R* products (Fig. 3e). It is plausible that a CAD-like reductase evaded our screening experiment or that this particular HYC3R is not CAD-related. The independent re-emergence of HYC3R function across species, as observed in *Mitragyna* and *Gelsemium*, suggests a recurring evolutionary adaptation that may be driven by selective

pressures favoring biosynthesis of particular MIAs. This phenomenon presents an intriguing area for further evolutionary and functional exploration.

STR, GS and reserpine BGC contribute to MIA emergence and evolution

Through detailed genome structural analysis, we were able to chart the evolution of the reserpine and GS BGCs within a diversity of Gentianales species, using *Vitis* as an outgroup. The genomic context for both BGCs was a single genomic block, dating back as early as 148 Mya (with the split of grapevine from other core eudicots), only recently being distributed into 3 distinct chromosomal regions in *C. roseus* via two species-specific translocation events following its common ancestry with *R. tetraphylla*. Given the similar gene content, close genomic proximity, the biochemical functional homology of the reserpine and GS BGCs, and the apparent rauvolfioid-specific insertion of the reserpine BGC into the ancestral block shared among all other Gentianales and *Vitis* (Fig. 2d), we hypothesize that the reserpine BGC represents a segmental (tandem) duplicate of the GS BGC that later diversified functionally to specialize in distinct MIA biosynthesis. The CADs encoded in these blocks largely retained the strictosidine aglycone reduction activity (Fig. 4c) while having evolved to contain other reduction activities to participate in many different MIA biosynthetic pathways. The BBE-like oxidases in these blocks, namely HYC3Os and ASOs, are gateway enzymes to the large 3*R*-MIA and oxindoles, and the widely distributed iboga, aspidosperma, and pseudoaspidosperma MIAs in Apocynaceae (Fig. 7).

The emergence of the GS BGC, from a genomic context traceable to before Rubiaceae diverged from other Gentianales, likely coincides with the origin of the STR BGC to form the foundation for the vast diversity of MIAs observed in Gentianales today. The retention and evolutionary dynamics of the region's basic CAD-encoding gene contents over many millions of years may reflect common as well as diversified selective pressures acting on multitudinous angiosperm lineages and species (i.e., within the stem lineage shared by *V. vinifera* plus Gentianales). The considerable homology of the reserpine and GS BGCs highlights the likely importance of segmental block duplication with local insertion in MIA biosynthetic evolution. Given the presence and CAD-specific nature of the *THAS4*/*HYS* and *Redox1*/*DPAS* clusters we located in *C. roseus*, and the existence of numerous CAD homologs in the entire syntenic region we examined among all species, it may be that these clusters represent further, smaller segmental block duplications that occurred within Apocynaceae. The genes encoding these CADs may have evolved instead from de novo, in situ duplications and neofunctionalizations of CAD family members belonging to ancient subclades of those enzymes that already peppered the region. Further research may be able to discern between these possibilities and others.

In summary, our evolutionary genomic analysis revealed previously unrecognized details about the genomic basis for MIA formation and diversification, suggesting that a tandem duplicative ancestry of the reserpine and GS BGCs could reflect a broader paradigm in which homologous gene duplications—and, in this case, small genomic block duplications as well—drive biosynthetic diversification in Gentianales [72,73]. The identification and biochemical characterization of diverse HYC3O and HYC3R enzymes uncovers the mechanistic basis for 3*R* MIA formation while elucidating the initial three steps of reserpine biosynthesis. These findings expand our understanding of MIA pathways beyond the well-characterized vinblastine and ajmaline biosynthetic routes. For reserpine biosynthesis, the remaining steps include 17-hydroxyl

methylation, indole C11 methoxylation, C18-hydroxylation, and final 18-OH esterification with eudesmic acid (3,4,5-trimethoxybenzoic acid). These transformations are likely catalyzed by cytochrome P450 enzymes, methyltransferases, and a unique transferase for eudesmic acid. Identifying and characterizing these enzymatic steps will pave the way for the total biosynthesis of reserpine through synthetic biology approaches. Identifying the missing HYC3R for corynanthe intermediate reduction would also facilitate the biosynthesis of intriguing corynanthe oxindoles, such as corynoxine, which possesses significant painkilling properties [74].

Materials and Methods:

Bioinformatics and cloning

The published genome of *R. tetraphylla* (GCA_030512225.1, available at NCBI) was reannotated (with default settings) using Gene Model Mapper version 1.9 by using available reference assembly annotations of several related species *C. roseus* (GCA_024505715.1), *A. syriaca* (CoGe gid61699), *O. pumila* (GCA_016586305.1), and *Calotropis gigantea* (NCBI PRJNA400797). For sequence discovery to synthesize enzymes for experimental analyses, the annotated genomes of *R. tetraphylla* [15], *C. roseus* [39], *G. sempervirens* [12] , *G. elegans* [75] , *M. speciosa* [40,49] were analysed with CoGeBLAST (https://genomevolution.org/coge/) [76]. The sequences of HYC3O/HYC3R have been deposited to NCBI GenBank (PP911565-911588, and OQ591889 for RsHYC3R). The open reading frames (ORFs) of all HYC3Os/HYC3Rs except Rs/CrHYC3R, RtYOS, SnvWS, were synthesized and subcloned within the BamHI/Sal sites of pESC-Leu vector (for HYC3O genes) or pESC-Ura vector (for CAD-like reductase genes) (Twist Bioscience, South San Francisco, CA, USA). *RsHYC3R*, *RsYOS*, *CrHYS*, *CrDCS* (*THAS1*), *CrTHAS2*, and *CrHYC3R* (*THAS3*) were amplified from plant cDNA using primer sets 1-12 (Supplementary table 1), and cloned within the BamHI/SalI sites of pESC-Ura and pET30b+ vectors. For yeast expression of HYC3Os and HYC3O/HYC3R combinations, the pESC vectors were mobilized to *Saccharomyces cerevisiae* strain BY4741 (MATα his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 YPL154c::kanMX4) using a standard lithium acetate/ polyethylene glycol transformation procedure. For de novo production of various reduced strictosidine aglycone products, the vectors containing CAD-like reductases were mobilized to *S. cerevisiae* strain AJM-dHYS [44].

Chemical standards and alkaloid purification from plants

Authentic chemical standards and substrates were purchased from commercial sources. These included ajmalicine (Sigma Aldrich, St. Louis, MO, USA), yohimbine and corynantheidine (Cayman Chemical, Ann Arbor, MI, USA), corynanthine and rauwolscine (Extrasynthese, Genay, France), and geissoschizine methyl ether (AvaChem Scientific, San Antonio, TX, USA). For alkaloid purification from plants, leaves (100 g) of *C. occidentalis* or *Luculia pinceana* (Rubiaceae; both greenhouse grown) were submerged in ethyl acetate for 1 hr to dissolve alkaloids. The *C. johimbe* bark powder (50 g) was submerged in ethanol for 10 min to dissolve alkaloids. The evaporated extracts were suspended in 1 M HCl and extracted with ethyl acetate. The aqueous phase was basified with NaOH to pH 8 and extracted with ethyl acetate to afford total alkaloids, which were further purified with preparative thin layer chromatography (TLC, Silica gel60 F254, Millipore Sigma, Rockville, MD, USA). For THA, *L. pinceana* alkaloids were separated on TLC with acetonitrile:toluene 1:1 (v/v) to give pure THA (Rf 0.77, 3.7 mg). For 3-

epi-ajmalicine, *C. occidentalis* alkaloids were separated on TLC with toluene: ethyl acetate: methanol 15:4:1 (v/v) to give pure 3-epi-ajmalicine (Rf 0.38, 3.3 mg). For alloyohimbine, the *C. johimbe* alkaloids were separated on TLC with three successive mobile phases. First, acetonitrile: toluene:methanol 15:4:1 (v/v) gave a band with Rf 0.2, which was further separated with acetonitrile:chloroform 1:1 (v/v) to give a band with Rf 0.16. Lastly, this was further separated with hexane:ethyl acetate: methanol 5:4:1 (v/v) to give pure alloyohimbine (Rf 0.37, 3.2 mg). The structures of purified alkaloids were confirmed with mass spectrometry and 1D/2D NMR analyses.

Yeast de novo alkaloid biosynthesis and alkaloid biotransformation

Single colonies of the yeasts carrying various vectors were inoculated in 1 mL synthetic complete (SC) media with 2% (w/v) glucose, and incubated at 30 °C, 200 rpm overnight. The yeasts were pelleted by centrifugation, washed once with water, resuspended in 1 mL SC media with 2% (w/v) galactose, and incubated at 30 °C, 200 rpm overnight. For yeast producing reduced strictosidine aglycone products, the media were directly mixed with equal volumes of methanol for LC-MS/MS analysis. For HYC3O/HYC3R yeasts, the cells were pelleted by centrifugation and resuspended in 0.1 mL 20 mM Tris-HCl pH 7.5 supplemented with 0.5-2 µg alkaloid substrates. The biotransformation took place at 30 °C, 200 rpm overnight, and was mixed with equal volume of methanol for LC-MS/MS analysis. For 3,14-dehydro-THA production, yeast cells carrying RtHYC3O from 200 ml culture were resuspended in 50 ml of 20 mM Tris-HCl pH 7.5 and fed with 5 mg THA at 30 °C, 200 rpm overnight. For 3-epi-rauwolscine, 3-epi-yohimbine, and 3-epi-THA production, in vitro reactions (50 ml) contained 20 mM Tris HCl pH 7.5, 1 mM NADPH yeast lysates from 200 ml cultures expressing RtHYC3O and RsHYC3R and 3-5 mg 3*S* substrates. After 1 hr reaction at 30 °C, the products were extracted with ethyl acetate and separated on preparative TLC (Silica gel60 F254, Millipore Sigma, Rockville, MD, USA) and methanol solvent. The purified products were subsequently analyzed by NMR.

