Discovering Dynamic Plant Enzyme Complexes in Yeast for Kratom Alkaloid Pathway Identification

Dr. Yinan Wu[+],

Chang Liu[+],

Dr. Anna Koganitsky,

Franklin L. Gong,

Prof. Sijin Li[*]

[*]Robert F. Smith School of Chemical and Biomolecular Engineering, Cornell University, 14853 Ithaca, NY (USA)

Abstract

Discovering natural product biosynthetic pathways of medicinal plants is challenging and laborious. Capturing the coregulation patterns of pathway enzymes, particularly transcriptomic regulation, has proven an effective method to accelerate pathway identification. In this study, we developed a yeast-based screening method to capture the protein-protein interactions (PPI) between plant enzymes, which is another useful pattern to complement the prevalent approach. Combining this method with plant multiomics analysis, we discovered four enzyme complexes and their organized pathways from kratom, an alkaloid-producing plant. The four pathway branches involved six enzymes, including a strictosidine synthase, a strictosidine β-D-glucosidase (MsSGD), and four medium-chain dehydrogenase/reductases (MsMDRs). PPI screening selected six MsMDRs interacting with MsSGD from 20 candidates predicted by multiomics analysis. Four of the six MsMDRs were then characterized as functional, indicating the high selectivity of the PPI screening method. This study highlights the opportunity of leveraging posttranslational regulation features to discover novel plant natural product biosynthetic pathways.

Graphical Abstract

Conflict of Interest
The authors declare no conflict of interest.
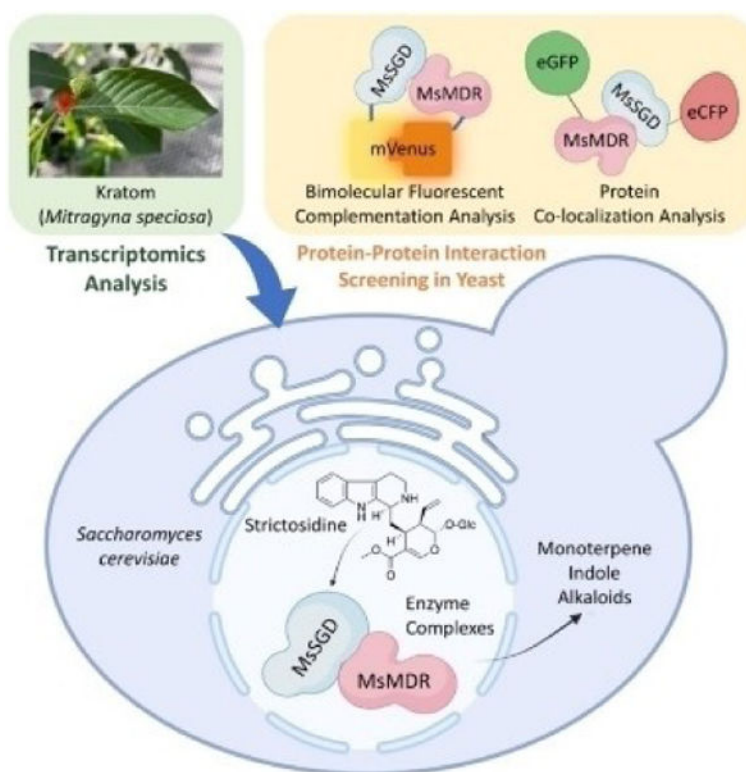
Yeast-based protein-protein interaction (PPI) screening identified dynamic enzyme complexes and pathways they organize from a rare plant, kratom. PPI screening identified four functional medium-chain dehydrogenases (MsMDRs) interacting with strictosidine β-D-glucosidase (MsSGD), leading to four novel pathway branches. This study highlights the opportunity of leveraging post-translational regulation features to accelerate plant pathway discovery.

Keywords

Biosynthesis; Monoterpene Indole Alkaloids; Natural Products; Pathway Discovery; Plant Enzymes

Introduction

Plant natural products (PNPs) from medicinal plants occupy unique structural space and important therapeutic niches. Identifying their biosynthetic pathways provides new biochemical knowledge for drug discovery and development and enables biomanufacturing of pharmaceutical PNPs. Methods to elucidate a PNP biosynthetic pathway usually begin with a hypothetical biochemical pathway, which consists of different types of enzymes to catalyze the sequential biosynthetic reactions. Tens to hundreds of candidate genes that might encode the predicted enzymes are then identified from sequenced plant transcriptomes, for example, using sequence alignment to predict the function of the enzymes encoded by the candidate genes. Finally, each candidate gene is functionally characterized *in planta* or using heterologous hosts, such as *Escherichia coli*, *Saccharomyces*

cerevisiae, and *Nicotiana benthamiana.* However, this process can be slow and costly. Functional characterization in the native host is limited to a handful of plants that can be engineered directly, while biochemical characterization through heterologous gene expression requires specific substrates, cofactors, and reaction conditions. Therefore, improving the accuracy of pathway gene prediction can significantly accelerate pathway identification and advance the understanding of plant specialized metabolism.

PNP pathways are regulated at various levels, including genomic, transcriptomic, epigenomic, and post-translational regulation. Plant omics analysis can capture multiple regulatory patterns to increase gene prediction accuracy. Transcriptomics analysis, such as co-expression analysis, has been primarily used to capture the similar expression patterns of genes at the RNA level. Combining transcriptomics analysis with plant metabolomics, which captures the pathway information at the metabolite level in different plant samples, [1] further increases the prediction accuracy. This multiomics approach has successfully elucidated many PNP pathways from various medicinal plants, such as mayapple,[2] kava,[3] and *Gloriosa superba*.[4] At the genomic level, patterns such as operon-like biosynthetic gene clusters (BGCs)[5] in plants have been found to be another useful cue for pathway predictions. Over 30 BGCs have been identified in the past decade,[6,7] leading to the discovery of diverse pathways from plants, such as the opium poppy,[8,9] cucumber,[10] St John's wort,[11] and Ginkgo.[12]

Post-translational regulation plays a crucial role in PNP biosynthesis, yet it has been less explored than other regulatory mechanisms. Specifically, the dynamic assembly of plant enzymes into enzyme complexes *in planta* is a protein-level cue that indicates the spatial organization of pathways but has been underutilized for pathway discovery. Plant enzyme complexes, or metabolons,[13] selectively assemble sequential enzymes in a pathway through proteinprotein interactions (PPIs) to regulate PNP biosynthesis. To date, over 30 enzyme complexes in diverse plants and PNP pathways have been characterized or partially identified, including the dhurrin pathway from sorghum,[14] the bitter acid pathway from hop,[15,16] isoflavonoid pathways from legumes,[17,18] and multiple phenylpropanoid pathways from *Arabidopsis thaliana*, snapdragon, torenia, and other plants.[19] In particular, three bipartite complexes are found in the medicinal plant *Catharanthus roseus*, which regulate the synthesis of pharmaceutical monoterpene indole alkaloids (MIAs).[20,21] Complex mapping studies have suggested the abundant existence of uncharacterized complexes from different plants.[22,23] These studies have indicated that dynamic plant enzyme complexes can be common in plant specialized metabolism. Thus, PPIs between enzymes in a complex can be leveraged as a protein-level pattern to search for putative pathway enzymes with high selectivity in different plants. Furthermore, PPI screening can serve as a convenient step to select genes predicted by multi-omics analysis before proceeding with biochemical characterization, as spontaneous PPIs between plant enzymes do not depend on unique substrates. However, identifying PPIs *in planta* is limited by throughput due to slow plant growth, and that the transient interactions between plant enzymes often lead to high false positive/negative rates with a single identification method. Baker's yeast (*S. cerevisiae*) has proven to be a feasible host for heterologous PPI identification. There are many modularized synthetic biology tools for rapidly constructing engineered yeasts to functionally express plant enzyme candidates. Multiple methods have

been developed in yeast for plant PPI identification, including the yeast 2-hybrid assay, bimolecular fluorescence complementation (BiFC), co-immunoprecipitation, and fusion fluorescent protein localization assays.[24] Therefore, combining yeast engineering and orthogonal PPI identification methods can enable a high-throughput substrate-free assay with high accuracy.

