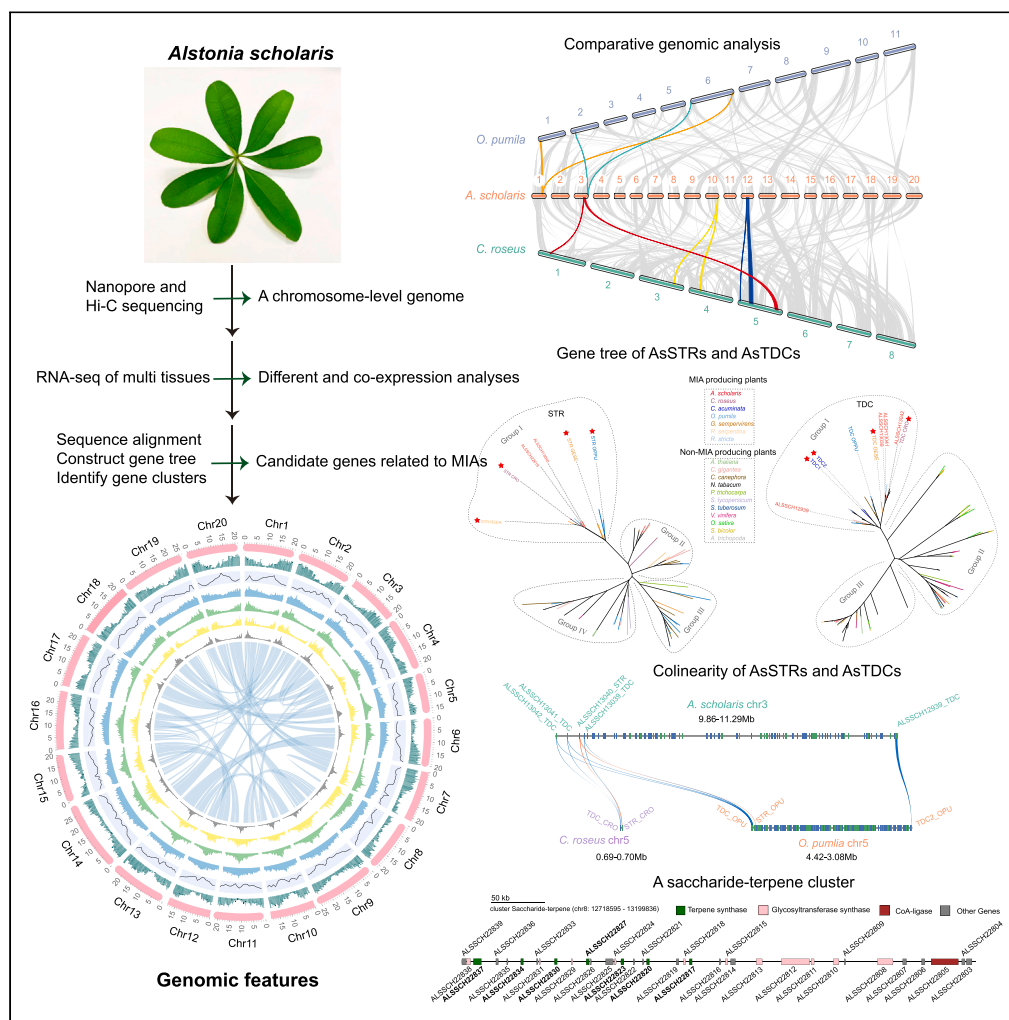


Article

Chromosome-level *Alstonia scholaris* genome unveils evolutionary insights into biosynthesis of monoterpenoid indole alkaloids



Haixia Chen, Sunil Kumar Sahu, Shujie Wang, ..., Le Cheng, Tsan-Yu Chiu, Huan Liu

qjucanyu@genomics.cn (T.-Y.C.)
liuhuan@genomics.cn (H.L.)

Highlights
Alstonia scholaris genome illuminates MIA biosynthesis evolution

Assembled 444.95 Mb genome into 20 chromosomes with high scaffold quality

Identified clusters for key enzymes in MIA biosynthetic pathway

Multi-omics data aid understanding of co-expression pattern of MIA genes



Article

Chromosome-level *Alstonia scholaris* genome unveils evolutionary insights into biosynthesis of monoterpenoid indole alkaloids

Haixia Chen,^{1,2,4} Sunil Kumar Sahu,^{1,2,4} Shujie Wang,¹ Jia Liu,³ Jinlong Yang,¹ Le Cheng,¹ Tsan-Yu Chiu,^{1,*} and Huan Liu^{1,5,*}

SUMMARY

***Alstonia scholaris* of the Apocynaceae family is a medicinal plant with a rich source of bioactive monoterpenoid indole alkaloids (MIAs), which possess anti-cancer activity like vinca alkaloids. To gain genomic insights into MIA biosynthesis, we assembled a high-quality chromosome-level genome for *A. scholaris* using nanopore and Hi-C data. The 444.95 Mb genome contained 35,488 protein-coding genes. A total of 20 chromosomes were assembled with a scaffold N50 of 21.75 Mb. The genome contained a cluster of strictosidine synthases and tryptophan decarboxylases with synteny to other species and a saccharide-terpene cluster involved in the monoterpenoid biosynthesis pathway of the MIA upstream pathway. The multi-omics data of *A. scholaris* provide a valuable resource for understanding the evolutionary origins of MIAs and for discovering biosynthetic pathways and synthetic biology efforts for producing pharmaceutically useful alkaloids.**

INTRODUCTION

The *Alstonia scholaris* from Apocynaceae family is commonly known as milkwood pine, blackboard tree, or devil tree and is widely distributed in the tropical regions of Africa and Asia.¹ Since ancient times, the use of natural products from terrestrial plants has been indispensable to humans in each civilization.² For instance, the bark of *A. scholaris* is used in traditional medicine in South and Southeast Asia to treat dysentery and malaria.³ The traditional system of Indian medicine named Ayurveda uses the bark in numerous compound formulations, including mahatikta ghrita, saptachchhadadi taila, saptaparnaghana vati, and saptachchhadadi kvatha⁴ (*A. scholaris* named Saptaparna or Saptaparn in Sanskrit language). The leaves of these plants are used in “Dai” ethnopharmacology to treat chronic respiratory diseases in Yunnan Province of China.⁵ Based on traditional utilization, the leaf extract has also been industrialized as an OTC (over the counter) drug in China, popularly known as “Deng-Tai-Ye” tablet.⁶ Additionally, it is also used to treat chronic bronchitis cough and was approved by the China Food and Drug Administration (CFDA).⁷ In addition, the extracts of *A. scholaris* have been shown to have anti-diabetic,⁸ anti-inflammatory,⁹ anti-tussive, anti-asthmatic,¹⁰ and, most importantly, anti-tumor activities.¹¹

Members of the Apocynaceae family serve as the major natural source for procuring monoterpenoid indole alkaloids (MIAs), which can be used to treat various human diseases. For example, catharanthine and vindoline from *Catharanthus roseus* can be used for diabetes treatment. Catharanthine, vinorelbine, and vincristine are currently being used for anti-cancer treatment.¹² Ajmalicine from *Rauvolfia verticillate* has neurological function and hypotensive effects.¹³ Camptothecin from *Camptotheca acuminata* also has anti-cancer effects.¹⁴ MIAs have a wide range of diverse and important pharmacological properties, some of which have been used clinically. The MIA biosynthesis pathways from *C. roseus* (vinblastine and vincristine)^{15–18} and *R. serpentina* (reserpine)¹⁹ were characterized at the molecular level. Several studies have also been carried out on *C. acuminata*²⁰ and *Ophiorrhiza pumila*²¹ to characterize the early steps of camptothecin biosynthesis.²²

Akuammiline alkaloids are a class of MIAs in *A. scholaris*, and more than 300 compounds with various pharmacological activities have been identified in *A. scholaris*.^{23,24} For example, echitamines exhibit both *in vitro* and *in vivo* cytotoxicity,²⁵ while strictamines²⁶ inhibit the transcription factor nuclear factor κ B (NF- κ B).²⁷ Additionally, the renal cortex protein SGLT2 is inhibited by the derivatives of picaline,^{28–30} whereas aspidophylline A reverses drug resistance in cancerous cell lines.³¹ Biogenetically, the akuammiline alkaloids are derived from geissoschizine, a key intermediate in the biosynthetic pathway of MIAs.³² The formation of strictosidine from secologanine and tryptamine, catalyzed by the enzyme strictosidine synthase (STR),³³ and by the presence of two enzymes function, strictosidine- β -D-glucosidase (SGD)^{34–36} and

¹State Key Laboratory of Agricultural Genomics, Key Laboratory of Genomics, Ministry of Agriculture, BGI Research, Shenzhen 518083, China

²BGI Research, Wuhan 430074, China

³Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin 150081, China

⁴These authors contributed equally

⁵Lead contact

*Correspondence: qiucanyu@genomics.cn (T.-Y.C.), liuhuan@genomics.cn (H.L.)

<https://doi.org/10.1016/j.isci.2024.109599>



Table 1. Statistics of the genome assembly

Assembly		<i>Alstonia scholaris</i>
Genome-sequencing depth (~124x)	Nanopore sequencing (Gb)	62
	Hi-C (Gb)	69
Estimated genome size (Mb)		489
Estimated heterozygosity (%)		0.835
Assembly size (Mb)		445
GC content (%)		34.73
Scaffold N50 (Mb)		13.244
BUSCO completeness of assembly (%)		98.3
Total length of pseudochromosome assembly (Mb)		445
Pseudochromosome number		20
Scaffold N50 of pseudochromosome assembly (Mb)		21.753
BUSCO completeness of pseudochromosome assembly (%)		98.3
The rate of pseudochromosome anchored genome (%)		99.9

geissoschizine synthase (GS),¹⁸ results in the production of geissoschizine. The intramolecular oxidative coupling between C7 and C16 of geissoschizine results in the formation of the framework of akuammiline.³² Moreover, this coupling forms the caged indolenine framework of (+)-rhazimal.³²

Elucidating the biosynthetic pathway of bioactive compounds will greatly benefit the development of synthetic biology tools for medicinal plants. The advancement of long-read sequencing has closed the gap in genomic information provided by short-read sequencing.^{20,37} Although there are several existing studies on various medicinal plants,^{38–42} the biosynthetic pathway of akuammiline alkaloids is still vague due to the lack of valid omics data. Therefore, these high-quality genome and transcriptome data of *A. scholaris* provide a solid foundation for identifying potential genes involved in the akuammiline alkaloid production pathway and advancing synthetic biology research on anti-cancer bioactives from *A. scholaris*.