Plant and yeast total protein extraction

Greenhouse-grown *M. speciosa* and *M. parvifolia* leaf tissues (3 g) and 0.2 g polyvinylpolypyrrolidone were ground in liquid nitrogen with mortar and pestle, and were extracted with ice-cold sample buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% (v/v) glycerol). The extracts were centrifuged at 15,000 g for 30 min, and desalted into the same sample buffer with a PD10 desalting column (Cytiva, Wilmington, DE, USA) according to the manufacturer's protocol. The total proteins were desalted once more, and the final samples were stored at -80°C. For HYC3O/HYC3R proteins, yeast cultures (50 ml) expressing various enzymes were pelleted, mixed with 1ml ice cold sample buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% (v/v) glycerol), and lysed by beating the cells at 4 °C with glass beads using a Qiagen tissuelyser II (Qiagen, Germantown, MD, USA). The lysate was centrifuged at 20,000g at 4 °C for 10 min. The aqueous phases containing total yeast soluble proteins were stored at -80°C.

In vitro enzyme assays

For HYC3O activity, the in vitro assay (100 µL) included 20 mM Tris HCl pH 7.5, 100 ng various alkaloid substrates, and 25-45 µL yeast crude lysate containing various HYC3Os (adjusted for the variance of HYC3O activity in each lysate). For each HYC3O, their lysate

amounts were kept constant. The triplicated assays took place at 30 °C for four hours, which was terminated by mixing with equal volumes of methanol for LC-MS analysis. For substrate preference analysis, the 3-dehydro product MS peak areas were compared.

For HYC3O/HYC3R coupled reactions, the in vitro assay (100 µL) included 20 mM Tris HCl pH 7.5, 1 mM NADPH, 100 ng of various alkaloid substrates, 25-45 µL yeast crude lysate containing various HYC3Os, 1-10 µL yeast crude lysate containing various HYC3Rs, and adjusted for the variance of HYC3R activity in each lysate. For reactions with *Mitragyna* leaf total proteins, 20 µg proteins were used instead of yeast HYC3R lysate. The assays took place at 30 °C for four hours, which was terminated by mixing with equal volume of methanol for LC-MS analysis.

To study HYC3R substrate preference, the seven 3-dehydro substrates were first produced using scaled up (25 ml) reactions with RtHYC3O under otherwise identical HYC3O reaction conditions. The reactions were evaporated to ~ 1 ml under vacuum as substrates. The in vitro assay (100 µL) included 20 mM Tris HCl pH 7.5, 1 mM NADPH, and 1-5 µL 3-dehydro MIA substrates in excess and adjusted for the variance in substrate amounts, and 1-10 µL yeast crude lysate containing various HYC3Rs. For each HYC3R, their lysate amounts were kept constant, and the substrates should not be completely consumed by end of the assays. The triplicated assays took place at 30 °C for four hours, which was terminated by mixing with equal volumes of methanol for LC-MS analysis. The 3*R* product MS peak areas were compared for HYC3R substrate preference.

For oxidation and reduction of monolignol and aldehydes, the in vitro assay (100 µL) included 20 mM Tris HCl pH 7.5, 1 mM NADPH (for reduction reactions), 2 µg substrates, 2 µg His-tagged and purified RsHYC3R, VvMLR1/2, or 25 µL yeast total lysate containing RtHYC3O. The assays took place at 30 °C for four hours before LC-MS analysis.

LC-MS/MS and NMR

The samples were analyzed using the Ultivo Triple Quadrupole LC-MS/MS system from Agilent (Santa Clara, CA, USA), equipped with an Avantor® ACE® UltraCore C18 2.5 Super C18 column (50×3 mm, particle size 2.5 μm) as well as a photodiode array detector and a mass spectrometer. For alkaloid analysis, the following solvent systems were used: Solvent A, methanol:acetonitrile:ammonium acetate (1M):water at 29:71:2:398; solvent B, methanol: acetonitrile:ammonium acetate (1M):water at 130:320:0.25:49.7. The following linear elution gradient was used: 0-5 min 80% A, 20% B; 5-5.8 min 1% A, 99% B; 5.8-8 min 80% A, 20% B; the flow during the analysis was constant and 0.6 ml/min. The photodiode array detector range was 200 to 500 nm. The mass spectrometer was operated with the gas temperature at 300°C and gas flow of 10 L/min. Capillary voltage was 4 kV from *m/z* 100 to *m/z* 1000 with scan time 500 ms, and the fragmentor performed at 135 V with positive polarity. The MS/MS was operated with gas temperature at 300°C, gas flow of 10 L min^{-1}, capillary voltage 4 kV, fragmentor 135 V, and collision energy 30 V with positive polarity. NMR spectra were recorded on an Agilent 400 MR and a Bruker Avance III HD 400 MHz NMR spectrometer in acetone-*d6* or $CDCl_3$.

Homology modeling and substrate docking studies

All computational experiments and visualizations were carried out with MOE version 2022.02 and AlphaFold 3 [77] on local computers. All molecular mechanics calculations and simulations were conducted using the AMBER14:EHT forcefield with Reaction Field solvation. Sequence homology searches for HYC3Rs on the Protein Data Bank found CrTHAS2 complexed with NADP+ as the template for homology modeling (PDB ID: 5H81) based on Hidden Markov Model energy scores. Sequence identity of RsHYC3R to the relevant 5H81 domain was 68%. Following a QuickPrep of the 5H81 template protein, homology models were then derived in MOE using default settings and scored using the GBVI/WSA dG method. NADPH was modeled and docked into each homology model using the location of NADP+ in the 5H81 enzyme as the general docking region and default MOE-Dock settings. 3,4-dehydro-HYC3R ligands were then modeled and docked to the HYC3R-NADPH complexes. HYC3O protein models were obtained in AlphaFold using the protein amino acid sequences and FADs as the selected ligands. The models were consequently transferred to MOE for docking. The models were then subjected to a QuickPrep using default settings. HYC3O ligands were then modeled and docked to the HYC3O-FAD complexes. For each ligand docking experiment, 17,000 docking poses were initially generated via the Triangle Matcher method and scored by the GBVI/WSA dG function. A subset comprising the best docking poses was refined by the induced fit method, where the bound ligands and active site residues were submitted to local geometry optimization and rescored once more by the GBVI/WSA dG scoring function. The top-scoring docking poses that were geometrically conducive to their respective reaction mechanisms were retained for subsequent analyses. Cartesian coordinates for all homology models and their ligand complexes can be found in the Supplementary data.

Phylogenetic analyses of enzyme families

Gene family representatives from the *R. tetraphylla* reserpine cluster (our GeMoMa gene model IDs: Catharanthus_roseus_rna_EVM0023822.1_R0, Catharanthus_roseus_rna_EVM0000382.1_R0, and Catharanthus_roseus_rna_EVM0027783.1_R2) were used as query sequences to perform CoGeBLAST searches using TBLASTX with default parameters. The searches were conducted against a selection of species, including *A. scholaris* (gid: 67811), *E. grandiflorum* (gid: 63766), *V. vinifera* (v12x; gid: 19990, *O. pumila* (vv1; gid: 63710), *G. elegans* (v1.0; gid: 64491), *A. syriaca* (v0.3; gid: 61699), *Mitragyna speciosa* (vv1; gid 63699), *R. tetraphylla* (v1.0; gid: 67314), and *C. roseus* (vASM2450571v1; gid: 65259). Sequences were aligned as proteins using Muscle. The alignments were trimmed to remove poorly aligned regions using GBlocks (all three sensitivity options boxes checked) in Seaview, a sequence processing tool. Phylogenetic trees were constructed using IQ-TREE 1.6.12 with the LG substitution model. A total of 1000 bootstrap replicates were performed to estimate support values for each node in the tree. Phylogenetic trees of HYC3O and HYC3R enzyme families (Figs. 2b,c) were constructed in the same manner as the large gene family trees. Inferred amino acid sequences for the latter are given in Supplementary data along with their GenBank accession numbers.

Plant genome structural analysis

To begin genome structural analyses, we reannotated the available genome of *R. tetraphylla* using GeMoMa, as described above. We checked the annotated assembly using SynMap in CoGe to ascertain its whole-genome duplication (WGD; polyploidy) status, to confirm no additional WGDs following the *gamma* hexaploidy event [78] at the base of all core eudicots (as

was reported in the *R. tetraphylla* genome paper [15]). While the *R. tetraphylla* genome publication reported that haplotypic contigs were purged and that no WGDs were observed, our own self:self syntenic analysis using SynMap in CoGe [79] showed the presence of internal syntenic blocks that could be ascribed to a recent WGD based on overall low Ks (synonymous substitution rate) values for homologous gene pairs. Surprisingly for a putative WGD, however, the numbers of such low-Ks pairs in *R. tetraphylla* closely matched the number of high-Ks pairs retained since the ancient *gamma* hexaploidy event, and *G. elegans* and *V. vinifera* self:self SynMaps and a *R. tetraphylla*:*C. roseus* plot also revealed old *gamma* peaks comprising similarly low numbers of gene pairs (Supplementary Fig. 76,77). Heavy fractionation (alternative homolog deletion on different subgenomes) of *gamma* over time was anticipated, but not similar fractionation for an extremely recent WGD, as would otherwise be suggested by the small number of low-Ks syntenic homolog pairs. Similar suspicions came from syntenic analyses executed in MCScan in the JCVI application [48] ; a self:self syntenic dotplot for *R. tetraphylla* showed the same internally duplicated blocks as SynMap did (Supplementary Fig. 78), but syntenic depth histograms generated by that software did not yield a substantial enough number of doubled blocks to suggest that a WGD accounted for their presence (Supplementary Figure 79). We further checked these results using the software Ksrates [80] (version 1.1.359), which corrects for unequal Ks rates in different taxa using a phylogenetic tree; it also provides relative timings via Ks values for species splits and any WGD events inferred. Coding sequence fasta files were extracted using AGAT version 1.0.065. See Supplementary figure 76. Based on this analysis, and the previous SynMap results, we conclude that the extremely recent "syntenic" gene pairs are instead regions of alternative haplotypes in the assembly that remained unpurged (and therefore partially diploid) despite use of the purge haplotigs software by the previous investigators. We confirmed this conclusion by examining two scaffolds, both containing a reserpine cluster; these were nearly identical in their overlapping regions (results not shown). We therefore focused our further structural analyses on the longer of the two *R. tetraphylla* scaffolds, which contained both the reserpine and GS BGCs.