This work presents a yeast-based PPI screening method that leverages protein-level patterns of PNP pathways for putative gene prediction. We chose to develop the proposed method on a valuable yet less-investigated medicinal plant named kratom[25,26] (*Mitragyna speciosa.* Figure 1A), which produces various MIAs with great pharmaceutical potential (e.g., mitragynine and speciogynine as opioid agonists[27] and biased analgesics[28]). Despite limited studies on *M. speciosa* MIA pathways[29,30] and no reports on MIA complexes, our discovery was guided by the identified *C. roseus* MIA enzyme complexes with conserved structures, that is, each containing a nuclear strictosidine β-D-glucosidase, SGD, and a medium-chain dehydrogenase/reductase, MDR.[20,21] In this study, we first developed a yeast-based method combining BiFC and fusion protein localization analyses for complex identification and validated it using a characterized complex derived from *C. roseus* MIA metabolism (Figure 1B). We then used this method to search for conserved complexes in *M. speciosa* based on gene candidates predicted by plant transcriptomics and metabolomics analyses. More specifically, multi-omics analysis of *M. speciosa* led to one functional strictosidine synthase (MsSTR) producing strictosidine (**1** in Figure 1C), one functional SGD (MsSGD), and 20 uncharacterized MsMDR candidates. PPI screening further refined the MsMDR candidates and identified six candidates that interact with MsSGD in yeast. Subsequent biochemical characterization confirmed that four out of the six interacting MsMDRs are functional enzymes leading to five known MIAs (**2**, **3**, **4**, **5**, **6** in Figure 1C) and one uncharacterized MIA (**7**), yielding a 67% success rate in pathway gene prediction. The discovery of four MsSGD-MsMDR complexes and their corresponding pathways demonstrates the feasibility of the yeast-based PPI screening method. This proof-of-concept discovery highlights the utility of PPI as an additional pattern for predicting PNP pathways and the potential of interactomics analysis to advance the discovery of PNP pathways.

Results and Discussion

Development of yeast-based PPI identification methods

C. roseus tetrahydroalstonine synthases 1 (CrTHAS1)[20] and 2 (CrTHAS2),[21] as well as heteroyohimbine synthase (CrHYS)[21] are CrMDRs known to interact with CrSGD. The binary PPIs have been proven by BiFC assays in *C. roseus* cells.[20,21] The subcellular localization of each enzyme was also characterized in *C. roseus* via the fusion and expression of fluorescent proteins.[21] However, the *inplanta* PPI identification methods suffer from low throughput and long turnaround time and have yet to be developed in most plants. We leveraged yeast as an efficient and heterologous platform to identify the PPIs between plant enzymes using BiFC. In parallel, we examined the localization of single and co-expressed plant enzymes in yeast using fluorescent fusion proteins. This method was validated using the reported PPI between CrSGD and CrHYS.

To conduct the BiFC assay, we constructed four sets of modular plasmids with different fusion designs in yeast. Split mVenus fragments (NV and CV)[31] were fused to the N/C terminus of CrSGD and CrHYS and compared (Figure S1). In the optimal design, CV was fused to the N-terminus of CrSGD (CV-CrSGD) to maintain the C-terminal nuclear localization signal (NLS).[32] NV was fused to the N-termini of CrHYS (NV-CrHYS) and a *Papaver somniferum* methyltransferase, Ps4′OMT[33] (NV-Ps4′OMT), which is unlikely to interact with *C. roseus* enzymes, as the negative control. Engineered yeasts were cultured for three days and analyzed using a confocal microscope. BiFC results showed that CrSGD and CrHYS interacted in the nucleus (Figure 2A), which is consistent with the BiFC result in *C. roseus.*[21] Notably, we did not detect any interactions between CrSGD and Ps4′OMT (Figure 2A). To further confirm the PPI, we performed a fusion protein co-localization assay, which showed that CrSGD was always targeted to the nucleus when expressed individually (Figures 2B1–2B4) or co-expressed with CrHYS (Figure 2C). CrHYS showed a nucleocytoplasmic localization when expressed alone, and the fluorescence signal was distributed evenly (Figures 2B5–2B8). Co-expressing CrHYS with CrSGD relocalized a significant portion of CrHYS to the nucleus (Figure 2C), indicating that the assembly of the CrSGD-CrHYS complex is driven by molecular-level PPI rather than simple protein compartmentalization. As previous *in-planta* assays showed that CrHYS has a preferential nuclear localization in *C. roseus*,[21] we speculate that single CrHYS was not targeted to the yeast nucleus preferentially due to its plant-specific NLS (a KKKR sequence)[34] that might not be recognized effectively in yeast.

Identification of MsSTR and MsSGD by metabolomics and transcriptomics analyses of M. speciosa

We analyzed the metabolite profiles of four tissues (mature leaf, young leaf, root, and flower) of *M. speciosa* plants. The metabolomics analysis showed that the leaf of *M. speciosa* accumulated more MIAs, such as strictosidine, mitragynine, speciogynine, and speciociliatine, compared to the root or flower tissues, and the MIA distributions in the young and mature leaves showed notable differences (Figure S2). Therefore, we used mature (M) and young (Y) leaf samples to assemble a de novo *M. speciosa* leaf transcriptome (Table S1). We used gene mining to identify MsSTR, which catalyzes the first committed step in the MIA biosynthetic pathway (Figure 1C). We used the characterized STRs from *C. roseus*[35] and *Rauvolfia serpentina*[36] (CrSTR and RsSTR) to search against the *M. speciosa* transcriptome and identified one candidate gene. The MsSTR candidate was functionally characterized in yeast fed with secologanin and tryptamine and was compared with a functional CrSTR variant (CrSTR) cloned from the de novo strictosidine-producing yeast.[37] Liquid chromatography-mass spectrometry (LC–MS) results showed that MsSTR produced strictosidine (**1**, *m/z* 531.2337, 20 ppm), which was not observed in the negative controls (cultural medium without yeast and with yeast harboring a blank vector) (Figure 3A). The characterized MsSTR was then used to synthesize strictosidine as the essential upstream precursor in yeast. We then identified MsSGD using the characterized CrSGD and RsSGD.[32] Engineered yeasts and a blank control were lysed and tested with approximately 0.1 mM of crude strictosidine synthesized by MsSTR at 30°C for 1 h. CrSGD, RsSGD, and MsSGD consumed 99%, 98%, and 70% of the strictosidine (Figure 3B). All SGDs′ products showed the same MS peaks identical as reported previously,[38] including one

major peak of *m/z* 351.1703 (Figure 3C), which is likely cathenamine, and multiple smaller peaks of *m/z* 351.1703 and *m/z* 369.1809 (Figure S3). MsSGD has the conserved plant β-glucosidase sequence and four identical active residues with CrSGD and RsSGD[39] (Figure S4) and a conserved C-terminal bipartite NLS (522-**KR**TLED-HEDFVS**KKR**L**R**Q-539). Constitutive expression in yeast (Figure 3D) and transient expression in *N. benthamiana* leaf (Figure 3E) further confirmed the localization of MsSGD in the nucleus, which aligned with the nuclear marker. As CrSGD and RsSGD self-assemble to form a multimer in the plant nucleus,[32] we validated the self-assembly of MsSGD in vitro using *E. coli*. As expected, MsSGD showed up as a multimer band in native electrophoresis (Figure 3F) and appeared as a monomer band after denaturation in SDS-PAGE (Figure 3G). We then searched for putative MsMDRs, which are potential MIA pathway enzymes downstream of MsSGD as identified in previous *C. roseus* MIA metabolism.[21,40] We identified 247 candidates that were functionally annotated as MDRs or functionally similar enzymes, such as alcohol dehydrogenases. We selected 83 candidates based on relatively high expression in the samples and prioritized them using differential expression analysis. Twenty candidates showed over four-fold difference in young and mature leaf samples and were named as MsMDR1-20. Although we performed a co-expression analysis of MsMDR and MsSGD, we could not further prioritize the MsMDR candidates due to the low abundance of MsSGD transcripts in the samples in comparison to other genes.