RESULTS

De novo genome assembly and pseudochromosome construction

We used approximately 45 Gb (~90x) of short reads for genome survey analysis, and the estimated genome size was 489 Mb based on k-mer analysis (Figure S1). Using a combination of 295 Gb (~590x) short and 62 Gb (~124x) long nanopore reads, we generated an assembly of 444,958,049 bp with a contig N50 size of 13.24 Mb (Table 1). We anchored the contig-level genome onto 20 pseudochromosomes with 69 Gb (~138x) of Hi-C (high-throughput/resolution chromosome conformation capture) data (Figure 1). The N50 value increased to 21.75 Mb (Table 1), and the length of the chromosomes ranged from 17.02 Mb to 29.20 Mb. The BUSCO (benchmarking universal single-copy orthologs)⁴³ results showed that 2,286 out of 2,326 plant BUSCOs (98.3%) could be found in both the contig and chromosome-level genome assemblies (Figure S2; Table S1).

Protein-coding gene prediction and functional annotation

We found 38.26% repetitive elements in the *A. scholaris* genome. The most abundant type was long terminal repeats (LTRs), accounting for 28.81% of the *A. scholaris* genome. DNA class repeat, LINE (long interspersed nuclear elements), and SINE (short interspersed nuclear elements) classes accounted for 5.03%, 2.21%, and 0.01%, respectively, of this genome (Table S2). A total of 35,488 genes with 5.64 exons per gene on average were predicted by combining three methods, namely *de novo*, homology, and transcriptome-based methods. The average lengths of the mRNAs, exons, and introns were 3,852 bp, 217 bp, and 565 bp, respectively (Table S3). The length distributions of the gene sets of *A. scholaris* and the other seven species (*C. gigantea*, *C. canephora*, *C. roseus*, *G. sempervirens*, *N. tabacum*, *S. lycopersicum*, and *O. pumila*) are shown in Figure S3. The complete and single-copy genes accounted for 92.6% of the predicted gene set (Figure S2; Table S1).

The functional annotation results revealed that approximately 96.70% of the genes had a conserved motif or homolog match in at least one of the public databases, including Swiss-Prot (78.16%), InterPro (93.53%), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (75.34%). For the non-coding RNAs, we also identified 142 microRNAs, 621 tRNAs, 135 rRNAs, and 829 small nuclear RNAs in the *A. scholaris* genome (Table S4).

Comparative genomic analysis

We compared the *A. scholaris* genome with 15 other sequenced genomes and identified 14,289 gene families. A total of 107 expanded gene families and 42 contracted families were significantly differentially expressed in *A. scholaris*. The 16 species family numbers and gene numbers are summarized in Figure 2A and Table S5. Gene ontology (GO) enrichment analysis of the 107 significantly expanded gene families revealed

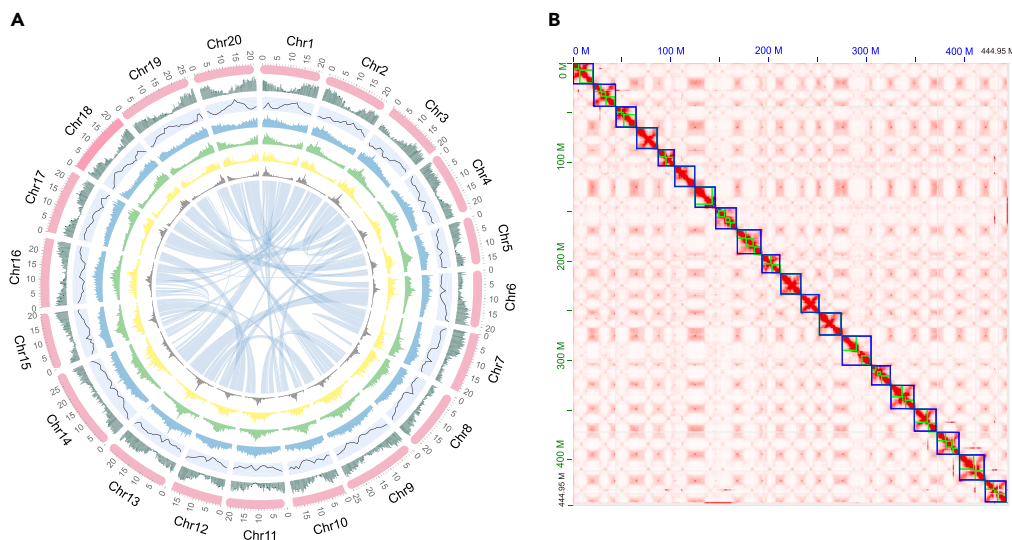


Figure 1. Genome information and morphological features of *A. scholaris*

(A) Characteristics of the 20 chromosomes of *A. scholaris*. The tracks from the outer to the inner regions of the circle individually represent the length of chromosomes (pink), gene numbers (dark green), the content of GC (black line), repeat sequences (blue), LTRs (green), LTR *Copia* (yellow), and LTR *Gypsy* (gray), and the links inside the circle show syntenic collinearity.

(B) Hi-C plot of the pseudochromosome-level assembly of *A. scholaris* genome. The axis refers to the genome size, and each blue box represents one chromosome.

25 GO terms. A portion of the expanded genes were enriched in binding terms in the molecular function category, such as ion binding (38), and organic cyclic compound binding (36). The other genes were enriched in several enzyme activity terms, including oxidoreductase activity (25), monooxygenase activity (14), and protein kinase activity (12) (Table S6). A total of seven expanded genes were located in the monoterpenoid biosynthesis pathway; three of these genes were annotated as 10 HGOs (ALSSCH34014, ALSSCH34015, and ALSSCH34016), and two of them are candidate G10Hs (ALSSCH04576 and ALSSCH08114) (Table S7). Additionally, these significantly contracted gene families were enriched in 69 GO categories and 35 pathways. For example, a total of 164 genes were enriched in the plant-pathogen interaction pathway, and 83 genes were enriched in the phenylpropanoid biosynthesis pathway (Tables S8 and S9).

A total of 102 single-copy orthologous groups were used for constructing a phylogenetic tree to estimate the divergence times of 16 plants. We found that *A. scholaris* was clustered with the Apocynaceae species group, which was separated approximately 67 million years ago from *C. gigantea*, while *R. serpentina* and *C. roseus* were closest and diverged almost 39 million years ago. These two species diverged 54 million years ago from *R. stricta* (Figure 2B).

Two whole-genome duplications (WGDs) shaped *A. scholaris* evolution

Ancient WGD events have contributed to plant adaptation and are prevalent in plants.⁴⁴ In our study, we used the k_s (a synonymous substitution per synonymous site) value to determine whether the *A. scholaris* genome had undergone WGD. We found a peak between k_s values of 0.3, indicating that a WGD event occurred approximately 35.1 million years ago (Figures 3A and 3B), which was later than the divergence time between *A. scholaris* and other Apocynaceae plants. We also performed a synteny analysis of the *A. scholaris* genes using MCScanx to confirm the collinearity relationship. We detected 4,543 syntenic blocks across the whole genome, including 26,485 genes (74.63%). Furthermore, a k_s plot of the paralogs of *C. canephora*, *C. roseus*, *O. pumila*, and *V. vinifera* confirmed that these species underwent WGD in accordance with previous reports (Figure 3A). Synteny analysis revealed 1:2 syntenic depth ratios in both *A. scholaris*-*C. roseus* and *A. scholaris*-*O. pumila* comparisons (Figures 3C and S4), which suggested that two WGD events occurred during the evolution of *A. scholaris*.

Gene clusters involved in MIAs biosynthesis in *A. scholaris*

Secologanin synthase (SLS) and STR, which were identified in Gentianales and catalyze the synthesis of strictosidine, were discovered in a previous study of *O. pumila* and demonstrated significant importance in enabling the evolution of novel enzymes for MIA biosynthesis and diversification.²¹ In our study, we used more species from Gentianales and found several STR copies in MIA-producing plants, including *A. scholaris*, *C. roseus*, *R. serpentina*, *C. gigantea*, *O. pumila*, *R. stricta*, and *G. sempervirens*. However, no STR was found in *Amborella trichopoda*, *Oryza sativa*, *Solanum tuberosum*, or *Sorghum bicolor* genomes. Phylogenetic analysis of the STRs revealed an MIA-specific plant gene family (Group I) that included previously functionally characterized STRs involved in the MIA biosynthesis pathway and two AsSTRs identified in our study. Another two AsSTRs (ALSSCH12919 and ALSSCH22548) clustered with *O. pumila*, *G. sempervirens*, *C. roseus*, and *R. serpentina* (Group II) (Figure 4A). Therefore, AsSTRs in Group I are more likely to have true STR activity. Groups III and IV contained homologous gene

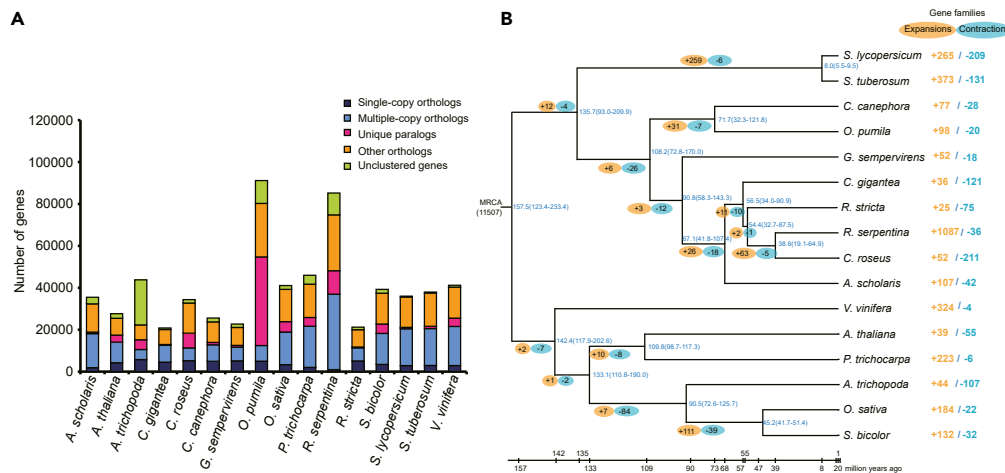


Figure 2. Gene family analysis and phylogenetic tree construction

(A) Bar chart of the orthology numbers in these 16 species.