The color highlighting of genes on the McScan visualization was achieved by identifying high-scoring pairs (HSPs) through synteny-based searches, focusing on conserved genomic regions between *R. tetraphylla* and related species. These HSPs were confirmed by locating their proximity to the refence genes within our large-scale phylogenetic gene trees to validate their evolutionary relationships. Syntenic analysis across the Gentianales order and *V. vinifera* provided a comparative view of evolutionarily conserved blocks. *R. tetraphylla* genes associated encoding homologs of both YOS and CADs were marked in dark yellow to indicate their HSP counterparts, while confirmed biosynthetic genes were connected with color-coded lines.

For the MCScan analyses noted above (syntenic dot plots, syntenic depth) as well as syntentic karyotype plots between assemblies, input annotation files were converted to gff3 format using AGAT version 1.0.065. AGAT was also used to extract CDS fasta files using each species' assembly and reformatted annotation. When detecting synteny between two species with the same ploidy level, a C-score cutoff of 0.99 (–cscore = 0.99) was used to filter out high-Ks pairs (i.e., greater than the *gamma* hexaploidy's expected Ks) for clearer connections between syntenic blocks. Otherwise, default options were used to generate figures.

Acknowledgements

This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant, a New Brunswick Innovation Foundation (NBIF) Research Assistantship Initiative Grant, and a NBIF Research Professional Initiative grant to Y.Q. V.A.A. acknowledges support from U.S. National Science Foundation grant 2030871. The authors thank the Chemical Computing Group ULC (www.chemcomp.com) for MOE licenses. This research was enabled in part by support provided by ACENET (www.ace-net.ca) and the Digital Research Alliance of Canada (alliancecan.ca).

Conflict of interest

The authors declare that there are no conflicts of interest.

Data availability

The data supporting the conclusion of this study are included in the article, supplementary information, and supplementary data file.

Table 1. ^{1}H NMR chemical shifts of alkaloids in this study.

	Tetrahydro-alstonine	3,14-dehydro-tetrahydro-alstonine	3-epi-ajmalicine		Akuammigine (3-epi-tetrahydroalstonine)			Pseudoyohimbine (3-epi-yohimbine)		3-epi-rauwolscine	Rauwolscine	alloyohimbine	
	This study	This study	This study	Ref [38]	This study 45°C	This study 45°C	Ref [26]	This study	Ref [36]	This study	This study	This study	This study
	$CDCl_3$	$CDCl_3$	$CDCl_3$	$CDCl_3$	acetone-*d*6	$CDCl_3$	$CDCl_3$	$CDCl_3$	$CDCl_3$	$CDCl_3$	$CDCl_3$	acetone-*d*6	$CDCl_3$
NH	7.77 br s	7.96 br s	8.17 br s	-	9.74 br s	7.94 br s	7.84 br s	7.99 br s	8.11	7.62 br s	7.69 br s	9.79 br s	7.78 s
3	3.36 ddd	-	4.55 qt	4.55 m	3.56 br m	3.60 br m	3.8	4.47 qt	4.46	4.43 br s	3.15 d	3.20 d	3.25 br s
5	2.56 m 2.96 ddd	3.09 dd, 3.07 ddd	3.32 dd	3.30 m	2.75 m 3.08 dd	2.91 br s 3.12 dd	2.8-3.2 m	2.53 dd 3.29 m	2.54 3.28	3.21 d 3.23 dd	2.52 dd 2.96 dd	2.47 dt 2.95 dd	2.54 m 2.97 dd
6	2.70 m 2.94 ddd	2.87 ddd 2.95 dd	2.55 dd 3.01 dddd	2.55 m 3.00 m	2.59 dd 2.93 m	2.65 dd 3.02 m	2.8-3.2 m	2.57 m 3.02 m	2.60 3.02	2.54 br ddd 3.01 m	2.68 m 2.95 m	2.64 dt 2.85 dddd	2.68 br d 2.95 m
9	7.45 d	7.46 d	7.49 dd	7.49 dd	7.39 d	7.46 d	7.46 dbr	7.49 br d	7.47	7.48 d	7.46 d	7.39 d	7.46 d
10	7.07 dt	7.06 dt	7.11 dt	7.10 dt	6.97 dt	7.08 dt	7.08 dd	7.11 dt	7.09	7.11 dt	7.08 dt	6.97 dt	7.08 t
11	7.13 dt	7.15 dt	7.17 dt	7.17 dt	7.03 dt	7.14 dt	7.14 dd	7.17 dt	7.15	7.17 dt	7.13 dt	7.05 dt	7.13 t
12	7.28 d	7.27 d	7.40 dd	7.40 dd	7.35 d	7.35 d	7.35 dbr	7.41 dt	7.4	7.33 d	7.31 d	7.29 d	7.31 d
14	2.49 dt 1.54 dd	4.98 dd	1.65 ddd 3.19 dt	1.65 ddd 3.19 d	1.98 br s 2.74 br m	2.11 br s 2.64 m	2.14 m 2.8-3.2 m	1.65 m 2.18 m	1.66 2.23	1.56 m 2.13 m	1.60 m 1.71 q	1.83 q 2.09 dt	1.89 m 1.89 m
15	2.77 m	3.50 m	1.99 dddd	2.00 m	2.75 br m	2.72 br s	2.75 m	1.62 m	1.61	2.17 br m	2.44 ddd	2.27 ddd	2.55 m
16	-	-	-	-	-	-	-	2.29 dd	2.28	2.42 dd	2.55 m	2.79 m	2.84 br s
17	7.56 s	7.60 s	7.50 d	7.50 d	7.47 s	7.55 d	7.54 s	4.22 qt	4.19	4.09 ddd	4.00 dt	3.88 quintet	3.84 br s
18	1.41 d	1.45 d	0.92 d	0.92 d	1.32 d	1.35 d	1.37 d	1.53 ddd 1.87 ddd	1.54 1.84	1.34 m 2.12 dt	1.37 m 2.05 m	1.73 m 2.00 m	1.81 m 1.92 m
19	4.50 dq	4.18 dq	4.34 dq	4.33 dq	4.43 quintet	4.44 quintet	4.44 qd	1.28 m 1.29 m	1.25 1.30	1.69 m 2.17 br m	1.56 m 2.11 ddd	1.52 m 2.02 m	1.56 m 2.09 br d
20	1.70 m	1.92 m	2.02 ddd	2.00 m	1.93 br s	1.85 br s	1.86 m	1.41 m	1.41	1.66 m	1.82 m	2.13 m	1.85 m
21	2.74 m 3.11 dd	3.22 d	2.56 t 2.63 dt	2.50t, 2.63 t	2.53 ddd 2.93 m	2.59 dd 2.97 dd	2.5-2.7 m 2.8-3.2 m	2.56 m 3.29 m	2.55 3.26	2.50 dd 3.01 dd	2.60 dd 2.85 dd	2.53 dd 2.82 m	2.56 m 2.81 br d
OCH_3	3.75 s	3.78 s	3.74 s	3.73 s	3.69 s	3.75 s	3.75 s	3.58 s	3.79	3.86 s	3.84 s	3.65 s	3.77 s

Table 2. ^{13}C NMR chemical shifts of alkaloids in this study.

	Tetrahydro-alstonine	3,14-dehydro-terahydroalstonine	3-epi-ajmalicine		Akuammigine (3-epi-tetrahydroalstonine)			Pseudoyohimbine (3-epi-yohimbine)		3-epi-rauwolscine		Rauwolscine	Alloyohimbine		
	This study	This study	This study	Ref [38]	This study 45°C	This study 45°C	Ref [37] >R.T.	This study	Ref [36]	This study	Ref [35]	This study	This study	This study	Ref [37]
	$CDCl_3$	$CDCl_3$	$CDCl_3$	$CDCl_3$	acetone -*d*6	$CDCl_3$	$CDCl_3$	$CDCl_3$	$CDCl_3$	$CDCl_3$	$CDCl_3$	$CDCl_3$	acetone -*d*6	$CDCl_3$	$CDCl_3$
2	134.68	129.88	132.76	132.50	135.41	133.63	132.8	133.15	123.3	132.23	132	134.61	-	134.65	134.4
3	59.98	134.69	54.26	54.10	55.98	54.81	54.5	54.29	54.3	54.04	53.8	60.32	60.84	60.54	60.1
5	53.72	51.41	51.07	50.90	53.55	52.54	52.2	51.19	51	51.22	51	53.34	53.19	53.29	52.8
6	21.92	21.53	16.98	16.80	20.84	19.49	19.2	17.1	16.9	16.88	16.6	21.75	21.69	21.8	21.3
7	108.27	110.71	106.89	106.70	108.27	108.22	106.8	108.03	107.2	108.48	108	108.61	-	108.65	107.1
8	127.32	127.14	127.88	127.70	128.6	127.88	127.2	127.96	127.5	127.92	127.6	127.46	-	127.5	126.8
9	118.21	118.6	118.08	117.90	118.34	118.11	117.7	118.01	117.9	118.22	118	118.25	117.6	118.29	117.5
10	119.56	119.67	119.5	119.30	119.57	119.61	119.1	119.52	119.2	119.69	119.6	119.57	118.62	119.65	118.6
11	121.56	122.79	121.64	121.40	121.54	121.62	121.2	121.54	121.4	121.68	121.7	121.59	120.71	121.6	120.5
12	110.93	110.94	111.42	111.20	112.03	111.2	110.8	111.53	111.4	111	110.9	110.84	110.87	110.92	110.6
13	136.14	136.89	135.99	135.80	137.62	136.24	135.7	136.06	136.1	135.81	135.7	136.11	-	136.17	135.8
14	34.41	96.39	31.37	31.20	31.85	31.1	30.6	32.6	32.1	23.88	23.6	27.84	31.41	31.38	31
15	31.52	30.29	26.27	26.00	25.95	26.2	25.7	31.62	31.3	32.67	32.2	38.04	38.25	38.17	37.4
16	109.62	108.97	107.84	107.60	108.27	107.16	107.6	52.2	52.1	54.14	54	54.83	51.54	50.92	50.6
17	155.87	155.51	154.76	154.60	154.95	155.31	154.8	67.47	67.2	66.09	65.8	66.13	66.38	67.4	66.7
18	18.67	18.44	15.2	15.00	18.75	18.74	18.4	31.77	31.6	33.45	33.2	33.2	30.78	31.79	30.2
19	72.62	72.99	74.02	73.80	74.7	73.59	73.2	23.16	23.1	24.41	24.1	24.63	25.11	25.35	24.8
20	38.60	36.7	41.54	41.30	38.41	37.8	37.2	40.39	39.9	36.06	35.6	36.69	32.32	32.85	32
21	56.48	51.7	47.56	47.30	N.D.	N.D.	50.3	51.45	51	49.86	49.5	60.64	60.5	60.35	59.6
CH_3-O-<u>C</u>O	168.12	168.36	167.54	167.30	167.91	167.95	167.5	175.08	175.4	174.91	173.8	174.83	-	174.56	174
<u>C</u>H_3-O-CO	51.26	51.33	51.19	50.90	51.13	51.18	50.9	52.04	51.9	52.08	52	52.04	50.57	52.05	51.5