Identification of four functional MsMDRs from six candidates interacting with MsSGD

We first screened all 20 MsMDR candidates′ interactions with MsSGD in yeast to identify novel MsSGD-MsMDR complexes by both BiFC (Figure 4A and Figure S5) and protein localization analysis (Figure 4B–4 C and Figure S6) using the modular plasmids established. Among the 20 MsMDR candidates screened, six showed positive PPI in both BiFC and co-localization assays (Figure 4), including MsMDR1 (later named MsTHAS), 4, 6, 11, MsMDR12 (later named MsHYS), and MsMDR20 (later named MsDCS). BiFC assays revealed that all PPIs took place in the yeast nucleus, while the negative control Ps4′OMT did not yield any fluorescence (as shown in Figure 2A2). Single protein localization assays confirmed that all six MsMDR candidates localized in the cytoplasm (Figure 4B). Co-expressing MsSGD with MsMDRs relocalized all six MsMDR candidates preferentially in the nucleus (Figure 4C). The relocalization was not observed in the remaining MsMDRs (Figure S6). We further validated the positive MsSGD-MsMDR interactions in *N. benthamiana* or in vitro. In the presence of MsSGD, all six interacting MsMDRs showed concentrated fluorescent signals in the nucleus of *N. benthamiana* (Figure S7). We further expressed and purified MsHYS, MsSGD, and Ps4′OMT in *E. coli* and confirmed the PPIs in vitro by gel electrophoresis. A new band indicating the PPI only showed up in the presence of both MsSGD and MsHYS and was analyzed by in-gel digestion and LC–MS/MS (Figure S8).We then characterized the biochemical activity of the six MsMDRs to prove the hypothesis that plant enzymes that can form a complex are functionally related. We used engineered yeast co-expressing MsSGD and MsMDR instead of purified enzymes to increase the characterization efficiency. As strictosidine aglycone is reactive and unstable, approximately 0.1 mM strictosidine was added to the crude yeast cell lysate and was converted to strictosidine aglycone by the MsSGD co-expressed in yeast. Four of the six interacting MsMDRs (MsHYS, MsTHAS, MsDCS, and MsMDR4) converted strictosidine

aglycone to downstream MIAs (Figure 5). MsHYS produced three heteroyohimbine-type MIAs, including ajmalicine (**2**), tetrahydroalstonine (**3**), and mayumbine (**4**). The products were identified by comparing their retention times (26.69 min, 27.06 min, and 27.25 min), MS (*m/z* 353.1860, 20 ppm), and tandem MS/MS spectra with standards (for ajmalicine and tetrahydroalstonine) or the products of CrHYS (for mayumbine)[21] (Figure 5A and Figure S9). MsTHAS produced **3** only (Figure 5B). MsMDR4 produced a trace amount of an unknown compound with the representative *m/z* of MIA of 353.1860 (20 ppm) in MS (**7**, Figure 5C) and indole-related moiety in MS/MS (*m/z* of 144.081) identical to those of ajmalicine and tetrahydroalstonine (Figure S9), indicating the production of an MIA. As this product was not present in the leaf extract of *M. speciosa* (Figure S10), it is likely a downstream intermediate from strictosidine aglycones. MsDCS produced two MS peaks with *m/z* of 355.2016 (20 ppm) (Figure 5D). The peak eluted at 27.4 min is identical to (20*R*)-dihydrocorynantheine (**6**), which was produced by the yeast co-expressing MsSGD and the characterized *Cinchona pubescens* dihydrocorynantheine synthase (CpDCS).[41] The peak eluted at 25.8 min was specifically produced by MsDCS and was also observed in the young leaf of *M. speciosa* (Figure S10), which we postulated is (20*S*)-dihydrocorynantheine (**5**), the putative precursor of mitragynine. The production of **5** and **6** was also reported in a recent study by Schotte and Jiang,[30] in which the enzyme was named *M. speciosa* dihydrocorynantheine synthase (MsDCS1).

In addition, we found that MsDCS produced **3**, while CrTHAS1 produced **6** (Figure S11).

Taken together, the PPI screening in yeast identified six interacting MsMDR candidates. Of these six candidates, four were subsequently confirmed as functional. In contrast, the remaining 14 MsMDR candidates that did not exhibit apparent PPI with MsSGD were screened, but no detectable products were observed (Figure S12). As a result, the additional PPI screening enhanced the efficiency of gene identification from 25% (4 out of 6 after PPI screening) in a substrate-free way, which can be applied as a complementary method to select pathway gene candidates and avoid large-scale biochemical characterization in the future.

Heteroyohimbine pathway development in yeast

We then reconstituted the heteroyohimbine-type MIA biosynthetic pathway in yeast to demonstrate the *M. speciosa* pathway′s productivity in vivo. We co-expressed MsSTR, MsSGD, and MsHYS in yeast and fed 0.5 mM secologanin and 0.5 mM tryptamine during fermentation. After 72 h, the engineered yeast produced 1.76 μM ajmalicine (**2**), 1.33 μM mayumbine (**4**), and 0.40 μM tetrahydroalstonine (**3**), respectively (Figure 6). The ratio among the three MIA products (**2** : **3** : **4** = 1 : 0.09 : 0.75) was different from that obtained in the in vitro cell lysate assay (1 : 0.16 : 0.06). As the in vivo fermentation exhibited an obvious preference for **4** over **3** throughout the 72 h fermentation process, we evaluated whether the product profile change was related to the change of pH during fermentation. We adjusted the pH in in vitro cell lysate assays to 4, 5.5, 7, 7.4, and 8, respectively, and observed no obvious difference in the composition of products (Figure S13). In addition, as the complex localizes in the yeast nucleus that usually maintains a neutral pH, we postulated that the change of pH, if existing, would not alter the product composition significantly.

Therefore, the formation of the MsSGD-MsHYS complex might be necessary to the varied product preference.

Collectively, this work presents a yeast-based PPI screening method to exploit the PPIs between plant pathway enzymes as a protein-level feature for PNP pathway identification from non-model plants, using a less-studied MIA-producer, *M. speciosa*, as an example. The effectiveness of this method was demonstrated by the identification of MsMDRs in this study. While identifying MsSTR and MsSGD was relatively efficient due to the limited number of candidates predicted by transcriptomics analysis, identifying MsMDRs downstream of MsSGD using traditional plant multi-omics analysis could be laborious and costly. With more than 200 genes annotated as *MsMDR*s, and the MDR-catalyzed reactions potentially involving unstable substrates (strictosidine aglycones) and products,[32] combining tissue-specific expression level analysis and MsSGD-MsMDR co-expression analysis still left 20 candidates to be screened biochemically. Our work demonstrated that yeast-based PPI screening narrowed these candidates to six, and further biochemical characterization of the six candidates identified four functional MsMDRs (MsTHAS, MsHYS, MsDCS, and MsMDR4). Combining the proposed substrate-free PPI screening with multi-omics analysis increased the selection ratio to 67% (4/6) compared to 20% (4/20) if only relying on the multi-omics analysis. The biochemical screening of the other 14 MsMDR candidates was conducted in this work for the complete characterization of the yeast-based PPI screening method and can be omitted in future applications. Moreover, the modular plasmids for BiFC and protein localization assays in yeast will enhance PPI screening efficiency, and higher-throughput analyses, such as flow cytometry, can be used to further optimize the throughput of PPI screening. Global PPI screening or interactomics analysis can be implemented to cover all potential interacting candidates in a plant in the future.

The proposed PPI screening method focuses on identifying PNP pathways spatially organized by enzyme complexes, and its future application depends on the abundance of complexes in nature. Multiple complexes have been identified in various PNP pathways from both model and medicinal plants.[14–19] Emerging plant interactomics studies have further found thousands of PPIs in diverse plants (e.g., 56000 PPIs from maize[22] and 6200 PPIs from *Arabidopsis*,[23] suggesting more complexes to discover. Therefore, although this PPI-centered method can only identify a certain type of PNP pathways, the unknown complexes remain a rich resource to exploit. First, information on known complexes will guide the discovery of conserved complexes across different plant species, as this proof-of-concept study has showcased that the known SGD-MDR complexes in *C. roseus* (including CrTHAS1,[20] CrTHAS2,[21] and CrHYS[21]) can guide the discovery of conserved *M. speciosa* complexes and pathways. In addition, the advantages of the yeast-based PPI screening method, such as high selectivity and substrate-free, can be leveraged to discover novel complexes without known conserved structures by largescale, high-throughput PPI screening in the future.