(B) Phylogenetic tree showing the sizes of significantly expanded and contracted gene families. The branch labels in yellow and blue represent the significantly expanded and contracted gene families (p value < 0.05), respectively, of each node. The right column shows significantly expanded and contracted gene families of individual species. Furthermore, the statistical method of enrichment analysis is χ^2 test. AdjustedPv is a corrected p value that is obtained by performing false discovery rate (FDR) testing on p values. AdjustedPv obtained from significance tests are generally considered to have significant statistical differences when AdjustedPv < 0.05 .

from *P. trichocarpa*, and *S. lycopersicum*, *V. vinifera*, respectively. However, SLS was almost expanded in all the MIA-producing plants and the other ten non-MIA-producing plants (Table S10). Tryptophan decarboxylase (TDC) also plays an essential role in strictosidine biosynthesis, and we found that TDC was expanded not only in all MIA-specific plants but also in other plants. Phylogenetic analysis of TDC genes showed three branches that included MIA-producing plants. Group I included four candidate TDC genes of *A. scholaris* and TDC1 and TDC2 from *C. acuminata*, one TDC from *O. pumila*, one TDC from *G. sempervirens*, also one TDC from *C. roseus*. In addition, Group I also includes other non-MIA-producing plants, such as *C. canephora* and *S. lycopersicum*. Group II and III consisted of genes from MIA-producing plants (*A. scholaris*, *C. roseus*, *G. sempervirens*, and *O. pumila*), and they included genes from non-MIA-producing plants (Figure 4B).

The four candidate AsTDCs exhibited collinearity with TDCs from both *C. roseus* and *O. pumila*. However, only one AsTDC from *O. pumila* displayed collinearity with TDC2. Furthermore, a single AsSTR showed collinearity with both STR_CRO and STR_OPU. AsTDCs and AsSTRs were located on chromosome 3, suggesting the possibility of forming a gene cluster (Figure 4C). In addition, a saccharide-terpene cluster on chromosome 8 was identified in our study; this cluster included seven terpene synthases, 11 glycosyltransferase synthases, one coenzyme A (CoA)-ligase, and 16 other genes (Figure 4D).

We compared the expression levels of the various genes in different tissues (Figure S5). The upregulated genes in both the leaf and branch, compared to those in the control (trunk bark), were predominantly enriched in GO terms related to the membrane, oxidoreductase activity, transported activity, and transmembrane transporter activity. The genes whose expression was significantly greater in the petiole than in the trunk bark were mostly enriched in metabolic processes, cellular metabolic processes, biosynthetic processes, and organic substance biosynthetic processes. More than two hundred genes were also enriched in the membrane and oxidoreductase activity terms. The results of the KEGG enrichment analysis showed that the metabolic pathways, photosynthesis pathway, and photosynthesis-antenna proteins pathway were more highly expressed in the leaves, branches, and petioles than in the trunk bark.

Additionally, compared with those in the branches, the upregulated genes in the trunk bark were enriched in the flavonoid biosynthesis pathway. Moreover, nine genes with higher expression levels in trunk bark than in leaf were enriched in the monoterpenoid biosynthesis pathway. Five (ID: ALSSCH22817, ALSSCH22820, ALSSCH22827, ALSSCH22830, and ALSSCH22834) of the nine genes were contained in the saccharide-terpene cluster mentioned earlier. Similarly, the upregulated genes in the petiole (compared to those in the branches) were enriched in the flavonoid biosynthesis pathway, as well as the monoterpenoid biosynthesis pathway. A total of six terpene synthases (ID: ALSSCH22817, ALSSCH22823, ALSSCH22827, ALSSCH22830, ALSSCH22834, and ALSSCH22837) were located on the saccharide-terpene cluster (Table S11). In particular, AsGESs (ALSSCH22817, ALSSCH22830, and ALSSCH22834) act on geranyl-FP to produce geraniol in the monoterpenoid biosynthesis pathway. These enzymes are located on the saccharide-terpene cluster. This cluster is the first terpene gene cluster in MIA-producing species. The results of all the GO and KEGG enrichment analyses between different tissues of *A. scholaris* are individually summarized in Tables S12–S32.

Candidate genes of the monoterpenoid indole alkaloid biosynthetic pathway

Previous studies have reported the composition and distribution of MIAs (picrinine, picralinal, echitamine, and akuammidine) in the leaves, flowers, trunk barks, and fruits of *A. scholaris*.⁴⁵ MIAs are a large group of plant-produced natural products of which more than 3,000 have been

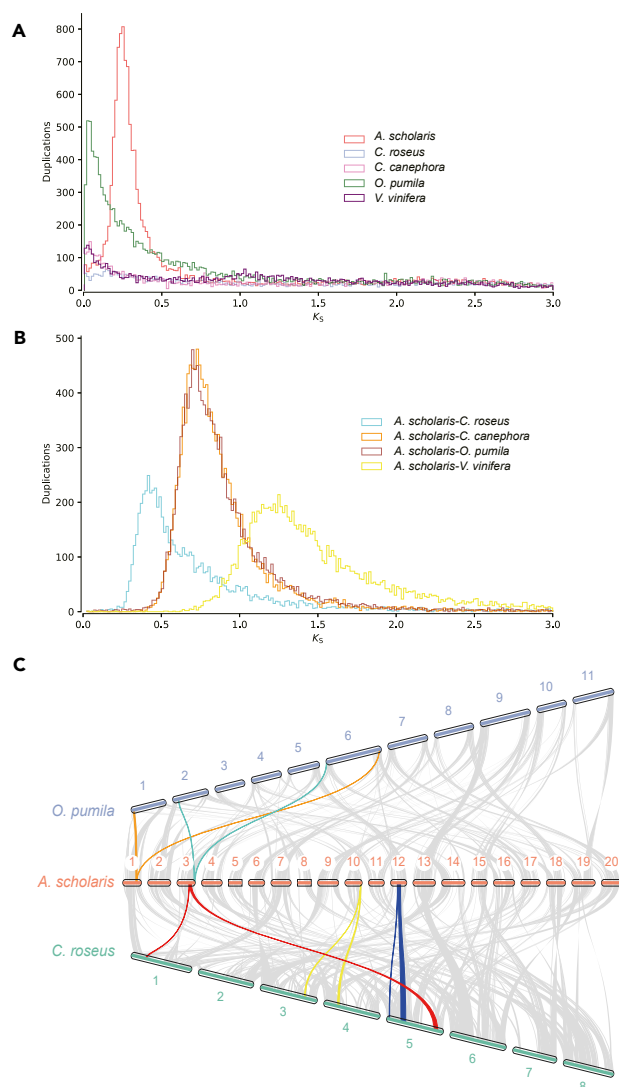


Figure 3. The analysis of whole-genome duplication in *A. scholaris*

(A) The distribution of synonymous substitution rate (K_s) distances observed for paralogs from *A. scholaris*, *C. canephora*, *C. roseus*, *O. pumila*, and *V. vinifera*. (B) The distribution of K_s values of orthologs between *A. scholaris* and the previously mentioned species.

(C) Synteny between genomic regions in *A. scholaris*, *C. roseus*, and *O. pumila*. The gray lines highlight major syntenic blocks spanning the genomes. The colored lines represent examples of syntenic genes found in two species that correspond to one copy in *A. scholaris*, and two in *C. roseus* and *O. pumila*.

identified,⁴⁶ mostly in Gentianale.⁴⁷ Additionally, a review paper summarized 444 monoterpene indole alkaloids that were reported from six genera of the Apocynaceae family between 2010 and 2020.⁴⁸ In this study, the extracted metabolites of *A. scholaris* leaves, trunk barks, and branches were compared with the ionic fragments and separation times of secologanin and tryptamine standards purchased from a certified vendor. The results showed that the same ionic fragments as the standard were found at the same separation time, which indicated the presence of two key precursors required for MIA biosynthesis (Figure S6).

The biosynthesis pathway of alkaloids from *A. scholaris* has not been elucidated. These pathways start with the common precursor stricatosidine, which undergoes several steps of reaction to form Rhazimal akuammiline (Figure 5A). Hence, we focused on identifying potential enzyme-coding genes involved in the akuammiline biosynthesis pathway. Initially, we compiled a preliminary gene list by aligning sequences with known genes from MIA biosynthesis pathways, and we filtered out genes with low or no expression. Furthermore, we screened candidate genes using qualitative protein data of multi-tissues of *A. scholaris*, resulting in the identification of 55 candidate genes in the *A. scholaris* genome (Table S33). The accession numbers, names and classification of all known MIA biosynthesis-related enzymes are summarized in Table S34. The processed protein group data are presented in Table S35.

The results of the co-expression analysis indicated strong correlations between modules and specific plant parts: the trunk bark (skyblue, darkmagenta, bisque4, darkorange), the petiole (darkorange2 and maroon), and the leaf (darkgreen and red) (Figure S7).