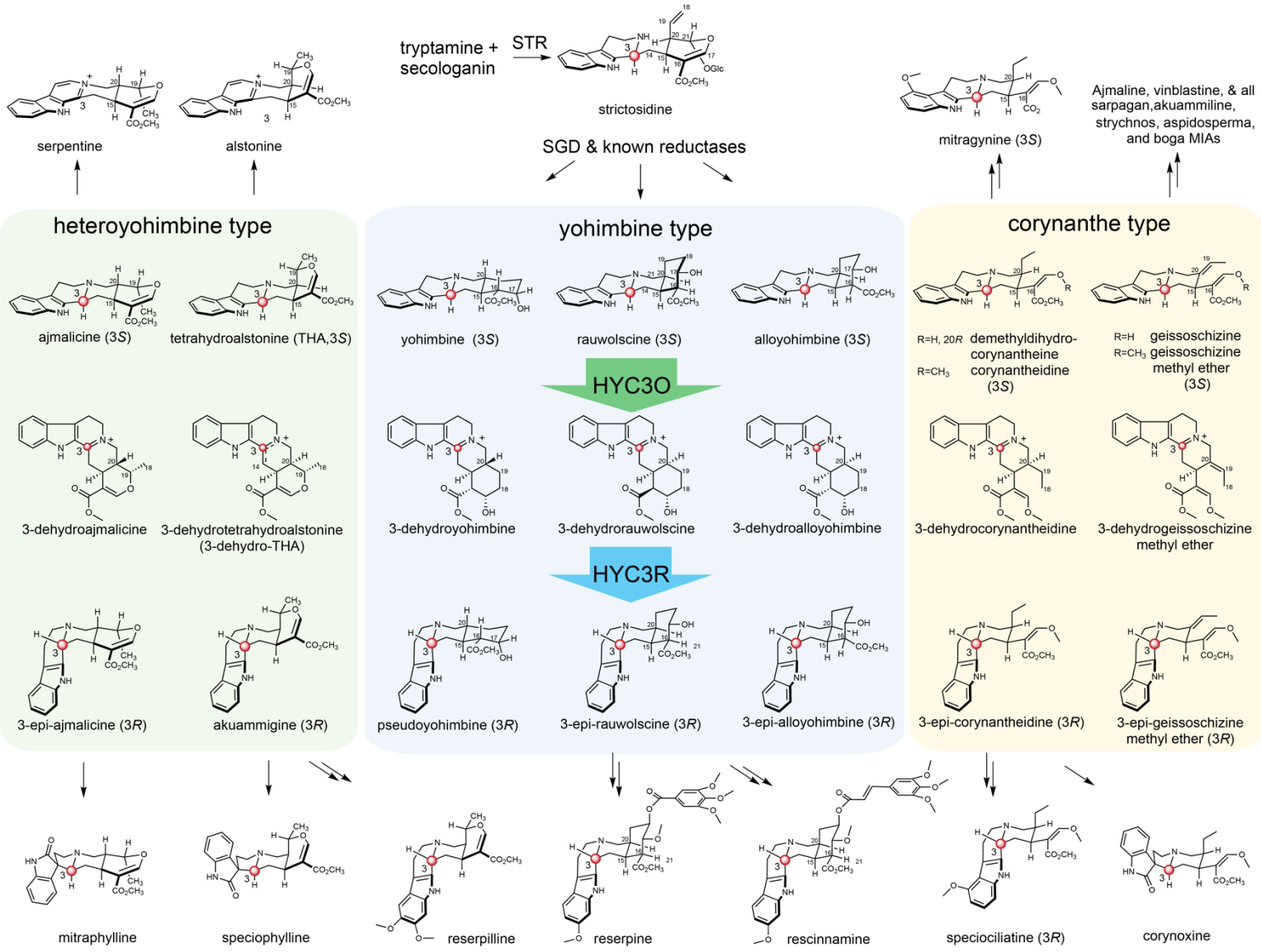

Figure 1. HYC3O and HYC3Rs are responsible for reserpine biosynthesis and the C3 stereochemistry inversion of diverse heteroyohimbine, yohimbine, and corynanthe types of MIAs. The 3*S* MIAs inherit their stereochemistry from the precursor strictosidine. The C3 stereochemistry inversion involves H3-hydride abstraction by HYC3Os, followed by *si*-face C3-reduction by HYC3Rs. The indoles of 3*R* MIAs are uniquely perpendicular to the rest molecule. Oxindoles with a 3*S* stereochemistry are derived from 3*R* precursors. STR: strictosidine synthase; SGD: strictosidine β-glucosidase.

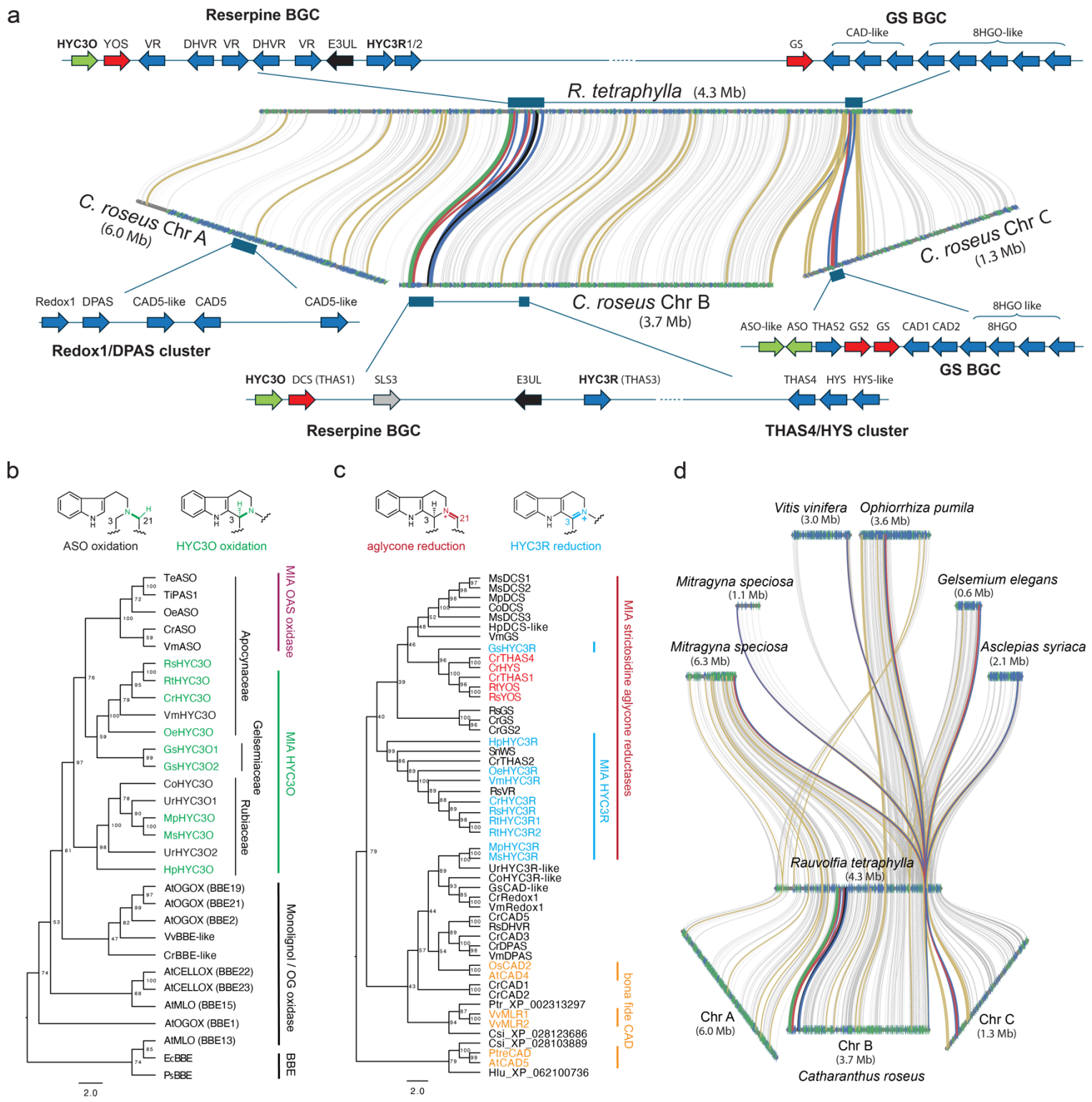

Figure 2. The reserpine (*HYC3O*/*HYC3R*) and GS clusters in *R. tetraphylla* and *C. roseus* trace their ancestry to the Gentianales common ancestor with grapevine (*V. vinifera*). (a) Synteny on a genomic scale between *R. tetraphylla* and *C. roseus* reveals a ca. 200 kpb reserpine