It is laborious to identify the transient and sometimes weak PPIs in plant enzyme complexes using the traditional *in planta* or in vitro method, as a single detection method usually leads to high false positive/negative rates, while the combination of multiple detection methods *in*

planta or in vitro is usually limited by the difficulty and time needed in plant engineering or protein purification. Performing orthogonal PPI assays in yeast can address this obstacle for multiple reasons. First, yeast is a faster host than plants for rapid in vivo PPI assay and is easy to engineer thanks to the abundant engineering tools and methods. For example, we have developed synthetic biology tools and methods to construct modular plasmids available for this study and future yeast-based PPI identification assays. The modular plasmid sets covered all four possible fusion protein designs in BiFC (Figure S1) and led to an optimal design (CV fused to the N-terminus of SGD and NV fused to the N-terminus of MDR) for all following SGD-MDR interactions. Second, the higher throughput of yeast than other hosts enables the combination of orthogonal PPI identification methods and is important to increase the PPI screening specificity. In this work, we used the BiFC assay to screen the putative PPIs at the molecular level due to its high throughput and ease of identification and used fusion protein localization assay to validate the result based on the organelle level protein localization. The combination of these two assays can capture PPIs involving cytosolic, membrane-bound, or nuclear proteins within a living cell and report the subcellular localization of the PPI with high accuracy. Particularly, the yeast-based PPI screening method is suitable to identify PPIs involving membrane-bound cytochrome P450s, which have been reported as the conserved scaffold of many plant complexes.[14,19] Protein localization assay proved important to exclude false positive PPIs obtained in BiFC assay. For example, six MsMDR candidates (MsMDR2, 3, 5, 7, 18, and 19) showed varied yet positive interactions with MsSGD in the BiFC assay (Figure S5). However, results from paralleled fusion protein localization assay provided more information to distinguish them from interacting MsMDRs. Fluorescent signals of individually expressed MsMDRs were detected in the yeast nucleus, indicating that these six MsMDRs are naturally nucleocytoplasmic or nuclear proteins. Co-expressing eCFP-MsSGD and eGFP-MsMDR then confirmed that none of the six MsMDR relocalized or aggregated more in the nucleus in the presence of MsSGD (Figure S6), unlike the obvious relocalization observed in the six interacting MsMDRs (Figure 4) or in previous CrSGD-CrHYS interaction assay.[21] These suggest that the positive BiFC results are likely due to the compartmentalization of these nuclear MsMDRs with MsSGD in the same organelle rather than the spontaneous PPI, and these six candidates are likely non-interacting MsMDRs or MsMDRs with undetected, weak interactions.

The MsSGD-MsMDR complexes identified regulate different pathway branches. MsSGD-MsTHAS and MsSGD-MsHYS are similar to the identified *C. roseus* complexes producing heteroyohimbine-type, while MsSGD-MsDCS and MsSGD-MsMDR4 have no known counterparts in *C. roseus.* It is plausible to conclude that the PPI between SGD and MDR is common in different types of MIA biosynthesis and different plants. The in vivo reconstruction of the MsSGD-MsHYS complex in yeast and subsequent fermentation indicated that this complex might redirect metabolic flux to tune the ratio of the products (ajmalicine, tetrahydroalstonine, and mayumbine) dynamically via spatial organization. Meanwhile, as the in vitro assay lacks intact nuclear context, it is possible that the MsSGD-MsHYS complex was not present or could not regulate the MIA biosynthesis efficiently in vitro. Investigation of the enzymatic characteristics of MsTHAS, MsDCS, and MsMDR4 and the function of corresponding complexes were limited by the activities

of these enzymes and might be obtained after further optimization in future studies. We tried to link the observed production of different MIAs in yeast with the metabolomic information of *M. speciosa* leaf samples. Young leaf samples accumulated more ajmalicine (**2**) and tetrahydroalstonine (**3**) than mature leaf samples, while mayumbine (**4**) was merely detectable (Figure S10), which might have been converted to downstream MIA derivatives by uncharacterized enzymes. (20*S*)-dihydrocorynantheine (**5**) accumulated more in young leaves, while (20*R*)-dihydrocorynantheine (**6**) was hard to observe in any of the plant samples (Figure S10). The different accumulations might also be related to the downstream pathway activities, as the concentration of speciogynine, which is the downstream product of **6**, is much higher than that of mitragynine, the downstream product of **5** (Figure S2). Similarly, the novel product of MsMDR4 was not observed in the metabolome, indicating a downstream pathway that might have converted the uncharacterized novel product to derivatives.

Conclusion

We have demonstrated a yeast-based PPI screening method to complement prevalent plant transcriptomics and metabolomics analyses for pathway discovery by providing posttranslational information about PNP pathways. The integration of pathway pattern information at RNA, metabolite, and protein levels led to the identification of six functional enzymes, including MsSTR, MsSGD, and four MsMDRs, which comprise four different MIA pathway branches. By capturing PPI patterns in yeast, the accuracy of pathway prediction was improved, and the effort and cost of subsequent biochemical screening were reduced. In this study, we also developed a modular synthetic biology approach combining BiFC assay and protein localization assay in yeast for more efficient PPI screening than prevalent *in planta* or in vitro methods. The established plasmids for PPI assays and the fast-growing yeast shorten the screening time and can be expanded for larger-scale PPI screening in the future. Inspired by the BGC-based methods that have provided additional genomic patterns for accelerated PNP pathway identification, we expect the PPIs to be another useful pattern at the protein level to accelerate PNP pathway discovery.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Data Availability Statement

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

References

[1]. Singh KS, Van der Hooft JJJ, Van Wees SCM, Medema MH, Nat. Prod. Rep 2022, 39, 1876–1896. [PubMed: 35997060]

[2]. Lau W, Sattely ES, Science 2015, 349, 1224–1228. [PubMed: 26359402]

[3]. Pluskal T, Torrens-spence MP, Fallon TR, de Abreu A, Shi CH, Nat. Plants 2019, 5, 867–878. [PubMed: 31332312]

[4]. Nett RS, Sattely ES, J. Am. Chem. Soc 2021, 143, 19454–19465. [PubMed: 34780686]

[5]. Nett RS, Lau W, Sattely ES, Nature 2020, 584, 148–153. [PubMed: 32699417]

[6]. Zhan C, Shen S, Yang C, Liu Z, Fernie AR, Graham IA, Luo J, Trends Plant Sci 2022, 27, 981–1001. [PubMed: 35365433]

[7]. Polturak G, Osbourni A, PLoS Pathog 2021, 17, e1009698. [PubMed: 34214143]

[8]. Chen X, Hagel JM, Chang L, Tucker JE, Shiigi SA, Yelpaala Y, Chen HY, Estrada R, Colbeck J, Enquist-Newman M, Ibáñez AB, Cottarel G, Vidanes GM, Facchini PJ, Nat. Chem. Biol 2018, 14, 738–743. [PubMed: 29807982]

[9]. Li Y, Smolke CD, Nat. Commun 2016, 7, 12137. [PubMed: 27378283]

[10]. Shang Y, Ma Y, Zhou Y, Zhang H, Duan L, Chen H, Zeng J, Zhou Q, Wang S, Gu W, Liu M, Ren J, Gu X, Zhang S, Wang Y, Yasukawa K, Bouwmeester HJ, Qi X, Zhang Z, Lucas WJ, Huang S, Science 2014, 346, 1084–1088. [PubMed: 25430763]

[11]. Wu S, Malaco Morotti AL, Wang S, Wang Y, Xu X, Chen J, Wang G, Tatsis EC, New Phytol 2022, 235, 646–661. [PubMed: 35377483]

[12]. Forman V, Luo D, Geu-Flores F, Lemcke R, Nelson DR, Kampranis SC, Staerk D, Møller BL, Pateraki I, Nat. Commun 2022, 13, 5143. [PubMed: 36050299]

[13]. Zhang Y, Fernie AR, Plant Commun 2021, 2, 100081. [PubMed: 33511342]

[14]. Laursen T, Borch J, Knudsen C, Bavishi K, Torta F, Martens HJ, Silvestro D, Hatzakis NS, Wenk MR, Dafforn TR, Olsen CE, Motawia MS, Hamberger B, Møller BL, Bassard JE, Science 2016, 354, 890–893. [PubMed: 27856908]