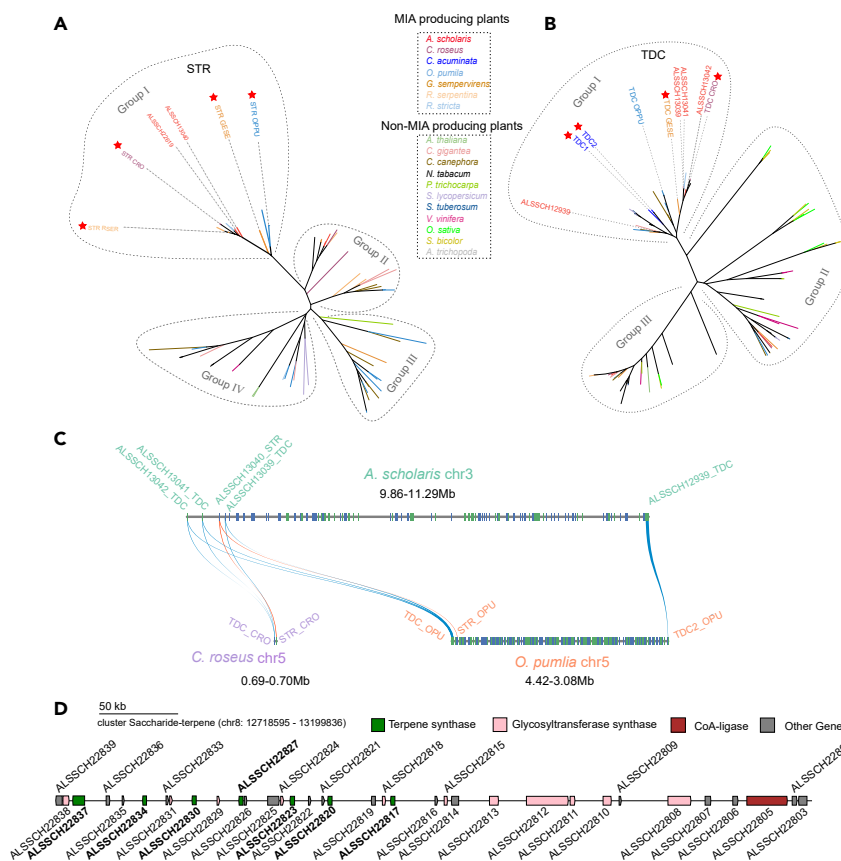


Figure 4. Key genes involved in strictosidine biosynthesis

(A and B) Maximum likelihood phylogenetic tree based on candidate STR and TDC gene families from these 18 species. “**” represents functionally characterized genes. Each species is represented by a different color.

(C) A gene cluster located on the third chromosome of the *A. scholaris* genome. The blue and orange lines show syntenic blocks of AsTDCs and AsSTRs with TDC and STR from *C. roseus* and *O. pumilia*, respectively.

(D) The saccharide-terpene cluster. The bolded gene IDs represent genes that are significantly highly expressed in petioles and trunk barks and are also genes that are enriched in the monoterpene biosynthesis pathway.

Within the skyblue, darkmagenta, bisque4, and darkorange modules, we identified the presence of AsGES, As10HGO, As7-DLH, AsTDC, AsSLS, and AsGS, while, in the darkgreen and red modules, we observed AsAS, AsIGPS, AsTSA, AsTSB, AsGO, AsIS, As7-DLH, and AsSLS. We also found AsIS and AsGO in the darkorange2 and maroon modules. These findings suggest a potential co-expression pattern of alkaloid biosynthesis genes in *A. scholaris*.

In addition, we compared the expression levels of those genes in the leaf, petiole, branch, and trunk bark tissues of *A. scholaris*. We found that 7DLGT/UGT6, TSB2, TSB3, TDC1, and TDC2 were more highly expressed in the petioles of *A. scholaris*. AsGOs were highly expressed in trunk bark and petioles. 7-DLH, PAT1, G10H, IGPS, IO, SGD, SLS, TSA, and TSB1 exhibited relatively average expression levels in all tissues, exhibiting a co-expression pattern in different modules (Figure 5B).

DISCUSSION

MIAs are natural compounds derived from secologanin and tryptamine that are subsequently obtained from tryptophan through decarboxylation. One of the most comprehensively elucidated MIA biosynthesis pathways is the vinca alkaloid biosynthesis pathway found in *C. roseus*, which leads to the production of compounds like vincristine, vinblastine, catharanthine, tarbersonine, and vindoline.^{15,46,49–52} A previous study has also identified various MIAs in different parts of *A. scholaris*.⁵³ Here, we assembled a chromosome-level genome of *A. scholaris*, an MIA-producing plant with multiple medicinal benefits. We detected the presence of the MIA precursor, secologanin, and tryptamine in various parts of *A. scholaris*. By performing an alignment with previously identified enzymes from MIA biosynthesis pathways, a series of candidate genes involved in MIA biosynthesis were identified in the *A. scholaris* genome. *A. scholaris* contains several unique alkaloids, such as 19-epi-scholaricine, scholaricine, 19,20-Z-vallesamine, and picrinine, which are the main medicinal components of the “Deng-Tai-Ye” tablet.⁵⁴ Because there is no established commercial standard for these alkaloids, detecting metabolic differences across multiple tissues in *A. scholaris* is challenging.

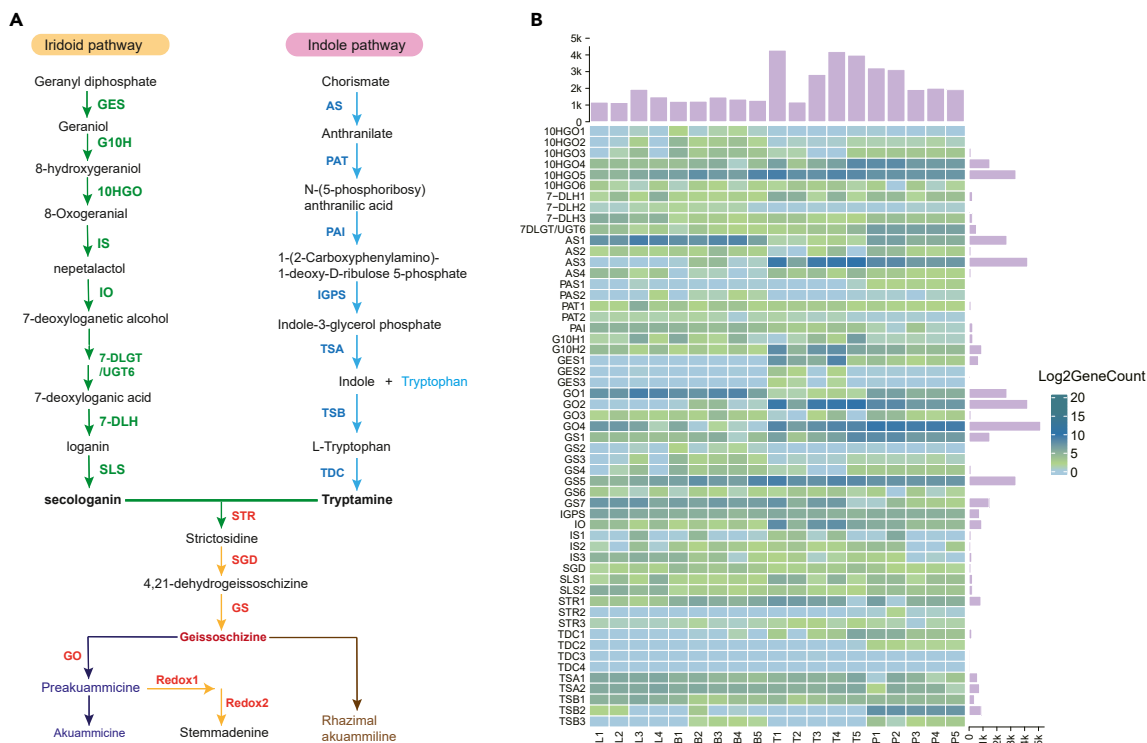


Figure 5. Biosynthesis pathway of akuammiline alkaloid and the expression levels of candidate enzymes in the pathway

(A) The akuammiline biosynthesis pathway.

(B) The expression levels of candidate genes in the *A. scholaris* genome. The abscissa of each heatmap indicates the different tissues of *A. scholaris*. B: branch, P: petiole, T: trunk bark, L: leaf. The number represents duplication. Log2GeneCount refers to the normalization of FPKM (fragments per kilobase million) values by the log method of each row. Different expressed genes match adjusted p value < 0.05 and a $|\log_2\text{FoldChange}| \geq 2$. Dark-blue color indicates a high expression level, and blue indicates a low expression level.

The biosynthesis pathway of camptothecin is similar to that of vinblastine/vincristine in *C. roseus* and involves the production of loganic acid.^{55–57} However, secologanic acid is transformed to strictosidinic acid by STRAS in *C. acuminata*, which has different modifications than *C. roseus*.²⁰ In *O. pumila*, STR has the same function as in *C. roseus* and similarly produces strictosine.²¹ We used additional Gentianales species for comparison with the *A. scholaris* genome and found STR expansion in MIA-specific plants, but not in non-MIA-producing plants. These findings indicate that STR is conserved in the MIA biosynthesis pathway. However, SLS was retained in all the species rather than in the MIA-producing plants. Besides, the existence of an alternate pathway for MIA biosynthesis in *C. acuminata* has been proven to occur through strictosidinic acid, which is synthesized by the condensation of secologanic acid with tryptamine by SLAS, an SLS-like enzyme,^{21,55} suggesting that SLS is not necessary for all MIA-producing plants. TDC also plays an essential role in strictosidine biosynthesis and was found to be expanded in all MIA-specific plants but also in other plants. The phylogenetic analysis of TDC genes showed a branch that included both MIA-specific plants and non-MIA-producing plants, suggesting an essential role of TDC in amino acid metabolism in plants.²¹

A recent study identified an STR-TDC cluster in the *C. roseus* v3 genome.⁵⁸ In our study, we discovered a gene cluster on the third chromosome of the *A. scholaris* genome that combines AsTDCs and one AsSTR. These genes exhibited collinearity with the TDC and STR genes from *C. roseus* and *O. pumila*, respectively. In addition, we also found a saccharide-terpene cluster on the eighth chromosome of the *A. scholaris* genome. This terpene cluster included seven terpene synthases that had higher expression levels in the trunk bark and petioles of *A. scholaris*. Taken together, these findings show that the monoterpenes of the MIA upstream pathway may be synthesized or transferred to the trunk bark and petiole. This cluster also included three AsGESs which catalyze the first step in the monoterpene synthesis pathway.