BGC, including genes encoding strictosidine aglycone reductases (*YOS* and DCS/THAS1) and the redox duo HYC3O/HYC3R for MIA C3 stereochemistry inversion. *R. tetraphylla* reserpine BGC contains genes coding for 8 CAD-like reductases, including VR and DHVR, which participate in ajmaline biosynthesis. The GS cluster is conserved between the two species, while *C. roseus* additionally contains BGCs encoding Redox1/DPAS and THAS4/HYS. (**b**) Phylogenetic analysis reveals a monophyletic origin for the MIA-oxidizing enzymes HYC3O and ASO within the Gentianales order. These enzymes evolved from broader families of BBE-like flavin-containing oxidases, which include monolignol oxidases and oligosaccharide oxidases from *Arabidopsis thaliana*. (**c**) Phylogenetic analysis demonstrates that MIA-reducing CAD-like reductases share a common ancestry with bona fide monolignol aldehyde reductases and other CAD-like enzymes across diverse plant species. Co: *Cephalanthus occidentalis*; Cr: *Catharanthus roseus*; Csi: *Camellia sinensis*; Gs: *Gelsemium sempervirens*; Hp: *Hamelia patens*; Hlu: *Humulus lupulus*; Ms: *Mitragyna speciosa*; Mt: *Mitragyna parvifolia*; Oe: *Ochrosia elliptica*; Rs: *Rauvolfia serpentina*; Rt: *Rauvolfia tetraphylla*; Rst: *Rhazya stricta*; Snv: *Strychnos nux-vomica*; Te: *Tabernaemontana elegans*; Ti: *Tabernanthe iboga*; Ur: *Uncaria rhynchophylla*; Vm: *Vinca minor*; At: *Arabidopsis thaliana*; Csi: *Camellia sinensis*; Ec: *Eschscholzia californica*; Os: *Oryza sativa*; Ps: *Papaver somniferum*; Pt: *Populus tremuloides*; Ptr: *Populus trichocarpa*; and Vv: *Vitis vinifera*. Sequences are provided in Supplementary data. (**d**) Synteny on a genomic scale between *R. tetraphylla* and other members of the Gentianales including *M. speciosa*, *O. pumila*, *G. elegans*, *A. syriaca*, *C. roseus*, and *V. vinifera* as outgroup. *R. tetraphylla* genes found in our large-scale gene trees (mostly encoding CADs) are highlighted with dark yellow syntenic lines connecting their corresponding High-Scoring Pair (HSP) gene counterparts in other species. In panels (a) and (d), colorations of arrows and syntenic lines correspond with *HYC30*-like (green), *HYC3R*-like (blue) and *THAS/GS*/*YOS* (red) genes.

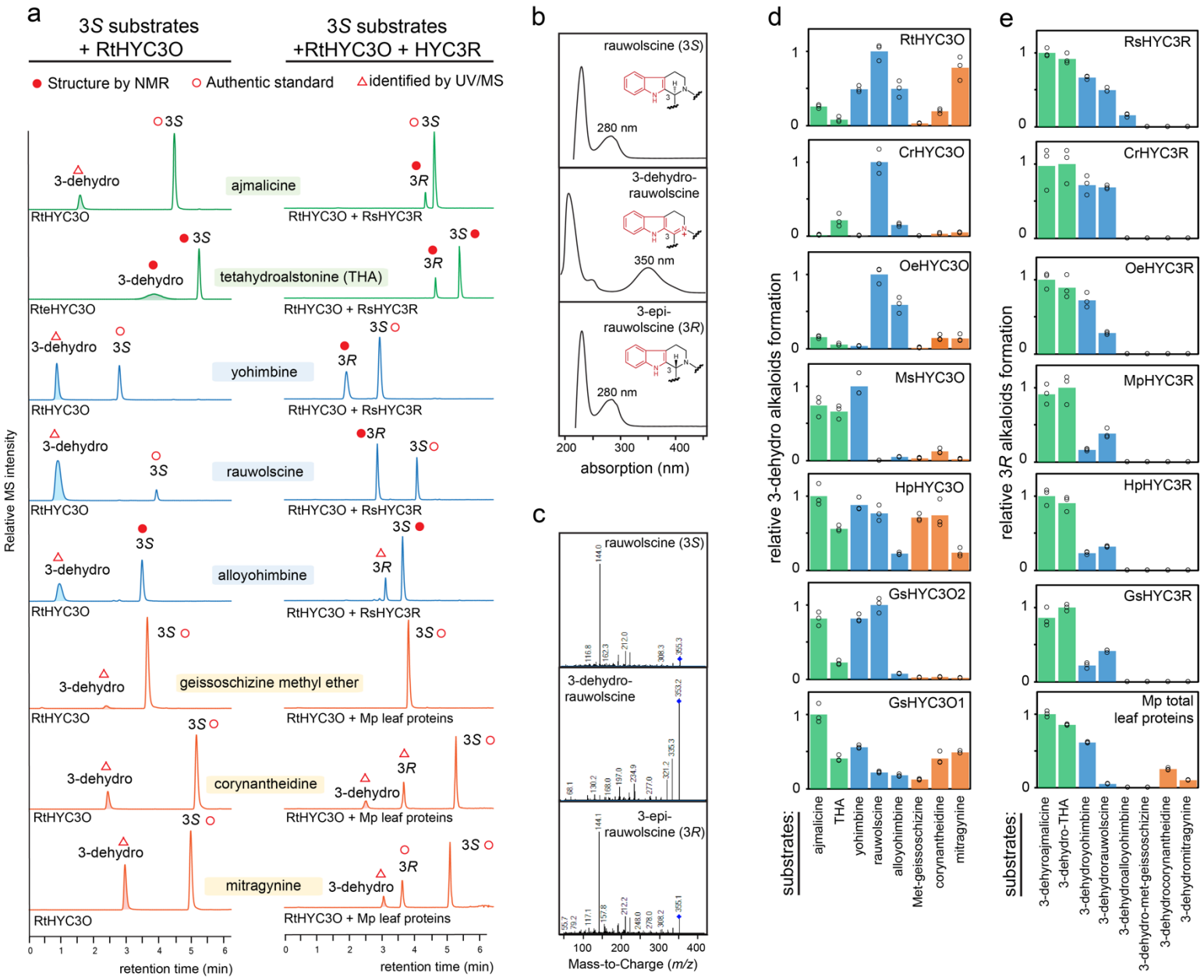

Figure 3. HYC3Os and HYC3Rs from three plant families catalyze C3 stereochemistry inversion across structurally diverse heteroyohimbine (green), yohimbine (blue), and corynanthe (orange) types of MIA substrates. (**a**) LC-MS/MS $[M+H]^+$ Multiple Reaction Monitoring (MRM) chromatograms show the *in vitro* reaction products generated by RtHYC3O alone, and RtHYC3O in combination with either RsHYC3R or *M. parvifolia* total leaf proteins, using various substrates. Ajmalicine and THA: MRM 353>144; Yohimbine, rauwolscine, and alloyohimbine: MRM 355/144; geissoschizine methyl ether: MRM 367>144; corynantheidine: MRM 369>144; mitragynine: MRM 399>174; 3-dehydroajmalicine and 3-dehydro-THA: MRM 351>265; 3-dehydro-yohimbine/rauwolscine/alloyohimbine: MRM 353>335; 3-dehydrogeissoschizine methyl ether: MRM 365>249; 3-dehydrocorynantheidine: MRM 367>251; 3-dehydromitragynine: MRM 399>227. (**b**) The UV absorption maximum for 3-

dehydro-rauwolscine shifts to 350 nm due to extended conjugation resulting from C3 dehydrogenation. (**c**) C3 dehydrogenation in rauwolscine leads to significant change in its MS/MS fragmentation pattern, compared to those of 3*S*/3*R*-rauwolscine epimers. (**d**) Relative enzyme activities of various HYC3Os with different substrates. (**e**) Relative enzyme activities of various HYC3Rs with different substrates. For (**d**) and (**e**), the activity for the substrate with the highest conversion rate was set as 1, with other rates normalized to this value. Data were generated from three technical replicates, and the bar graphs display mean values, with individual data points shown as circles for each mean (Supplementary data). Product ion scans used for MRM and additional UV absorption profiles are provided in Supplementary Fig. 1