[15]. Li H, Ban Z, Qin H, Ma L, King AJ, Wang G, Plant Physiol 2015, 167, 650–659. [PubMed: 25564559]

[16]. Ban Z, Qin H, Mitchell AJ, Liu B, Zhang F, Weng JK, Dixon RA, Wang G, Proc. Natl. Acad. Sci. USA 2018, 115, E5223–E5232. [PubMed: 29760092]

[17]. Mameda R, Waki T, Kawai Y, Takahashi S, Nakayama T, Plant J 2018, 96, 56–74. [PubMed: 29979476]

[18]. Dastmalchi M, Bernards MA, Dhaubhadel S, Plant J 2016, 85, 689–706. [PubMed: 26856401]

[19]. Nakayama T, Takahashi S, Waki T, Front. Plant Sci 2019, 10, 821. [PubMed: 31338097]

[20]. Stavrinides A, Tatsis EC, Foureau E, Caputi L, Kellner F, Courdavault V, O′Connor SE, Chem. Biol 2015, 22, 336–341. [PubMed: 25772467]

[21]. Stavrinides A, Tatsis EC, Caputi L, Foureau E, Stevenson CEM, Lawson DM, Courdavault V, O′Connor SE, Nat. Commun 2016, 7, 12116. [PubMed: 27418042]

[22]. Han L, Zhong W, Qian J, Jin M, Tian P, Zhu W, Zhang H, Sun Y, Feng JW, Liu X, Chen G, Farid B, Li R, Xiong Z, Tian Z, Li J, Luo Z, Du D, Chen S, Jin Q, Li J, Li Z, Liang Y, Jin X, Peng Y, Zheng C, Ye X, Yin Y, Chen H, Li W, Chen LL, Li Q, Yan J, Yang F, Li L, Nat. Genet 2023, 55, 144–153. [PubMed: 36581701]

[23]. Dreze M, Carvunis A-R, Charloteaux B, Galli M, Pevzner SJ, Tasan M, Ahn Y-Y, Balumuri P, Barabási A-L, Bautista V, Braun P, Byrdsong D, Chen H, Chesnut JD, Cusick ME, Dangl JL, de los Reyes C, Dricot A, Duarte M, Ecker JR, Fan C, Gai L, Gebreab F, Ghoshal G, Gilles P, Gutierrez BJ, Hao T, Hill DE, Kim CJ, Kim RC, Lurin C, MacWilliams A, Matrubutham

U, Milenkovic T, Mirchandani J, Monachello D, Moore J, Mukhtar MS, Olivares E, Patnaik S, Poulin MM, Przulj N, Quan R, Rabello S, Ramaswamy G, Reichert P, Rietman EA, Rolland T, Romero V, Roth FP, Santhanam B, Schmitz RJ, Shinn P, Spooner W, Stein J, Swamilingiah GM, Tam S, Vandenhaute J, Vidal M, Waaijers S, Ware D, Weiner EM, Wu S, Yazaki J, Science 2011, 333, 601–607. [PubMed: 21798944]

[24]. Bassard JE, Halkier BA, Phytochem. Rev 2018, 17, 211–227. [PubMed: 29755303]

[25]. Boyer EW, Babu KM, Adkins JE, McCurdy CR, Halpern JH, Addiction 2008, 103, 1048–1050. [PubMed: 18482427]

[26]. Veltri C, Grundmann O, Subst. Abuse Rehabil 2019, 10, 23–31. [PubMed: 31308789]

[27]. Suhaimi FW, Yusoff NHM, Hassan R, Mansor SM, Navaratnam V, Müller CP, Hassan Z, Brain Res. Bull 2016, 126, 29–40. [PubMed: 27018165]

[28]. Todd DA, Kellogg JJ, Wallace ED, Khin M, Flores-Bocanegra L, Tanna RS, McIntosh S, Raja HA, Graf TN, Hemby SE, Paine MF, Oberlies NH, Cech NB, Sci. Rep 2020, 10, 19158. [PubMed: 33154449]

[29]. Charoonratana T, Wungsintaweekul J, Keawpradub N, Verpoorte R, Acta Physiol. Plant 2013, 35, 2611–2621.

[30]. Schotte C, Jiang Y, Grzech D, Dang T-TT, Laforest LC, León F, Mottinelli M, Nadakuduti SS, McCurdy CR, O′Connor SE, J. Am. Chem. Soc 2023, 145, 4957–4963. [PubMed: 36883326]

[31]. Gookin TE, Assmann SM, Plant J 2014, 80, 553–567. [PubMed: 25187041]

[32]. Guirimand G, Courdavault V, Lanoue A, Mahroug S, Guihur A, Blanc N, Giglioli-Guivarc′h N, St-Pierre B, Burlat V, BMC Plant Biol 2010, 10, 182. [PubMed: 20723215]

[33]. Li Y, Li S, Thodey K, Trenchard I, Cravens A, Smolke CD, Proc. Natl. Acad. Sci. USA 2018, 115, E3922–E3931. [PubMed: 29610307]

[34]. Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita M, Yanagawa H, J. Biol. Chem 2009, 284, 478–485. [PubMed: 19001369]

[35]. Pasquali G, Goddijn OJM, De Waal A, Verpoorte R, Schilperoort RA, Hoge JHC, Memelink J, Plant Mol. Biol 1992, 18, 1121–1131. [PubMed: 1600148]

[36]. Ma X, Panjikar S, Koepke J, Loris E, Stöckigt J, Plant Cell 2006, 18, 907. [PubMed: 16531499]

[37]. Brown S, Clastre M, Courdavault V, O′Connor SE, Proc. Natl. Acad. Sci. USA 2015, 112, 3205–3210. [PubMed: 25675512]

[38]. Carqueijeiro I, Koudounas K, De Bernonville TD, Sepúlveda LJ, Mosquera A, Bomzan DP, Oudin A, Lanoue A, Besseau S, Cruz PL, Kulagina N, Stander EA, Eymieux S, Burlaud-Gaillard J, Blanchard E, Clastre M, Atehortùa L, St-Pierre B, Giglioli-Guivarc′h N, Papon N, Nagegowda DA, O′Connor SE, Courdavault V, Plant Physiol 2021, 185, 836–856. [PubMed: 33793899]

[39]. Barleben L, Panjikar S, Ruppert M, Koepke J, Stöckigta J, Plant Cell 2007, 19, 2886–2897. [PubMed: 17890378]

[40]. Tatsis EC, Carqueijeiro I, Dugé De Bernonville T, Franke J, Dang TTT, Oudin A, Lanoue A, Lafontaine F, Stavrinides AK, Clastre M, Courdavault V, O′Connor SE, Nat. Commun 2017, 8, 316. [PubMed: 28827772]

[41]. Trenti F, Yamamoto K, Hong B, Paetz C, Nakamura Y, O′Connor SE, Org. Lett 2021, 23, 1793–1797. [PubMed: 33625237]

[42]. Bolger AM, Lohse M, Usadel B, Bioinformatics 2014, 30, 2114–2120. [PubMed: 24695404]

[43]. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A, Nat. Protoc 2013, 8, 1494–1512. [PubMed: 23845962]

[44]. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM, Bioinformatics 2015, 31, 3210–3212. [PubMed: 26059717]

[45]. Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, Lee TJ, Leigh ND, Kuo TH, Davis FG, Bateman J, Bryant S, Guzikowski AR, Tsai SL, Coyne S, Ye WW, Freeman RM, Peshkin L, Tabin CJ, Regev A, Haas BJ, Whited JL, Cell Rep 2017, 18, 762–776. [PubMed: 28099853]

[46]. Li B, Dewey CN, BMC Bioinf 2011, 12, 323.