To maximize the value of our genomic data in the akuammilan alkaloid synthetic pathway analysis of *A. scholaris*, we published the genomic data of short reads in advance. A study reported the discovery of a series of new enzymes involved in akuammilan alkaloid biosynthesis by using our publicly available *A. scholaris* genome data. Among these enzymes, AsRHS and AsGO share a significant sequence identity of 62.4%. Notably, the amino acid residue at position 372 plays a crucial role in regulating the geissoschizine reaction by altering the distance between C-2 and C-7 in relation to the heme. Consequently, one enzyme may predominantly oxidize C2 (GO), while the other may target C7 (RHS) of geissoschizine.⁵⁹ We analyzed to investigate the co-expression patterns of these genes, revealing several modules that

exhibited strong correlations with the leaf, trunk bark, and petiole parts of *A. scholaris*. Among these modules, we identified candidate MIA genes, suggesting a potential co-expression pattern for the alkaloid biosynthesis genes in *A. scholaris*. Overall, our findings contribute to a deeper understanding of *A. scholaris* and pave the way for innovative applications in the field of alkaloid biosynthesis, setting the stage for exciting future research in this domain. In addition, our data are conducive to revealing the mechanism of MIA evolution.

Limitations of the study

While the current study identified putative candidate genes implicated in the MIA biosynthetic pathway, functional characterization of these genes in heterologous hosts such as tobacco or *E. coli* was not performed. The incorporation of metabolomic data could further elucidate the MIA landscape in *A. scholaris*. Despite the lack of reference standards for most *A. scholaris* MIAs, future investigations could employ total ion chromatograms coupled with mass spectrometric analysis to tentatively annotate major alkaloid peaks based on their *m/z* values.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability statements
 - Data and code availability
- METHOD DETAILS
 - Plant sample collection and sequencing
 - Genome assembly and chromosome anchoring
 - Identification of repetitive sequences
 - RNA-seq analysis
 - Gene model prediction and functional annotation
 - Candidate MIA gene prediction
 - Gene family analysis and evolutionary tree construction
 - Protein detection
 - Metabolite detection
- QUANTIFICATION AND STATISTICAL ANALYSES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2024.109599>.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (no. 2019YFC1711000) and Shenzhen-Hong Kong-Macao Science and Technology Innovation Project (Category C) (ref no.: EF038/ICMS-LMY/2021/SZSTIC). This work is part of the 10KP project (<https://db.cngb.org/10kp/>). This work is also supported by the high-level talent training support plan of Yunnan Province to L.C. (2020) and China National GeneBank (CNGB; <https://www.cngb.org/>).

AUTHOR CONTRIBUTIONS

H.C., L.C., and H.L. designed the study and all the experiments. H.C., S.K.S., and T.-Y.C. performed the data analysis. H.C., S.W., and J.L. collected samples and did metabolome analysis. H.C., T.-Y.C., and S.K.S. wrote the manuscript. All the authors have read and agreed to the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 17, 2023

Revised: January 25, 2024

Accepted: March 25, 2024

Published: March 27, 2024

REFERENCES

- Baliga, M.S. (2012). Review of the phytochemical, pharmacological and toxicological properties of *Alstonia scholaris* Linn. R. Br (Saptaparna). *Chin. J. Integr. Med.* <https://doi.org/10.1007/s11655-011-0947-0>.
- Richardson, M.A., Sanders, T., Palmer, J.L., Greisinger, A., and Singletary, S.E. (2000). Complementary/alternative medicine use in a comprehensive cancer center and the implications for oncology. *J. Clin. Oncol.* **18**, 2505–2514. <https://doi.org/10.1200/jco.2000.18.13.2505>.
- Salim, A.A., Garson, M.J., and Craik, D.J. (2004). New Indole Alkaloids from the Bark of *Alstonia scholaris*. *J. Nat. Prod.* **67**, 1591–1594. <https://doi.org/10.1021/np0498612>.
- Pagariya, A., Jain, N., and Mahajan, M.P. (2020). Saptaparni a Traditional Medicinal Plant—a Concise Review. *Int. J. Biol. Pharm. Allied Sci.* **9**, 828–838. <https://doi.org/10.31032/ijbps/2020/9.4.5038>.
- Khyade, M.S., Kasote, D.M., and Vaikos, N.P. (2014). *Alstonia scholaris* (L.) R. Br. and *Alstonia macrophylla* Wall. ex G. Don: A Comparative Review on Traditional Uses, Phytochemistry and Pharmacology. *J. Ethnopharmacol.* **153**, 1–18. <https://doi.org/10.1016/j.jep.2014.01.025>.
- Cai, X.-H., Du, Z.-Z., and Luo, X.-D. (2007). Unique Monoterpenoid Indole Alkaloids from *Alstonia scholaris* (L.) R. Br. *1817–1820*. <https://doi.org/10.1021/ol0705301>.
- Yang, Z., Sun, L., Liang, C., Xu, Y., Cao, J., Yang, Y., and Gu, J. (2016). Simultaneous quantitation of the diastereoisomers of scholarisine and 19-epischolarisine, vallesamine, and picirnine in rat plasma by supercritical fluid chromatography with tandem mass spectrometry and its application to a pharmacokinetic study. *J. Sep. Sci.* **39**, 2652–2660. <https://doi.org/10.1002/jssc.201600243>.
- El-Askary, H.I., El-Olemy, M.M., Salama, M.M., Sleem, A.A., and Amer, M.H. (2012). Bioguided isolation of pentacyclic triterpenes from the leaves of *Alstonia scholaris* (Linn.) R. Br. growing in Egypt. *Nat. Prod. Res.* **26**, 1755–1758. <https://doi.org/10.1080/14786419.2011.608848>.
- Shang, J.H., Cai, X.H., Feng, T., Zhao, Y.L., Wang, J.K., Zhang, L.Y., Yan, M., and Luo, X.D. (2010). Pharmacological evaluation of *Alstonia scholaris*: anti-inflammatory and analgesic effects. *J. Ethnopharmacol.* **129**, 174–181. <https://doi.org/10.1016/j.jep.2010.02.011>.
- Shang, J.H., Cai, X.H., Zhao, Y.L., Feng, T., and Luo, X.D. (2010). Pharmacological evaluation of *Alstonia scholaris*: anti-tussive, anti-asthmatic and expectorant activities. *J. Ethnopharmacol.* **129**, 293–298. <https://doi.org/10.1016/j.jep.2010.03.029>.
- Jagetia, G.C., and Baliga, M.S. (2006). Evaluation of anticancer activity of the alkaloid fraction of *Alstonia scholaris* (Saptaparna) *in vitro* and *in vivo*. *Phytother. Res.* **20**, 103–109. <https://doi.org/10.1002/ptr.1810>.
- Almagro, L., Fernández-Pérez, F., and Pedreño, M.A. (2015). Indole alkaloids from *Catharanthus roseus*: bioproduction and their effect on human health. *Molecules* **20**, 2973–3000. <https://doi.org/10.3390/molecules20022973>.
- Liu, W., Chen, R., Chen, M., Zhang, H., Peng, M., Yang, C., Ming, X., Lan, X., and Liao, Z. (2012). Tryptophan decarboxylase plays an important role in ajmalicine biosynthesis in *Rauvolfia verticillata*. *Planta* **236**, 239–250. <https://doi.org/10.1007/s00425-012-1608-z>.
- Lorence, A., and Nessler, C.L. (2004). Camptothecin, over four decades of surprising findings. *Phytochemistry* **65**, 2735–2749. <https://doi.org/10.1016/j.phytochem.2004.09.001>.
- Qu, Y., Easson, M.E.A.M., Simionescu, R., Hajicek, J., Thamm, A.M.K., Salim, V., and De Luca, V. (2018). Solution of the multistep pathway for assembly of corynanthean, strychnos, iboga, and aspidosperma monoterpenoid indole alkaloids from 19E-geissoschizine. *Proc. Natl. Acad. Sci. USA* **115**, 3180–3185. <https://doi.org/10.1073/pnas.1719979115>.
- Caputi, L., Franke, J., Farrow, S.C., Chung, K., Payne, R.M.E., Nguyen, T.-D., Dang, T.-T.T., Soares Teto Carqueijeiro, I., Koudounas, K., Dugé de Bernonville, T., et al. (2018). Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science* **360**, 1235–1239. <https://doi.org/10.1126/science.aat4100>.
- Levac, D., Murata, J., Kim, W.S., and De Luca, V. (2008). Application of carborundum abrasion for investigating the leaf epidermis: molecular cloning of *Catharanthus roseus* 16-hydroxytabersonine-16-O-methyltransferase. *Plant J.* **53**, 225–236. <https://doi.org/10.1111/j.1365-313X.2007.03337.x>.
- Tatsis, E.C., Carqueijeiro, I., Dugé de Bernonville, T., Franke, J., Dang, T.T.T., Oudin, A., Lanoue, A., Lafontaine, F., Stavrinides, A.K., Clastre, M., et al. (2017). A three enzyme system to generate the strychnos alkaloid scaffold from a central biosynthetic intermediate. *Nat. Commun.* **8**, 316. <https://doi.org/10.1038/s41467-017-00154-x>.
- Bayer, A., Ma, X., and Stöckigt, J. (2004). Acetyltransfer in natural product biosynthesis—functional cloning and molecular analysis of vinorine synthase. *Bioorg. Med. Chem.* **12**, 2787–2795. <https://doi.org/10.1016/j.bmc.2004.02.029>.
- Kang, M., Fu, R., Zhang, P., Lou, S., Yang, X., Chen, Y., Ma, T., Zhang, Y., Xi, Z., and Liu, J. (2021). A chromosome-level *Camptotheca acuminata* genome assembly provides insights into the evolutionary origin of camptothecin biosynthesis. *Nat. Commun.* **12**, 3531. <https://doi.org/10.1038/s41467-021-23872-9>.
- Rai, A., Hirakawa, H., Nakabayashi, R., Kikuchi, S., Hayashi, K., Rai, M., Tsugawa, H., Nakaya, T., Mori, T., Nagasaki, H., et al. (2021). Chromosome-level genome assembly of *Ophiorrhiza pumila* reveals the evolution of camptothecin biosynthesis. *Nat. Commun.* **12**, 405. <https://doi.org/10.1038/s41467-020-20508-2>.
- Luca, V.D. (2011). Monoterpenoid Indole Alkaloid Biosynthesis. In *Plant Metabolism and Biotechnology*, H. Ashihara, A. Crozier, and A. Komamine, eds., pp. 263–291.
- Reddy, D.S. (2016). Phytochemical analysis of active constituents of *Alstonia scholaris* and their cytotoxicity *in vitro*. *Int. J. Pharmaceut. Sci. Res.* **7**, 3262–3273. [https://doi.org/10.13040/ijpsr.0975-8232.7\(8\).3262-73](https://doi.org/10.13040/ijpsr.0975-8232.7(8).3262-73).
- Zhang, L., Zhang, C.-J., Zhang, D.-B., Wen, J., Zhao, X.-W., Li, Y., and Gao, K. (2014). An unusual indole alkaloid with anti-adenovirus and anti-HSV activities from *Alstonia scholaris*. *Tetrahedron Lett.* **55**, 1815–1817. <https://doi.org/10.1016/j.tetlet.2014.01.122>.
- Jagetia, G.C., Baliga, M.S., Venkatesh, P., Ullloor, J.N., Mantena, S.K., Genebriera, J., and Mathuram, V. (2005). Evaluation of the cytotoxic effect of the monoterpene indole alkaloid echitamine *in-vitro* and in tumour-bearing mice. *J. Pharm. Pharmacol.* **57**, 1213–1219. <https://doi.org/10.1211/jpp.57.9.0017>.
- Schnoes, H.K., Biemann, K., Mokry, J., Kompis, I., Chatterjee, A., and Ganguli, G. (1966). Strictamine. *J. Org. Chem.* **31**, 1641–1642. <https://doi.org/10.1021/jo01343a507>.
- Hou, Y., Cao, X., Wang, L., Cheng, B., Dong, L., Luo, X., Bai, G., and Gao, W. (2012). Microfractionation bioactivity-based ultra performance liquid chromatography/quadrupole time-of-flight mass spectrometry for the identification of nuclear factor-kappaB inhibitors and beta2 adrenergic receptor agonists in an alkaloidal extract of the folk herb *Alstonia scholaris*. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **908**, 98–104. <https://doi.org/10.1016/j.jchromb.2012.10.004>.
- Britten, A.Z., and Smith, G.F. (1963). Akumma Alkaloids. Part VI.* The Reactions of Picraline. *J. Chem. Soc.* **3860**, 3850–3854. <https://doi.org/10.1039/JR9630003850>.
- Arai, H., Hirasawa, Y., Rahman, A., Kusumawati, I., Zaini, N.C., Sato, S., Aoyama, C., Takeo, J., and Morita, H. (2010). Alstiphylanines E–H, picraline and ajmaline-type alkaloids from *Alstonia macrophylla* inhibiting sodium glucose cotransporter. *Bioorg. Med. Chem.* **18**, 2152–2158. <https://doi.org/10.1016/j.bmc.2010.01.077>.
- Meng, W., Ellsworth, B.A., Nirschl, A.A., McCann, P.J., Patel, M., Girotra, R.N., Wu, G., Sher, P.M., Morrison, E.P., Biller, S.A., et al. (2008). Discovery of Dapagliflozin: A Potent, Selective Renal Sodium-Dependent Glucose Cotransporter 2 (SGLT2) Inhibitor for the Treatment of Type 2 Diabetes. *J. Med. Chem.* **51**, 1145–1149. <https://doi.org/10.1021/jm701272q>.
- Subramaniam, G., Hiraku, O., Hayashi, M., Koyano, T., Komiyama, K., and Kam, T.-S. (2007). Biologically Active Aspidofractinine, Rhazinilam, Akumamine, and Vincorine Alkaloids from *Kopsia*. *J. Nat. Prod.* **70**, 1783–1789. <https://doi.org/10.1021/np0703747>.
- Smith, J.M., Moreno, J., Boal, B.W., and Garg, N.K. (2015). Cascade reactions: a driving force in akuammiline alkaloid total synthesis. *Angew. Chem. Int. Ed. Engl.* **54**, 400–412. <https://doi.org/10.1002/anie.201406866>.
- Stöckigt, J., Antonchick, A.P., Wu, F., and Waldmann, H. (2011). The Pictet–Spengler Reaction in Nature and in Organic Chemistry. *Angew. Chem. Int. Ed. Engl.* **50**, 8538–8564. <https://doi.org/10.1002/anie.201008071>.
- Hemscheidt, T., and Zenk, M.H. (1980). Glucosidases involved in indole alkaloid biosynthesis of *Catharanthus* cell cultures. *FEBS Lett.* **110**, 187–191. [https://doi.org/10.1016/0014-5793\(80\)80069-X](https://doi.org/10.1016/0014-5793(80)80069-X).
- Luijendijk, T.J., Stevens, L.H., and Verpoorte, R. (1998). Purification and characterization of strictosidine β-D-glucosidase from *Catharanthus roseus* cell suspension cultures. *Plant Physiol. Biochem.* **36**, 419–425. [https://doi.org/10.1016/S0981-9428\(98\)80205-2](https://doi.org/10.1016/S0981-9428(98)80205-2).
- Barleben, L., Ma, X., Koepke, J., Peng, G., Michel, H., and Stöckigt, J. (2005). Expression, purification, crystallization and preliminary X-ray analysis of strictosidine glucosidase, an