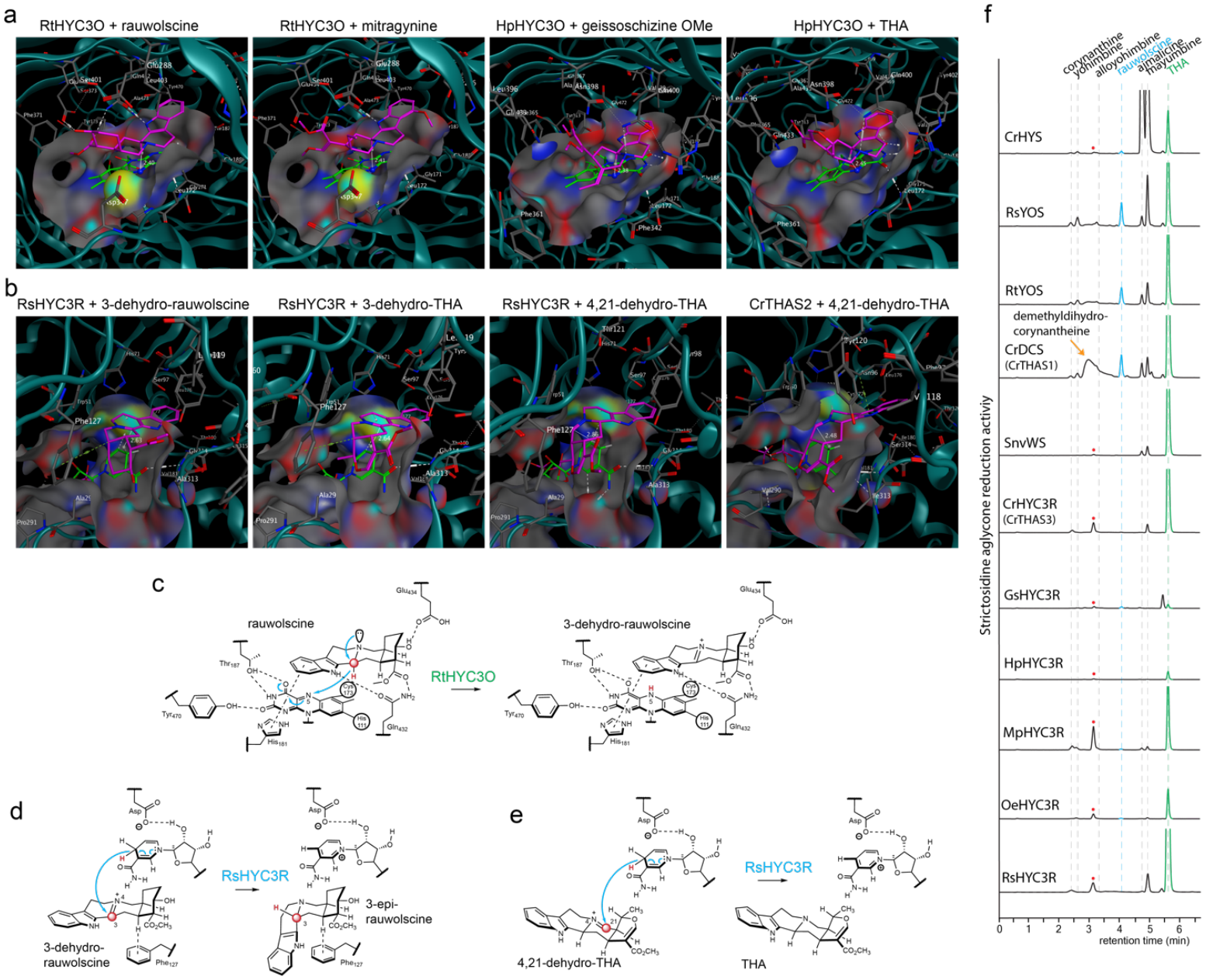

Figure 4. Homology modeling and substrate docking experiments reveal the basis for substrate promiscuity of HYC3Os and dual catalytic activity of HYC3Rs for both 3-dehydro MIA and strictosidine aglycone substrates. (**a**) Docking studies of rauwolscine, mitragynine, geissoschizine methyl ether, and tetrahydroalstonine (THA) at the active sites of RtHYC3O and HpHYC3O homology models. Hydrogen bond networks and π-stacking interactions orient the substrates for efficient H3 hydride abstraction by FAD. Variations in active site architecture correlate with the observed substrate preferences (**b**) Docking studies of 3-dehydrorauwolscine, 3-dehydro-THA, and 4,21-dehydro-THA at the active sites of the RsHYC3R homology model and CrTHAS2 crystal structure demonstrate key differences in substrate accommodation. The spacious active site in RsHYC3R supports the binding and

reduction of both 3-dehydro-THA and 4,21-dehydro-THA (a form of strictosidine aglycone mixture), accounting for its dual catalytic activity. In contrast, the narrower active site of CrTHAS2 selectively facilitates the reduction of 4,21-dehydro-THA but not 3-dehydro-THA. FAD is shown in green, and alkaloid substrates are in magenta. The surface of active site is illustrated: oxygen-red, nitrogen-blue, sulfur-yellow, carbon-grey. White dotted lines show the hydrogen bonds. The distances between FAD N5 and alkaloid substrates H3 or H21 are shown. (**c**) Illustration of rauwolscine binding and oxidation by FAD at the RtHYC3O active site. (**d**) Illustration of 3-dehydrorauwolscine binding and reduction by NADPH at the RsHYC3R active site. (**e**) Illustration of 3-dehydro-THA binding and reduction by NADPH at the RsHYC3R active site. (**f**) LC-MS/MS MRM $[M+H]^+$ (355>144 and 353>144) chromatograms show the *in vivo* reduction of strictosidine aglycone by strictosidine aglycone reductases and HYC3Rs. These reductases were expressed in yeast strain AJM-dHYS engineered for *de novo* production of strictosidine aglycone. The reductases exhibited diverse product spectra, reflecting the structural diversity of strictosidine aglycone in equilibrium. An unknown *m/z* 353 peak is labeled with a red dot.

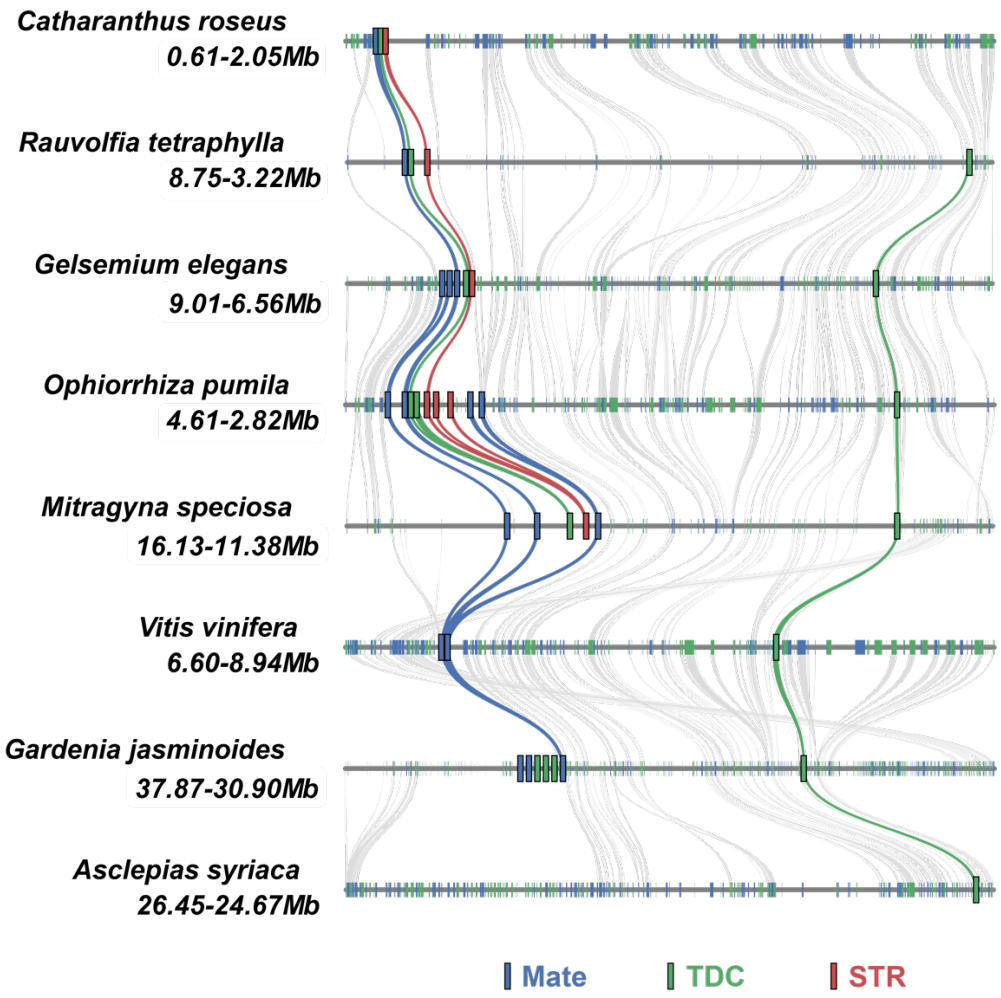

Figure 5. Genomic-scale synteny around the STR BGC for members of the order Gentianales, with grapevine (*Vitis vinifera*) as an outgroup. A syntentic view reveals conservation of the genomic block surrounding the *TDC*/*STR*/*MATE* gene cluster responsible for biosynthesis of the MIA precursor strictosidine. All MIA-producing species in Rubiaceae (*O. pumila* and *M. speciosa*), Gelsemiaceae (*G. elegans*), and Apocynaceae (*C. roseus* and *Rauvolfia tetraphylla*) have complete STR BGCs, with parsimony suggesting the BGC assembled in the Gentianales stem lineage. Evolutionary inferences include: (1) a *MATE* was already in position in the Gentianales/*V. vinifera* common ancestor, (2) a *TDC* also existed downstream of the *MATE*, (3) this (or another) *TDC* duplicated in the Gentianales stem lineage and translocated proximal to the preexisting *MATE*, completing the BGC, (4) a *STR* translocated to the region containing the *MATE* and *TDC* duplicate, completing the BGC, the genes of which subsequently (5) underwent alternative tandem duplications, most notably within Rubiaceae species, (6) the *STR* was deleted in the non-MIA-producing Rubiaceae species *G. jasminoides*, (7) the entire STR BGC was deleted in the non-MIA Apocynaceae species *A. syriaca*, and (8) the downstream *TDC* was deleted in MIA-producing *C. roseus*. Syntenic lines are colored green for *TDC* genes, red for *STR*s, and blue for *MATE*s; grey represents other homologous genes. Where colored lines are multiple within the STR BGC, tandem duplications of *TDC*/*STR*/*MATE* genes can be inferred. Only one of the two subgenomes of tetraploid *M. speciosa* is shown. Numbers below species names indicate lengths of chromosomal regions in Mb.