[47]. Robinson MD, McCarthy DJ, Smyth GK, Bioinformatics 2010, 26, 139–140. [PubMed: 19910308]

[48]. Alberti S, Gitler AD, Lindquist S, Yeast 2007, 24, 913–919. [PubMed: 17583893]

[49]. Wu Y, Chen MN, Li S, Metab. Eng. Commun 2022, 14, e00195. [PubMed: 35287355]
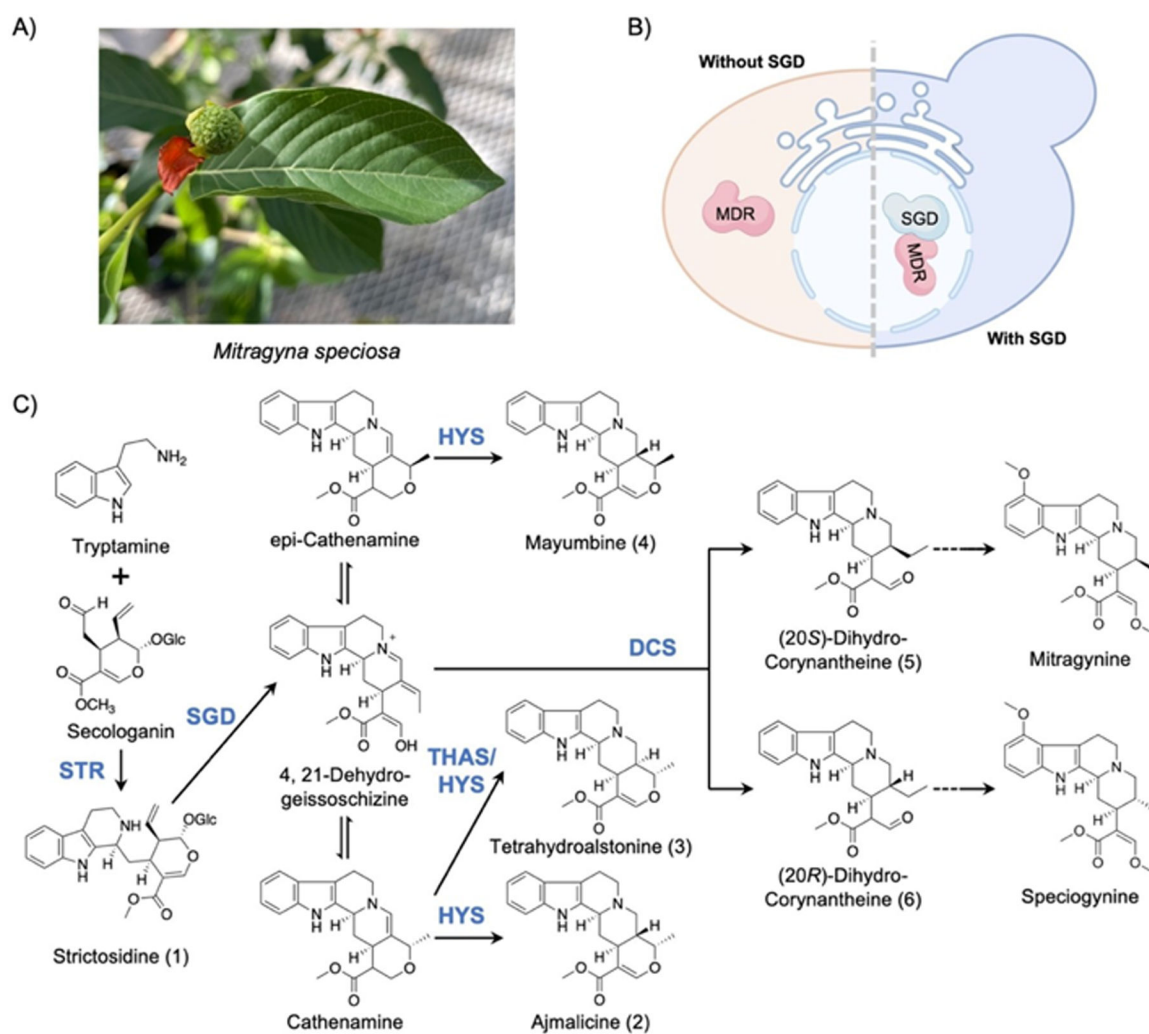
Figure 1.
Discovering *M. speciosa* SGD-MDR complexes and pathways in yeast. A) A photograph
of *M. speciosa* leaf and flower. B) Recruitment of MDR into the yeast nucleus by
SGD to form a complex. C) Representative *M. speciosa* MIA pathways involving
identified complexes. Solid arrows indicate reactions catalyzed by MsMDRs identified
in this study. Dashed arrows are uncharacterized reactions. Abbreviations: strictosidine
β-D-glucosidase (SGD), medium-chain dehydrogenase/reductase (MDR), strictosidine
synthase (STR), heteroyohimbine synthase (HYS), tetrahydroalstonine synthase (THAS),
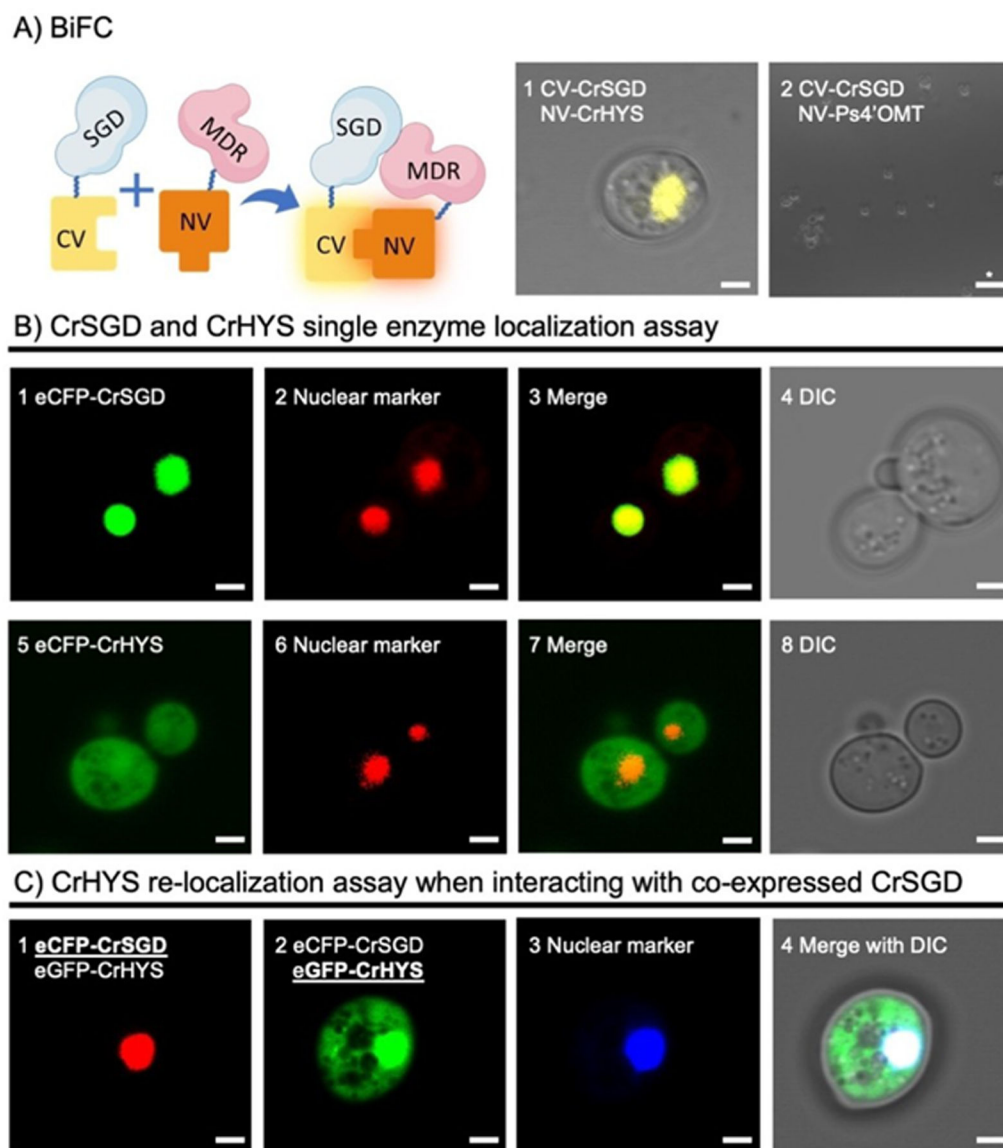dihydrocorynantheine synthase (DCS).