- enzyme initiating biosynthetic pathways to a unique diversity of indole alkaloid skeletons. *Biochim. Biophys. Acta* 1747, 89–92. <https://doi.org/10.1016/j.bbapap.2004.09.026>.
37. Sahu, S.K., and Liu, H. (2023). Long-read sequencing (method of the year 2022): the way forward for plant omics research. *Mol. Plant* 16, 791–793. <https://doi.org/10.1016/j.molp.2023.04.007>.
 38. Guo, X., Fang, D., Sahu, S.K., Yang, S., Guang, X., Folk, R., Smith, S.A., Chanderbali, A.S., Chen, S., Liu, M., et al. (2021). Chloranthus genome provides insights into the early diversification of angiosperms. *Nat. Commun.* 12, 6930. <https://doi.org/10.1038/s41467-021-26922-4>.
 39. Wang, S., Liang, H., Wang, H., Li, L., Xu, Y., Liu, Y., Liu, M., Wei, J., Ma, T., Le, C., et al. (2022). The chromosome-scale genomes of *Dipterocarpus turbinatus* and *Hopea hainanensis* (Dipterocarpaceae) provide insights into fragrant oleoresin biosynthesis and hard wood formation. *Plant Biotechnol. J.* 20, 538–553. <https://doi.org/10.1111/pbi.13735>.
 40. Fan, Y., Sahu, S.K., Yang, T., Mu, W., Wei, J., Cheng, L., Yang, J., Liu, J., Zhao, Y., Lisby, M., and Liu, H. (2021). The *Clausena lansium* (Wampee) genome reveal new insights into the carbazole alkaloids biosynthesis pathway. *Genomics* 113, 3696–3704. <https://doi.org/10.1016/j.ygeno.2021.09.007>.
 41. Fan, Y., Sahu, S.K., Yang, T., Mu, W., Wei, J., Cheng, L., Yang, J., Mu, R., Liu, J., Zhao, J., et al. (2020). Dissecting the genome of star fruit (*Averrhoa carambola* L.). *Hortic. Res.* 7, 94. <https://doi.org/10.1038/s41438-020-0306-4>.
 42. Sahu, S.K., Liu, M., Yssel, A., Kariba, R., Muthemba, S., Jiang, S., Song, B., Hendre, P.S., Muchugi, A., Jamnadass, R., et al. (2019). Draft Genomes of two Artocarpus plants, Jackfruit (*A. heterophyllus*) and Breadfruit (*A. altissilis*). *Genes* 11, 27. <https://doi.org/10.3390/genes11010027>.
 43. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
 44. Wu, S., Han, B., and Jiao, Y. (2020). Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol. Plant* 13, 59–71. <https://doi.org/10.1016/j.molp.2019.10.012>.
 45. Mahar, R., Manivel, N., Kanojjiya, S., Mishra, D.K., and Shukla, S.K. (2022). Assessment of Tissue Specific Distribution and Seasonal Variation of Alkaloids in *Alstonia scholaris*. *Metabolites* 12, 607. <https://doi.org/10.3390/metabo12070607>.
 46. Pan, Q., Mustafa, N.R., Tang, K., Choi, Y.H., and Verpoorte, R. (2016). Monoterpenoid indole alkaloids biosynthesis and its regulation in *Catharanthus roseus*: a literature review from genes to metabolites. *Phytochem. Rev.* 15, 221–250. <https://doi.org/10.1007/s11101-015-9406-4>.
 47. De Luca, V., Salim, V., Levac, D., Atsumi, S.M., and Yu, F. (2012). Discovery and functional analysis of monoterpenoid indole alkaloid pathways in plants. *Methods Enzymol.* 515, 207–229. <https://doi.org/10.1016/B978-0-12-394290-6.00010-0>.
 48. Mohammed, A.E., Abdul-Hameed, Z.H., Alotaibi, M.O., Bawakid, N.O., Sobahi, T.R., Abdel-Lateff, A., and Alarif, W.M. (2021). Chemical Diversity and Bioactivities of Monoterpene Indole Alkaloids (MIAs) from Six Apocynaceae Genera. *Molecules* 26, 488. <https://doi.org/10.3390/molecules26020488>.
 49. Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., Woittiez, L., van der Krol, S., Lugan, R., Ilc, T., et al. (2014). The seco-iridoid pathway from *Catharanthus roseus*. *Nat. Commun.* 5, 3606. <https://doi.org/10.1038/ncomms4606>.
 50. Franke, J., Kim, J., Hamilton, J.P., Zhao, D., Pham, G.M., Wiegert-Rininger, K., Crisovan, E., Newton, L., Vaillancourt, B., Tatsis, E., et al. (2019). Gene Discovery in Gelsemium Highlights Conserved Gene Clusters in Monoterpene Indole Alkaloid Biosynthesis. *Chembiochem* 20, 83–87. <https://doi.org/10.1002/cbic.201800592>.
 51. Nakabayashi, R., Mori, T., Takeda, N., Toyooka, K., Sudo, H., Tsugawa, H., and Saito, K. (2020). Metabolomics with 15N Labeling for Characterizing Missing Monoterpene Indole Alkaloids in Plants. *Anal. Chem.* 92, 5670–5675. <https://doi.org/10.1021/acs.analchem.9b03860>.
 52. Sharma, A., Amin, D., Sankaranarayanan, A., Arora, R., and Mathur, A.K. (2020). Present status of *Catharanthus roseus* monoterpene indole alkaloids engineering in homo- and hetero-logous systems. *Biotechnol. Lett.* 42, 11–23. <https://doi.org/10.1007/s10529-019-02757-4>.
 53. Kaushik, D., Rana, A.C., Kaushik, P., and Sharma, N. (2011). *Alstonia scholaris*: It's Phytochemistry and pharmacology. *Chron. Young Sci.* 2, 71. <https://doi.org/10.4103/2229-5186.82970>.
 54. Pandey, K., Shevkar, C., Bairwa, K., and Kate, A.S. (2020). Pharmaceutical perspective on bioactives from *Alstonia scholaris*: ethnomedicinal knowledge, phytochemistry, clinical status, patent space, and future directions. *Phytochem. Rev.* 19, 191–233. <https://doi.org/10.1007/s11101-020-09662-z>.
 55. Sadre, R., Magallanes-Lundback, M., Pradhan, S., Salim, V., Mesberg, A., Jones, A.D., and DellaPenna, D. (2016). Metabolite diversity in alkaloid biosynthesis: a multilane (diastereomer) highway for camptothecin synthesis in *Camptotheca acuminata*. *Plant Cell* 28, 1926–1944. <https://doi.org/10.1105/tpc.16.00193>.
 56. Murata, J., Roepke, J., Gordon, H., and De Luca, V. (2008). The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell* 20, 524–542. <https://doi.org/10.1105/tpc.107.056630>.
 57. Salim, V., Yu, F., Altarejos, J., and De Luca, V. (2013). Virus-induced gene silencing identifies *Catharanthus roseus* 7-deoxyloganic acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis. *Plant J.* 76, 754–765. <https://doi.org/10.1111/tpj.12330>.
 58. Li, C., Wood, J.C., Vu, A.H., Hamilton, J.P., Rodriguez Lopez, C.E., Payne, R.M.E., Serna Guerrero, D.A., Gase, K., Yamamoto, K., Vaillancourt, B., et al. (2023). Single-cell multiomics in the medicinal plant *Catharanthus roseus*. *Nat. Chem. Biol.* 19, 1031–1041. <https://doi.org/10.1038/s41589-023-01327-0>.
 59. Wang, Z., Xiao, Y., Wu, S., Chen, J., Li, A., and Tatsis, E.C. (2022). Deciphering and reprogramming the cyclization regioselectivity in bifurcation of indole alkaloid biosynthesis. *Chem. Sci.* 13, 12389–12395. <https://doi.org/10.1039/d2sc03612f>.
 60. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. <https://doi.org/10.1038/nmeth.3317>.
 61. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
 62. Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., and Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18, 188–196. <https://doi.org/10.1101/gr.6743907>.
 63. Lomsadze, A., Ter-Hovhannisyian, V., Chernoff, Y.O., and Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* 33, 6494–6506. <https://doi.org/10.1093/nar/gki937>.
 64. Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinf.* 5, 59. <https://doi.org/10.1186/1471-2105-5-59>.
 65. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. <https://doi.org/10.1093/nar/28.1.27>.
 66. Bairoch, A., and Apweiler, R. (2000). The Swiss-Prot protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 28, 45–48. <https://doi.org/10.1093/nar/28.1.45>.
 67. Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., Flood, E.W., Gardner, P.P., Jones, T.A., Tate, J., and Finn, R.D. (2015). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* 43, D130–D137. <https://doi.org/10.1093/nar/gku1063>.
 68. Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435. <https://doi.org/10.1093/nar/gkn176>.
 69. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238. <https://doi.org/10.1186/s13059-019-1832-y>.
 70. Zwaenepoel, A., and Van de Peer, Y. (2019). WGD—simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* 35, 2153–2155. <https://doi.org/10.1093/bioinformatics/bty915>.
 71. Sahu, S.K., Thangaraj, M., and Kathiresan, K. (2012). DNA extraction protocol for plants with high levels of secondary metabolites and polysaccharides without using liquid nitrogen and phenol. *ISRN Mol. Biol.* 2012, 205049. <https://doi.org/10.5402/2012/205049>.
 72. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 3, 95–98. <https://doi.org/10.1016/j.cels.2016.07.002>.
 73. Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). De novo assembly of the genome using Hi-C yields chromosome-length scaffolds. *Science* 356, 92–95. <https://doi.org/10.1126/science.aal3327>.

74. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. <https://doi.org/10.1101/gr.092759.109>.
75. Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics Chapter 4. Unit 4.10*. <https://doi.org/10.1002/0471250953.bi0410s05>.
76. Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110, 462–467. <https://doi.org/10.1159/000084979>.
77. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. <https://doi.org/10.1093/nar/27.2.573>.
78. Edgar, R.C., and Myers, E.W. (2005). PILER: identification and classification of genomic repeats. *Bioinformatics* 21 (Suppl 1), i152–i158. <https://doi.org/10.1093/bioinformatics/bti1003>.
79. Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. <https://doi.org/10.1093/nar/gkm286>.
80. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. <https://doi.org/10.1093/gigascience/giab008>.
81. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. <https://doi.org/10.1038/nbt.3122>.
82. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. <https://doi.org/10.1186/1471-2105-9-559>.
83. Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. <https://doi.org/10.1093/nar/28.1.33>.
84. Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. <https://doi.org/10.1093/nar/25.5.955>.
85. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. <https://doi.org/10.1093/molbev/mst010>.
86. Alexandros, S. (2014). RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
87. Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., and Warnow, T. (2014). ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics* 30, i541–i548. <https://doi.org/10.1093/bioinformatics/btu462>.
88. De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22, 1269–1271. <https://doi.org/10.1093/bioinformatics/btl097>.
89. Niu, L., Yuan, H., Gong, F., Wu, X., and Wang, W. (2018). Protein Extraction Methods Shape Much of the Extracted Proteomes. *Front. Plant Sci.* 9, 802. <https://doi.org/10.3389/fpls.2018.00802>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Leaf, petiole, branch and Trunk bark of <i>Alstonia scholaris</i> .	Ruili Botanical Garden in Ruili, Yunnan Province, China	N/A
Deposited data		
Genome sequencing and assembly data	This study	CNSA with project accession CNP0002381. (https://db.cngb.org/search/?q=CNP0002381)
Experimental models: Organisms/strains		
<i>Alstonia scholaris</i> ecotype f Yunnan Province, China	Ruili Botanical Garden	N/A
Software and algorithms		
Nextdenovo (v 2.3.0)	N/A	https://github.com/Nextomics/NextDenovo
NextPolish (v 1.3.1)	N/A	https://github.com/Nextomics/NextPolish
Juicer (v 1.6)	Durand et al. ⁶⁰	N/A
Juicerbox	N/A	https://github.com/aidenlab/juicerbox
JCVI	N/A	https://github.com/tanghaibao/jcvi
RepeatMasker (v 4.0.6)	Chen et al. ⁶¹	N/A
RepeatProteinMask (v 4.0.6)	Chen et al. ⁶¹	N/A
Tandem Repeats Finder (v 4.07b)	Benson et al. ⁶²	N/A
Piler (v1.0)	Edgar & Myers ⁶³	N/A
LTR-FINDER (v 1.06)	Xu et al. ⁶⁴	N/A
Hisat2 (v 2.1.0)	Kim et al. ⁶⁵	N/A
StringTie (v 1.3.3b)	Pertea et al. ⁶⁶	N/A
WGCNA	Langfelder & Horvath ⁶⁷	N/A
Maker (v 2.31)	Cantarel et al. ⁶⁸	N/A
OrthoFinder (v 2.3.3)	Emms and Kelly ⁶⁹	N/A
WGD	Zwaenepoel et al. ⁷⁰	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests can be directed to Prof. Huan Liu (liuhuan@genomics.cn).

Materials availability statements

The study did not generate new unique reagents.

Data and code availability

- The raw data of genome, transcriptome sequencing and assembly data of *A. scholaris* are deposited at CNSA (<https://db.cngb.org/cnsa/>) under the project accession number CNP0002381, and all datasets are publicly available before the date of publication.
- The DESeq2 and WGCNA analysis R scripts are provided in [Data S1](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

METHOD DETAILS

Plant sample collection and sequencing

The *Alstonia scholaris* (ID 52822) plant cultivated in the Ruili Botanical Garden of Yunnan Province, China, was used in this study. We collected fresh and young leaves for Nanopore, Hi-C and WGS sequencing. The purity, concentration and integrity of the extracted DNA were tested by Nanodrop, Qubit and Agarose Gel Electrophoresis, respectively. The library was constructed by using SQK-LSK109 kit and the PromethION platform was used for ONT sequencing. For Hi-C experiment, we cut the fresh leaves into fragments with 50 ml of MC buffer and 1.39 ml of

37% methanol to infiltrate those fragments. The methanol-processed tissues were ground to powder in liquid nitrogen for DNA extraction using the CTAB (cetyl trimethyl ammonium bromide) method.⁷¹ The Hi-C library was constructed and sequenced on the BGISEQ-500 platform, according to their standard protocol.

We collected leaf, petiole, branch, and trunk bark samples to extract total RNA by using the CTAB-BBIOZOL method. The RNA quality was evaluated by Nanodrop, Qubit 2.0 and Agilent 2100 instruments to ensure that the RNA was suitable for library construction and sequencing. Then, 4 μ L fragmentation buffer was added to the mRNA sample liquid under 150 bp fragmentation conditions for eight minutes. The fragmented samples were mixed with RT buffer to start reverse transcription to obtain the second chain product. After purification, the "A" and adapter bases were added, and after PCR amplification and enzyme cleavage, the RNA-seq library was constructed.

Genome assembly and chromosome anchoring

A total of 295 Gb of short reads and 62 Gb of long reads were generated for the genome assembly. Nextdenovo (v 2.3.0) software (<https://github.com/Nextomics/NextDenovo>) was used for *de novo* assembly of *A. scholaris* genome via a "correct-then-assemble" strategy. NextPolish (v 1.3.1) (<https://github.com/Nextomics/NextPolish>) was used to fix base errors in the genome generated by noisy long reads with a combination of short and long-reads data.