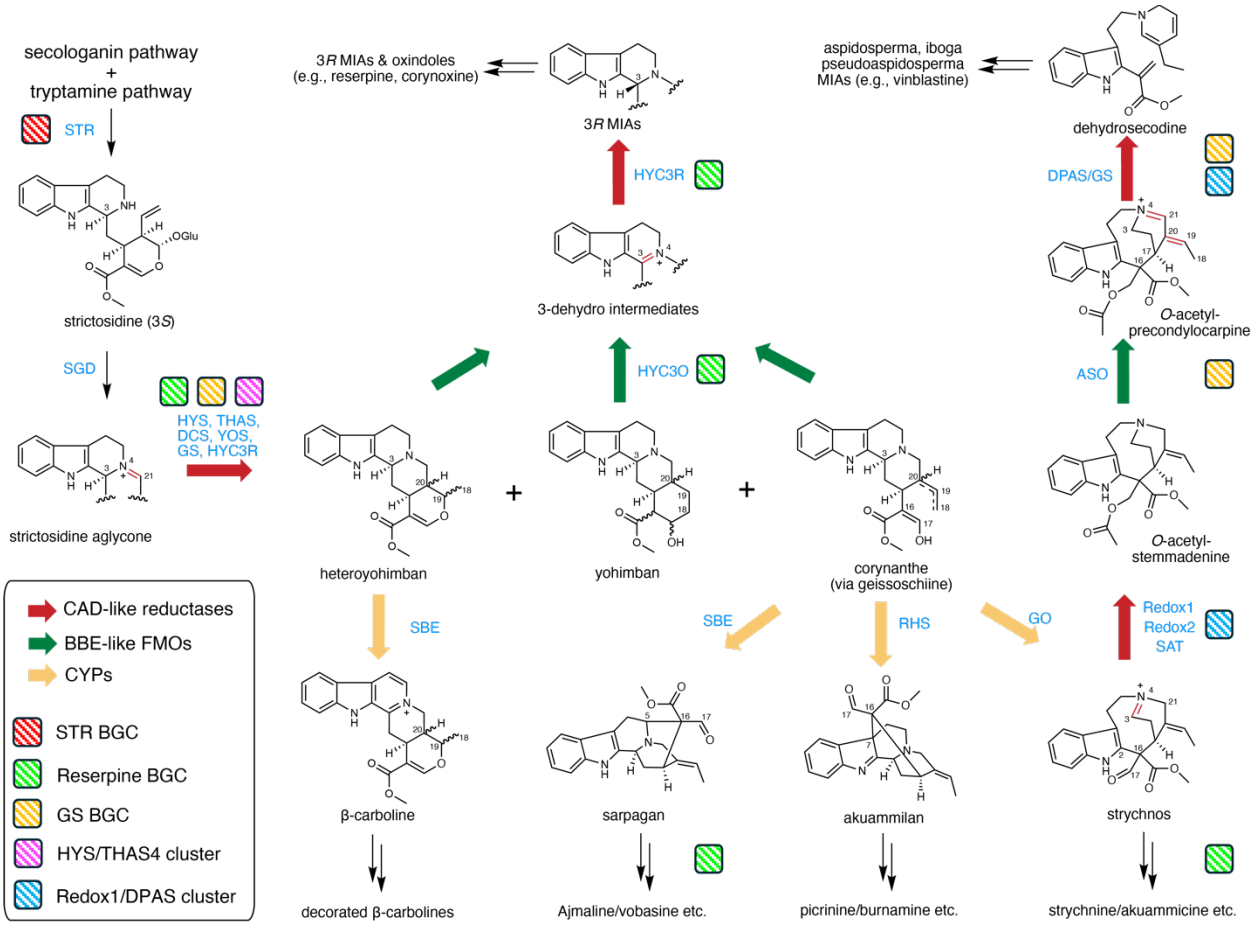

Figure 6. MIA skeleton diversification is primarily driven by the activities of physically clustered CAD-like reductases, BBE-like oxidases, and non-clustered CYPs. SGD: strictosidine β-glucosidase; GS: geissoschizine synthase; THAS: tetrahydroalstonine synthase; HYS: heteroyohimbine synthase; YOS: yohimban synthase; DCS: demethyldihydrocorynantheine/demethylcorynantheidine synthase; SBE: sarpagan bridge enzyme; RHS: rhazimal synthase; GO: geissoschizine oxidase; SAT: stemmadinine *O*-acetyltransferase; ASO: *O*-acetylstemmadenine oxidase; DPAS: dihydroprecondylocarpine synthase; HYC3O: heteroyohimbine/yohimbine/corynanthe 3-oxidase; HYC3R: heteroyohimbine/yohimbine/corynanthe 3-reductase.

Reference

1. Reed, J. *et al.* Elucidation of the pathway for biosynthesis of saponin adjuvants from the soapbark tree. *Science* **379**, 1252–1264 (2023).

2. Li, Q. *et al.* Gene clustering and copy number variation in alkaloid metabolic pathways of opium poppy. *Nat. Commun.* **11**, 1190 (2020).

3. Winzer, T. *et al.* A Papaver somniferum 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science (New York, NY)* **336**, 1704–1708 (2012).

4. Kellner, F. *et al.* Genome-guided investigation of plant natural product biosynthesis. *The Plant journal : for cell and molecular biology* **82**, 680–692 (2015).

5. Qu, Y., Safonova, O. & Luca, V. D. Completion of the canonical pathway for assembly of anticancer drugs vincristine/vinblastine in Catharanthus roseus. *The Plant journal : for cell and molecular biology* **97**, 257–266 (2018).

6. Qu, Y. *et al.* Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proceedings of the National Academy of Sciences* **112**, 6224–6229 (2015).

7. Qu, Y. *et al.* Solution of the multistep pathway for assembly of corynanthean, strychnos, iboga, and aspidosperma monoterpenoid indole alkaloids from 19E-geissoschizine. *Proc. Natl. Acad. Sci.* **115**, 3180–3185 (2018).

8. Li, C. *et al.* Single-cell multi-omics in the medicinal plant Catharanthus roseus. *Nat. Chem. Biol.* **19**, 1031–1041 (2023).

9. Caputi, L. *et al.* Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science (New York, NY)* **360**, 1235–1239 (2018).

10. Li, F. *et al.* Characterization of a vacuolar importer of secologanin in Catharanthus roseus. *Commun. Biol.* **7**, 939 (2024).

11. Payne, R. M. E. *et al.* An NPF transporter exports a central monoterpene indole alkaloid intermediate from the vacuole. *Nature Plants* **3**, 1–9 (2017).

12. Franke, J. *et al.* Gene Discovery in Gelsemium Highlights Conserved Gene Clusters in Monoterpene Indole Alkaloid Biosynthesis. *ChemBioChem* **20**, 83–87 (2019).

13. Rai, A. *et al.* Chromosome-level genome assembly of Ophiorrhiza pumila reveals the evolution of camptothecin biosynthesis. *Nature communications* **12**, 1–19 (2021).

14. Zuntini, A. R. *et al.* Phylogenomics and the rise of the angiosperms. *Nature* **629**, 843–850 (2024).

15. Stander, E. A. *et al.* The Rauvolfia tetraphylla genome suggests multiple distinct biosynthetic routes for yohimbane monoterpene indole alkaloids. *Commun. Biol.* **6**, 1197 (2023).

16. Lezin, E. *et al.* A chromosome-scale genome assembly of Rauvolfia tetraphylla facilitates identification of the complete ajmaline biosynthetic pathway. *Plant Commun.* **5**, 100784 (2024).

17. Edge, A. *et al.* A tabersonine 3-reductase Catharanthus roseus mutant accumulates vindoline pathway intermediates. *Planta* **25**, 1–15 (2017).

18. Qu, Y. *et al.* Geissoschizine synthase controls flux in the formation of monoterpenoid indole alkaloids in a Catharanthus roseus mutant. *Planta* **25**, 1–10 (2017).

19. Guo, J., Gao, D., Lian, J. & Qu, Y. De novo biosynthesis of antiarrhythmic alkaloid ajmaline. *Nat. Commun.* **15**, 457 (2024).

20. Hong, B. *et al.* Biosynthesis of strychnine. *Nature* **607**, 617–622 (2022).

21. Stavrinides, A. *et al.* Structural investigation of heteroyohimbine alkaloid synthesis reveals active site elements that control stereoselectivity. *Nature communications* **7**, 12116 (2016).

22. Kim, K. *et al.* Biosynthesis of kratom opioids. *N. Phytol.* **240**, 757–769 (2023).

23. Lorensen, M. D. B. B. *et al.* Spatial localization of monoterpenoid indole alkaloids in Rauvolfia tetraphylla by high resolution mass spectrometry imaging. *Phytochemistry* **209**, 113620 (2023).

24. Ma, X., Panjikar, S., Koepke, J., Loris, E. & Stöckigt, J. The structure of Rauvolfia serpentina strictosidine synthase is a novel six-bladed beta-propeller fold in plant proteins. *The Plant cell* **18**, 907–920 (2006).

25. Treimer, J. F. & Zenk, M. H. Purification and properties of strictosidine synthase, the key enzyme in indole alkaloid formation. *European journal of biochemistry / FEBS* **101**, 225–233 (1979).

26. Kohl, W., Witte, B. & Höfle, G. Alkaloide aus Catharanthus roseus-Zellkulturen, II [1] / Alkaloids from Catharanthus roseus Tissue Cultures, II [1]. *Zeitschrift Für Naturforschung B* **36**, 1153–1162 (1981).

27. Duwiejua, M., Woode, E. & Obiri, D. D. Pseudo-akuammigine, an alkaloid from Picralima nitida seeds, has anti-inflammatory and analgesic actions in rats. *J. Ethnopharmacol.* **81**, 73–79 (2002).

28. Bruyn, A. D., Zhang, W. & Buděšinský, M. NMR study of three heteroyohimbine derivatives from Rauwolfia serpentina: Stereochemical aspects of the two isomers of reserpiline hydrochloride. *Magn. Reson. Chem.* **27**, 935–940 (1989).

29. Kouamo, K., Creche, J., Chénieux, J. C., Rideau, M. & Viel, C. Alkaloid Production by Ochrosia elliptica Cell Suspension Cultures. *J. plant Physiol.* **118**, 277–83 (1984).

30. Takayama, H. Chemistry and Pharmacology of Analgesic Indole Alkaloids from the Rubiaceous Plant, Mitragyna speciosa. *Chem Pharm Bulletin* **52**, 916–928 (2004).

31. Pandey, R., Singh, S. C. & Gupta, M. M. Heteroyohimbinoid type oxindole alkaloids from Mitragyna parvifolia. *Phytochemistry* **67**, 2164–2169 (2006).

32. Paniagua-Vega, D., Cerda-García-Rojas, C. M., Ponce-Noyola, T. & Ramos-Valdivia, A. C. A New Monoterpenoid Oxindole Alkaloid from Hamelia Patens Micropropagated Plantlets. *Nat Prod Commun* **7**, 1934578X1200701109 (2012).