Figure 2.
CrSGD-CrHYS interaction validated in yeast using BiFC and localization analysis. A) Positive BiFC took place in yeast nucleus with fluorescence signal shown in yellow false color (A1). No fluorescence signal was detected in the negative control (A2). B) Single enzyme localization assay. Fluorescence signals were shown in red (nuclear marker) or green (CrSGD or CrHYS) false colors. C) Localization of CrHYS (C2) when co-expressed CrSGD (C1) in the presence of the nuclear marker (C3). Fluorescence signals were shown in red (CrSGD), green (CrHYS) or blue (nuclear marker) false colors. All signals were merged in C4. All scale bars are 2 μm except for the one labeled with an asterisk (15 μm).
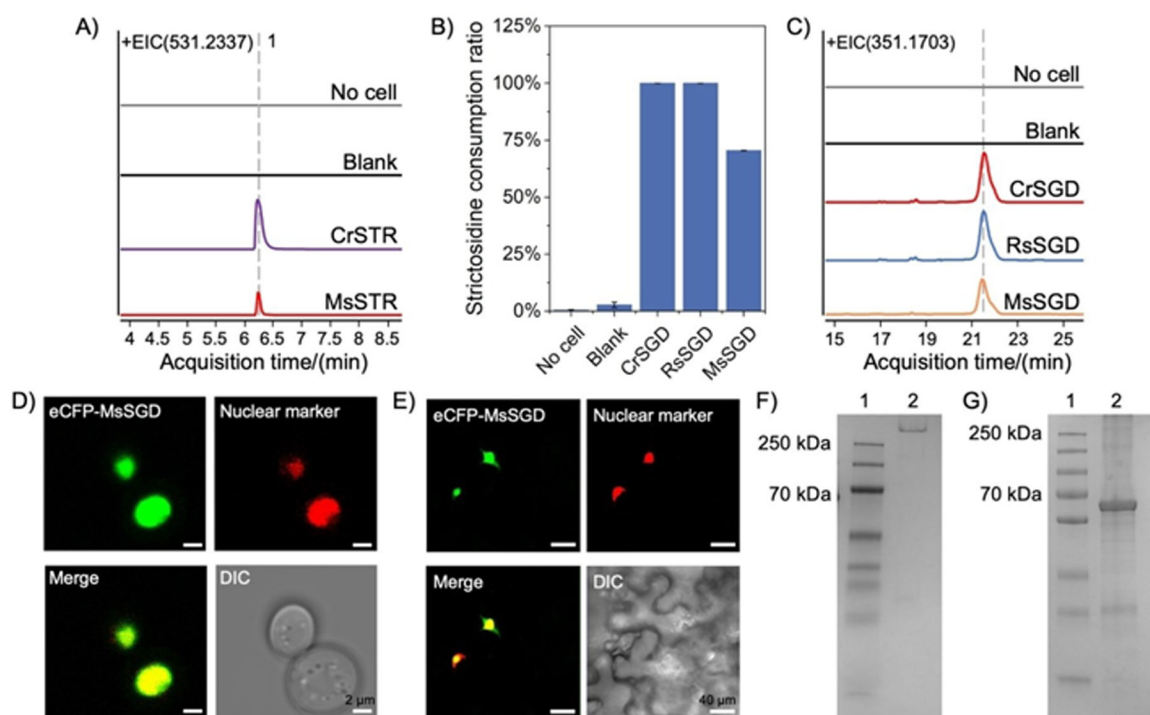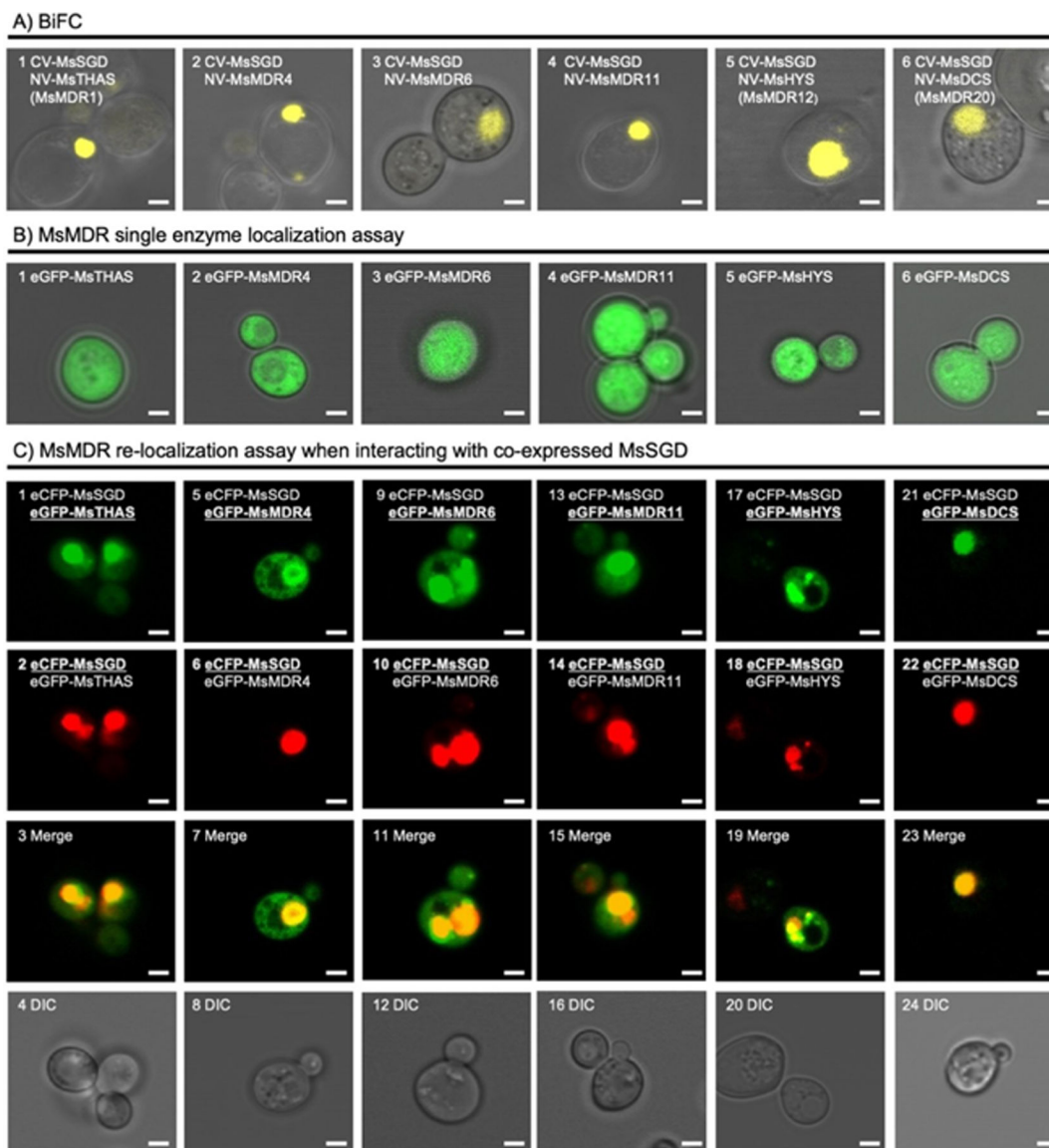
Figure 3.

Characterization of MsSTR and MsSGD using yeast, *N. benthaiana*, and *E. coli.* A) Production of strictosidine (**1**) by MsSTR and CrSTR. B) Deglycosylation of strictosidine by MsSGD, compared with characterized CrSGD and RsSGD. Error bars represent standard deviation ($n = 3$). C) Production of an *m/z* 351.1703 peak. D), E) Nuclear localization of MsSGD identified in yeast (D) and *N. benthamiana* (E) using green (MsSGD) or red (nuclear marker) false colors. F), G) Recombinant His-MsSGD protein (62.3 kDa) analyzed by native PAGE (F) or SDS-PAGE (G) followed by Coomassie blue staining. 1, protein ladder; 2, His-MsSGD multimer (F) or His-MsSGD monomer (G).