A total of 69 Gb data were generated for the Hi-C maps. Juicer⁷² (v 1.6) software was used to map the Hi-C data to the assembled genome, and sorting and merging steps generated the input file of the 3D *de novo* assembly (3D-DNA) pipeline⁷³ which assembles an accurate genome with chromosome-length scaffolds. Juicerbox (<https://github.com/aidenlab/juicerbox>) was used for manual correction, after which 3D-DNA was reanalyzed to generate the final assembled genome. The completeness of the genome assembly was assessed against the eudicot database (odb10) by BUSCO with default settings.

The method for identifying repeat sequences is described in the section titled "Identification of Repetitive Sequences." Gene density was calculated based on the gene positions within each window. A window size of 1 Mb with a step size of 1 Mb was used for sliding windows to calculate the GC content. The colinear regions of chromosomes were obtained using JCVI (<https://github.com/tanghaibao/jcvi>). Finally, the Circos software⁷⁴ was employed to combine all the results and generate the figure.

Identification of repetitive sequences

RepeatMasker (v 4.0.6) and RepeatProteinMask (v 4.0.6)⁷⁵ were used to search in the Repase⁷⁶ database to identify TEs in the DNA and protein leaves. Tandem Repeats Finder (v 4.07b)⁷⁷ was used to identify tandem repeats. Our strategy for identifying repeat sequences involves a combination of *de novo* and homology-based methods. *De novo* identification was performed with Piler (v1.0)⁷⁸ and LTR-FINDER (v 1.06),⁷⁹ and RepeatMasker. By using the previously constructed libraries as a database, RepeatMasker was used to identify and classify the final repeats in the *A. scholaris* genome.

RNA-seq analysis

Hisat2 (v 2.1.0)⁸⁰ was used to map the clean RNA-seq data to the *A. scholaris* genome with the following parameters: hisat2-2.1.0/hisat2-aligns -wrapper basic -0 -t -x index -1 clean.read1.fq.gz -2 clean.read2.fq.gz -S clean.sam. Then, we used samtools (v 1.7)⁸⁰ software to sort the bam files as an input file of StringTie (v 1.3.3b)⁸¹ to predict each sample's transcript in bulk and integrate them into one nonredundant transcript. The parameters were as follows: stringtie sorted.bam -p 15 -G genome.gtf -o sorted.bam.gtf; stringtie -merge -p 20 -G genome.gtf -o merged_stringtie.gtf mergelist.txt. Finally, each sample's gene expression was quantified and integrated by the following procedure: stringtie -e -B -p 8 -G merged_stringtie.gtf -o ballgown/output_merge.gtf sorted.bam; stringtie/prepDE.py -i ballgown.

We used DESeq2 R package⁸¹ to perform differential expression analysis with gene count data. A gene matching adjusted p value < 0.05 and a $|\log_2\text{FoldChange}| \geq 2$ were considered to indicate differentially expressed genes. The co-expression analysis was conducted using the WGCNA package⁸² in R software. All FPKM values data from 19 samples were used as the input file, and a power of 8 (soft thresholding power) was used for correlation coefficient analysis to determine the difference between gene correlations. Additionally, we defined clear Pearson correlation coefficient thresholds, such as $|r| > 0.6$ and $p < 0.05$. The DESeq2 and WGCNA Rscript data are supplied in [Data S1](#).

Gene model prediction and functional annotation

Maker (v 2.31)⁶² was used for gene annotation with homology, *de novo* and transcriptome-based prediction evidence. We used the protein sequences of *A. thaliana*, *C. gigantea*, *C. roseus*, *C. canephora*, *G. sempervirens*, *N. tabacum*, *O. pumila*, *O. sativa*, *R. serpentina*, *S. lycopersicum* and known MIA-related genes from the uniprot database as homologous species. Genemark-ES (v 4.21)⁶³ was used for unsupervised self-training of the eukaryote genome with the default criteria. The first round of MAKER analysis was run with EST sequences, homologous species sequences, GeneMark HMMs and Augustus training HMMs of *A. scholaris*. SNAP⁶⁴ was subsequently trained with the first round of results. The second round of MAKER was run with the above data and the gff file generated by the first-round analysis.

We aligned the predicted protein sequences against the KEGG,⁶⁵ COG,⁸³ SwissProt,⁶⁶ TrEMBL, InterPro, and NR protein databases by BLASTP (E-value $\leq 1e-05$). RNAscan-SE v1.3.1⁸⁴ was used for tRNA gene identification. We aligned the assembled genome against the plant rRNA and Rfam⁶⁷ databases using BLASTN (E-value $\leq 1e-05$) for rRNA, snRNA and miRNA annotation.

Candidate MIA gene prediction

We downloaded all the identified MIA biosynthesis-related protein sequences (Table S34) as query sequences and performed BLASTP analysis (identity > 40, e-value > 1e-20) with *A. scholaris*, *A. thaliana*, *Amborella trichopoda*, *C. gigantea*, *C. roseus*, *C. canephora*, *G. sempervirens*, *N. tabacum*, *O. pumila*, *O. sativa Japonica Group*, *Populus trichocarpa*, *R. serpentina*, *Rhazya stricta*, *S. lycopersicum*, *Solanum tuberosum*, *Sorghum bicolor*, and *Vitis vinifera* protein sequences. Moreover, the Interpro annotation information (Ipr, Pfam, and GO) was combined to find the best match sequences for the species-specific candidate genes.

Gene clusters of *A. scholaris* were predicted by Plantismash (<http://plantismash.secondarymetabolites.org/>) software with default parameters. The input files include assembled genome with fasta format and the annotation file with gff format.

Gene family analysis and evolutionary tree construction

For the gene family clustering analysis, 15 plant genome sequences, namely those of *C. gigantea*, *C. roseus*, *R. serpentina*, *A. thaliana*, *A. trichopoda*, *R. stricta*, *C. canephora*, *G. sempervirens*, *O. pumila*, *P. trichocarpa*, *V. vinifera*, *S. lycopersicum*, *S. tuberosum*, *S. bicolor* and *O. sativa*, were used with the *A. scholaris* genome. OrthoFinder software (v 2.3.3)⁶⁸ was used for gene family cluster identification. The output of OrthoFinder was subsequently passed to identify gene families. If an orthologous group contained more than or equal to eight species then that orthologous gene was considered to be a single-copy ortholog. MAFFT (v 7.310)⁶⁵ was used to align single-copy genes in all species. RAxML⁸⁶ (v 8.2.4) was used to construct each a gene tree with the PROTCATGTR model. Astral (v 5.5.9)⁸⁷ with 100 bootstrap replicates was used to construct the species phylogenetic tree.

The gene tree about AsSTRs and AsTDCs were also constructed by MAFFT and RAxML softwares, then polished by iTOL (<https://itol.embl.de/>).

MCMCTREE⁸⁶ was used to estimate the divergence time between *A. scholaris* and other species with the default parameters. CAFÉ⁸⁸ was used to predict the expansion and contraction of gene family numbers by employing a phylogenetic tree and gene family statistics. WGD software⁷⁰ was used to perform the Ks distribution analysis.

Protein detection

First, proteins from leaf, branch and trunk bark samples were extracted by using the short gradient phenol extraction method.⁸⁹ Next, target proteins were detected by using label-free technology with a bottom-up strategy. MaxQuant software was subsequently used to search for proteins in the target database (annotated protein sequence file of *A. scholaris* by ourselves) with the following parameters: Mass accuracy of MS and MS/MS of 20 ppm and 0.5 da Orbitrap, respectively.

Metabolite detection

The leaf, branch and trunk bark samples were aliquots of 0.5 g each. The samples were triturated with 10 ml of 70% methyl alcohol and incubated under quiescent conditions in an ultrasonic cleaner for 45 mins. Subsequently, the samples were centrifuged at 6000 rpm for 15 min, after which the liquid supernatant was collected. The residue was removed, and the previous steps were repeated. The extracted supernatant was combined and drained overnight with a vacuum draining machine. The samples were redissolved in 1 ml of 70% methyl alcohol and placed into an ultrasonic cleaner for 1~2 hours. Next, the samples were transferred to 2 ml centrifuge tubes and centrifuged at 12,000 rpm for 10 minutes. The liquid supernatant was collected and stored at -20°C.

We used authentic standards purchased from a certified vendor (<https://www.rmuu.com/>), including tryptamine and secologan, to carry out targeted metabolomics analysis.

Mass spectrometry detection was performed on a quadrupole mass spectrometer, Q-Exacte (Thermo Fisher Scientific) equipped with a heater-electrospray (HESI) source in positive mode for parallel reaction monitoring (PRM)-MS analysis. The conditions of the MS/MS detector were as follows: the flow rate of sheath gas (nitrogen) was 40 arb and the flow rate of auxiliary gas (nitrogen) was 11 arb; the capillary temperature was 320°C; the spray voltage was 3.8 kV; the probe heater temperature was 320°C and the S-lens RF level was 50. LC-MS/MS chromatography was performed on a Kinetex® 1.7 μm EVO C18(100×2.1 mm) column (Phenomenex). The column temperature was 30°C. The injection volume was 10 μl. The solvents used were H₂O + 0.1% formic acid as Solvent A and 100% acetonitrile (LC-MS grade) as Solvent B, with a flow rate of 0.3 ml/min. The gradient elution program was as follows: 0 min, 10% B; hold for 2 min; hold for 2 to 5 min, linear gradient to 30 % B; from 5 min to 8 min, linear gradient to 35% B; to 8.5 min, linear gradient to 100% B; hold for 1.5 min; and from 10 min to 11 min, back to 10% B for 2 min to re-equilibrate the column.

QUANTIFICATION AND STATISTICAL ANALYSES

Bioinformatic analysis was described in the [method details](#) section. The standardized thresholds about screening related modules for tissues of co-expression analysis is $|r| > 0.6$, $p < 0.05$. The statistical method of enrichment analysis is χ^2 test. AdjustedPv is a corrected p-value that is obtained by performing false discovery rate (FDR) testing on p-values. AdjustedPv obtained from significance tests are generally considered to have significant statistical differences when AdjustedPv < 0.05.