33. Nguyen, T.-A. M. *et al.* Discovery of a cytochrome P450 enzyme catalyzing the formation of spirooxindole alkaloid scaffold. *Front Plant Sci* **14**, 1125158 (2023).

34. Winkler, A. *et al.* A concerted mechanism for berberine bridge enzyme. *Nature chemical biology* **4**, 739–741 (2008).

35. Falkenhagen, H., Stckigt, J., Kuzovkina, I. N., Alterman, I. E. & Kolshorn, H. Indole alkaloids from “hairy roots” of Rauwolfia serpentina. *Can. J. Chem.* **71**, 2201–2203 (1993).

36. Stahl, R. & Borschberg, H. A Reinvestigation of the Oxidative Rearrangement of Yohimbane-Type Alkaloids. Part A. Formation of pseudoindoxyl (= 1,2-dihydro-3H-indol-3-one) derivatives. *Helvetica Chim. Acta* **77**, 1331–1345 (1994).

37. Wenkert, E. *et al.* General methods of synthesis of indole alkaloids. 14. Short routes of construction of yohimboid and ajmalicinoid alkaloid systems and their carbon-13 nuclear magnetic resonance spectral analysis. *J Am Chem Soc* **98**, 3645–3655 (1976).

38. Carbonezi, C. A. *et al.* Determinação por RMN das configurações relativas e conformações de alcalóides oxindólicos isolados de Uncaria guianensis. *Química Nova* **27**, 878–881 (2004).

39. Sun, S. *et al.* Single-cell RNA sequencing provides a high-resolution roadmap for understanding the multicellular compartmentation of specialized metabolism. *Nat. Plants* **9**, 179–190 (2023).

40. Brose, J. *et al.* The Mitragyna speciosa (Kratom) Genome: a resource for data-mining potent pharmaceuticals that impact human health. *G3: GenesGenomesGenet.* **11**, jkab058 (2021).

41. Wu, Y., Liu, C., Koganitsky, A., Gong, F. L. & Li, S. Discovering Dynamic Plant Enzyme Complexes in Yeast for Kratom Alkaloid Pathway Identification**. *Angew. Chem. Int. Ed.* **62**, e202307995 (2023).

42. Winkler, A., Hartner, F., Kutchan, T. M., Glieder, A. & Macheroux, P. Biochemical Evidence That Berberine Bridge Enzyme Belongs to a Novel Family of Flavoproteins Containing a Bi-covalently Attached FAD Cofactor*. *J. Biol. Chem.* **281**, 21276–21285 (2006).

43. Gao, D. *et al.* De Novo Biosynthesis of Vindoline and Catharanthine in Saccharomyces cerevisiae. *Biodesign Res* **2022**, (2022).

44. Liu, T. *et al.* Construction of ajmalicine and sanguinarine de novo biosynthetic pathways using stable integration sites in yeast. *Biotechnol Bioeng* (2022) doi:10.1002/bit.28040.

45. Krithika, R. *et al.* Characterization of 10-Hydroxygeraniol Dehydrogenase from Catharanthus roseus Reveals Cascaded Enzymatic Activity in Iridoid Biosynthesis. *Sci. Rep.* **5**, 8258 (2015).

46. Tang, H. *et al.* SynFind: Compiling Syntenic Regions across Any Set of Genomes on Demand. *Genome Biol. Evol.* **7**, 3286–3298 (2015).

47. Albert, V. A. & Krabbenhoft, T. J. Navigating the CoGe Online Software Suite for Polyploidy Research. *Methods Mol. Biol.* **2545**, 19–45 (2023).

48. Tang, H. *et al.* JCVI: A versatile toolkit for comparative genomics analysis. *iMeta* **3**, e211 (2024).

49. Pootakham, W. *et al.* A Chromosome-Scale Genome Assembly of Mitragyna speciosa (Kratom) and the Assessment of Its Genetic Diversity in Thailand. *Biology* **11**, 1492 (2022).

50. Keilwagen, J., Hartung, F. & Grau, J. Gene Prediction, Methods and Protocols. *Methods Mol. Biol.* **1962**, 161–177 (2019).

51. Antonelli, A. *et al.* Settling a family feud: a high-level phylogenomic framework for the Gentianales based on 353 nuclear genes and partial plastomes. *Am. J. Bot.* **108**, 1143–1165 (2021).

52. Sabir, J. S. M. *et al.* The nuclear genome of Rhazya stricta and the evolution of alkaloid diversity in a medically relevant clade of Apocynaceae. *Sci. Rep.* **6**, 33782 (2016).

53. Cuello, C. *et al.* The Madagascar palm genome provides new insights on the evolution of Apocynaceae specialized metabolism. *Heliyon* **10**, e28078 (2024).

54. Rueffer, M., Nagakura, N. & Zenk, M. H. Strictosidine, the common precursor for monoterpenoid indole alkaloids with 3 α and 3 β configuration. *Tetrahedron Lett.* **19**, 1593–1596 (1978).

55. Kutchan, T. M., Hampp, N., Lottspeich, F., Beyreuther, K. & Zenk, M. H. The cDNA clone for strictosidine synthase from Rauvolfia serpentina DNA sequence determination and expression in Escherichia coli. *Febs Lett* **237**, 40–44 (1988).

56. Guirimand, G. *et al.* Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"? *BMC Plant Biology* **10**, 182 (2010).

57. Wang, Z. *et al.* Deciphering and Reprogramming the Cyclization Regioselectivity in Bifurcation of Indole Alkaloids Biosynthesis. *Chem Sci* (2022) doi:10.1039/d2sc03612f.

58. Dang, T.-T. T. *et al.* Sarpagan bridge enzyme has substrate-controlled cyclization and aromatization modes. *Nat Chem Biol* **14**, 760–763 (2018).

59. Mann, S. G. A. *et al.* Stereochemical Insights into Sarpagan and Akuammiline Alkaloid Biosynthesis. (2024) doi:10.1101/2024.12.19.629490.

60. Toplak, M. *et al.* The single berberine bridge enzyme homolog of Physcomitrella patens is a cellobiose oxidase. *Febs J.* **285**, 1923–1943 (2018).

61. Benedetti, M. *et al.* Four Arabidopsis berberine bridge enzyme-like proteins are specific oxidases that inactivate the elicitor-active oligogalacturonides. *Plant J.* **94**, 260–273 (2018).

62. Locci, F. *et al.* An Arabidopsis berberine bridge enzyme-like protein specifically oxidizes cellulose oligomers and plays a role in immunity. *Plant J.* **98**, 540–554 (2019).

63. Costantini, S. *et al.* Berberine bridge enzyme–like oxidases of cellodextrins and mixed-linked β-glucans control seed coat formation. *Plant Physiol.* **194**, 296–313 (2023).

64. Daniel, B. *et al.* Oxidation of Monolignols by Members of the Berberine Bridge Enzyme Family Suggests a Role in Plant Cell Wall Metabolism*. *J. Biol. Chem.* **290**, 18770–18781 (2015).

65. Vries, S. de *et al.* The evolution of the phenylpropanoid pathway entailed pronounced radiations and divergences of enzyme families. *Plant J.* **107**, 975–1002 (2021).

66. Eng, J. G. M. *et al.* A Catharanthus roseus Fe(II)/α-ketoglutarate-dependent dioxygenase catalyzes a redox-neutral reaction responsible for vindolinine biosynthesis. *Nat Commun* **13**, 3335 (2022).

67. Mai, Z. *et al.* Oxidation of four monoterpenoid indole alkaloid classes by three cytochrome P450 monooxygenases from Tabernaemontana litoralis. *Plant J.* (2024) doi:10.1111/tpj.17145.

68. Kamileen, M. O. *et al.* Recycling Upstream Redox Enzymes Expands the Regioselectivity of Cycloaddition in Pseudo-Aspidosperma Alkaloid Biosynthesis. *J Am Chem Soc* (2022) doi:10.1021/jacs.2c08107.

69. Borges, J., Manresa, M. T., Ramón, J. L. M., Pascual, C. & Rumbero, A. Two new oxindole alkaloids isolated from Hamelia Patens Jacq. *Tetrahedron Lett* **20**, 3197–3200 (1979).

70. Ali, Z., Demiray, H. & Khan, I. A. Isolation, characterization, and NMR spectroscopic data of indole and oxindole alkaloids from Mitragyna speciosa. *Tetrahedron Lett.* **55**, 369–372 (2014).

71. Flores-Bocanegra, L. *et al.* The Chemistry of Kratom [Mitragyna speciosa]: Updated Characterization Data and Methods to Elucidate Indole and Oxindole Alkaloids. *J Nat Prod* **83**, 2165–2177 (2020).

72. Denoeud, F. *et al.* The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* **345**, 1181–1184 (2014).

73. Xu, Z. *et al.* Tandem gene duplications drive divergent evolution of caffeine and crocin biosynthetic pathways in plants. *BMC Biol.* **18**, 63 (2020).

74. Chakraborty, S. *et al.* Kratom Alkaloids as Probes for Opioid Receptor Function: Pharmacological Characterization of Minor Indole and Oxindole Alkaloids from Kratom. *ACS Chem. Neurosci.* **12**, 2661–2678 (2021).

75. Liu, Y. *et al.* Whole-genome sequencing and analysis of the Chinese herbal plant Gelsemium elegans. *Acta Pharm. Sin. B* **10**, 374–382 (2019).

76. Lyons, E. & Freeling, M. How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J.* **53**, 661–673 (2008).

77. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

78. Jiao, Y. *et al.* A genome triplication associated with early diversification of the core eudicots. *Genome Biol.* **13**, R3 (2012).

79. Lyons, E., Pedersen, B., Kane, J. & Freeling, M. The Value of Nonmodel Genomes and an Example Using SynMap Within CoGe to Dissect the Hexaploidy that Predates the Rosids. *Trop. Plant Biol.* **1**, 181–190 (2008).

80. Sensalari, C., Maere, S. & Lohaus, R. ksrates: positioning whole-genome duplications relative to speciation events in KS distributions. *Bioinformatics* **38**, 530–532 (2021).