Figure 4.
Identified MsSGD-MsMDR interactions. A) MsSGD-MsMDR interactions identified by BiFC in the yeast nucleus. B) Localization of eGFP-MsMDRs (green) when expressed alone. C) Relocalization of MsMDRs (green) in the presence of MsSGD (red). Merged colors in yellow. Cell morphology is observed with differential interference contrast (DIC). Scale bar: 2 μm.
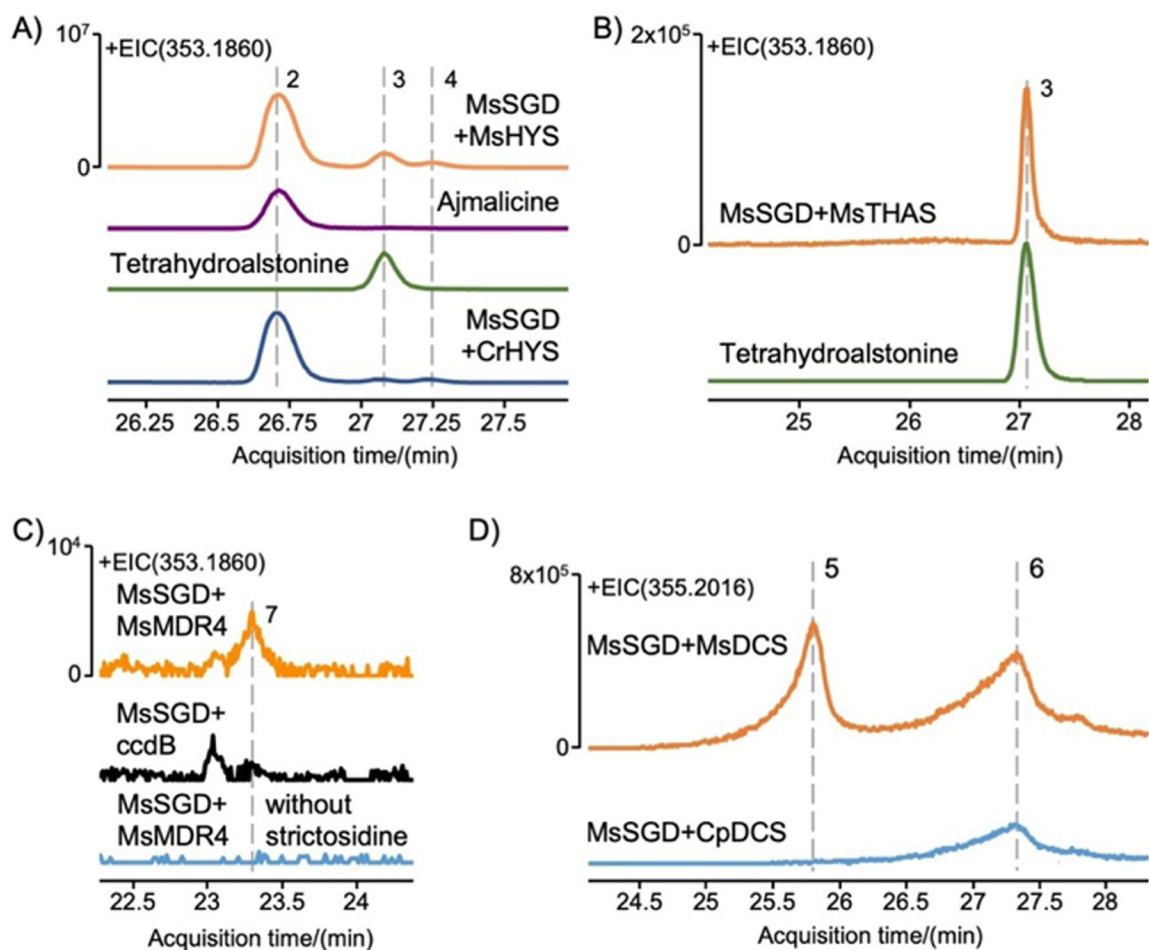
Figure 5.
Biochemical characterization of MsHYS, MsTHAS, and MsMDR4 and MsDCS. A) MsHYS produced ajmalicine (**2**), tetrahydroalstonine (**3**), and mayumbine (**4**). B) MsTHAS produced tetrahydroalstonine (**3**). C) MsMDR4 produced an unknown MIA (**7**) with *m/z* of 353.1860 (20 ppm). D) MsDCS produced (20*S*)-dihydrocorynantheines (**5**) and (20*R*)-dihydrocorynantheines (**6**).
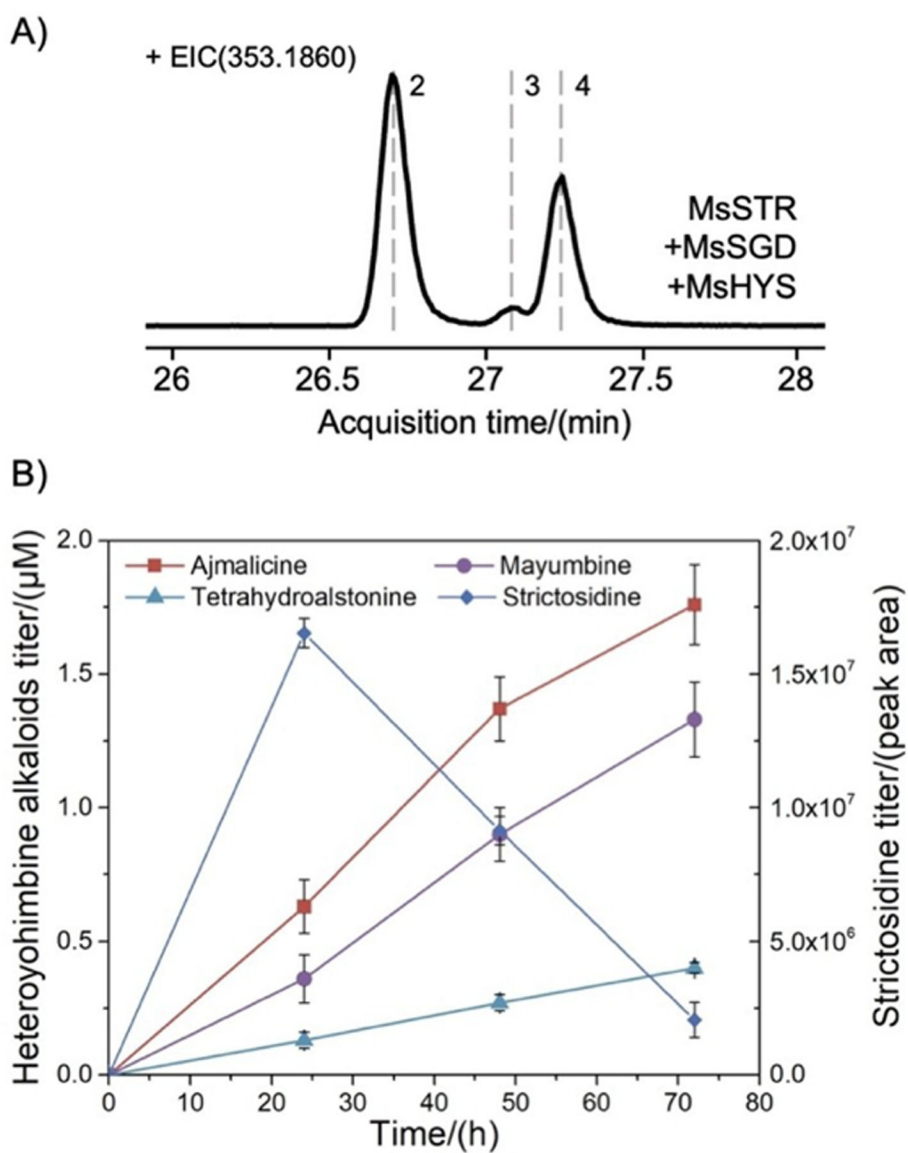
Figure 6.
Heteroyohimbine-type MIAs biosynthesis in yeast by reconstructed *M. speciosa* pathway. A) EIC (*m/z* 353.1860) of ajmalicine, tetrahydroalstonine, and mayumbine (**2**, **3**, and **4**) after 24 h fermentation. B) Time curves of strictosidine (**1**) and heteroyohimbine alkaloids production in 72 h fermentation. Error bars represent standard deviations ($n = 3$).