Characterization of *O*-Methyltransferases in the Biosynthesis of Phenylphenalenone Phytoalexins Based on the Telomere-to-Telomere Gap-less Genome of *Musella lasiocarpa*

Wanli Zhao [1, †], Junzhi Wu [1, 4, †], Mei Tian [1], Shu Xu [1], Shuaiya Hu [3], Zhiyan Wei [3], Guyin Lin [1], Liang Tang [1], Ruiyang Wang [1], Boya Feng [1], Bi Wang [1], Hui Lyu [2], Christian Paetz [2], Xu Feng [1], Jia-Yu Xue [3, *], Pirui Li [1, *], Yu Chen [1, *]

[1] Jiangsu Key Laboratory for the Research and Utilization of Plant Resources, Jiangsu Province Engineering Research Center of Eco-cultivation and High-value Utilization of Chinese Medicinal Materials, Institute of Botany, Jiangsu Province and Chinese Academy of Sciences (Nanjing Botanical Garden Mem. Sun Yat-Sen), 210014 Nanjing, China

[2] NMR/Biosynthesis Group, Max-Planck-Institute for Chemical Ecology, Hans-Knöll-Straße 8, 07745 Jena, Germany

[3] College of Horticulture, Bioinformatics Center, Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, 210095 Nanjing, China

[4] Nanjing University of Chinese Medicine, 210023 Nanjing, China

[†] the authors contributed equally to this work.

[*]Corresponding authors.

[*]Corresponding authors telephone and fax number: Yu Chen, Email, ychen@jib.ac.cn, Telephone, 025-84347116, fax number 025-84347084; Pirui Li, Email, lipirui@cnbg.net, Telephone, 025-84347159, fax number 025-84347084; Jia-Yu Xue, Email, xuejy@njau.edu.cn, Telephone, 15195886485, fax number 025-84395869;

The authors email: Wanli Zhao (zhaowanli@jib.ac.cn), Junzhi Wu (15535430762@163.com), Mei Tian (tianmei@cnbg.net), Shu Xu (shuxu@cnbg.net), Shuaiya Hu (whosy@stu.njau.edu.cn), Zhiyan Wei (2022104136@stu.njau.edu.cn), Guyin Lin (lgy_grace@sina.com), Liang Tang (liangtang0107@163.com), Ruiyang Wang (shiqi059@gmail.com), Boya Feng (boyafeng@cnbg.net), Bi Wang (wangbi@cnbg.net)), Hui Lyu (hlyu@ice.mpg.de), Christian Paetz (cpaetz@ice.mpg.de), Xu Feng (fengxucnbg@cnbg.net).

Note: Since the official email address of the institute is only open to on-the-job employees, here are Junzhi Wu, Guyin Lin, Liang Tang, and Ruiyang Wang using their private email addresses.

Abstract

Phenylphenalenones (PhPNs), phytoalexins in wild bananas (Musaceae), are known to act against various pathogens. However, the abundance of PhPNs in many Musaceae plants of economic importance is low. Knowledge of the biosynthesis of PhPNs and the application of biosynthetic approaches to improve their yield is vital for fighting banana diseases. However, the processes of PhPNs biosynthesis, especially those involved in methylation modification, remain unclear. *Musella lasiocarpa* is an herbaceous plant belonging to Musaceae; and due to the abundant PhPNs, the biosynthesis in *M. lasiocarpa* has been the subject of much attention. In this study, we assembled a telomere-to-telomere gap-less genome of *M. lasiocarpa* as the reference, and further integrated transcriptomic and metabolomic data to mine the candidate genes involved in PhPN biosynthesis. To elucidate the diversity of PhPNs in *M. lasiocarpa*, three screened *O*-methyltransferases (Ml01G0494, Ml04G2958, and Ml08G0855) by phylogenetic and expressional clues were applied to *in vitro* enzymatic assays. The results show that the three were all novel *O*-methyltransferases involved in the biosynthesis of PhPN phytoalexins, among which, Ml08G0855 was proved to function as a multifunctional enzyme targeting multiple hydroxyl groups in the PhPNs structure. Moreover, we tested the antifungal activity of PhPNs against *Fusarium oxysporum* and found that the methylated modification of PhPNs enhanced their antifungal activity. These findings provide valuable genetic resources in banana breeding and lay a foundation for improving disease resistance through molecular breeding.

Keywords: *Musella lasiocarpa*; Telomere-to-Telomere gap-less genome assembly; Phytoalexins; Phenylphenalenones; Methyltransferase

Introduction

Bananas (*Musa* spp.), which originated in Southeast Asia, are one of the most important commercial crops in the world [1]. China is the world's second-biggest producer of bananas after India, harvesting about 11.7 million tons in 2021 [2]. However, banana yields are severely curtailed by diseases caused by fungi, viruses, and plant-parasitic nematodes. Notably, the Banana Fusarium Wilt (BFW), caused by *Fusarium oxysporum* f. sp. *cubense* tropical race 4, is one of cultivated banana's most destructive diseases; despite decades of researches, few effective options for managing this disease have been developed. Planting resistant cultivars is widely considered to be the prior strategy in affected areas [3]. The wild relatives of commercial crops usually possess more genetic diversity and so are often useful for developing more resistant varieties [4], in contrast to cultivated plants, which often produce less variety and/or fewer defense-related secondary metabolites than their wild-type relatives [5,6]. The use of defense metabolites, e.g., phenylphenalenones (PhPNs), can be an important strategy in overcoming current problems of banana cultivation.

Phenylphenalenone-type secondary metabolites, which consist of a tricyclic phenalene nucleus and a lateral phenyl ring (Figure 1B), occur mainly in monocot taxa, like Strelitziaceae and Musaceae [7]. Because PhPNs in wild banana plants were reported to be important phytoalexins and phytoanticipins, they were considered valuable resources for breeding disease-resistant banana cultivars [8,9]. However, cultivated banana plants contain PhPNs in low concentrations and low structural variety. Genetic engineering could be used to alter that imbalance, influencing the biosynthesis of PhPNs and the enzymes involved in order to tailor disease-resistant plants. Previous studies suggested that PhPNs are biosynthetically derived from the phenylpropanoid pathway and that their linear precursors are transformed through an intramolecular Diels-Alder cyclization [7]. Only a chalcone synthase (CHS) *WtPKS1* catalyzing the first step in diarylheptanoid biosynthesis was characterized from *Wachendorfia thyrsiflora* [10]. All other biosynthetic enzymes that contribute to the formation of PhPNs are still unknown.

Musella lasiocarpa is an endemic plant in China and the only member of the genus *Musella* (Figure 1A). It is mainly distributed in southwestern areas of the country, such as the Yunnan and the Sichuan provinces [11,12]. Because of its long flowering period and beautiful appearance (e.g., large and golden inflorescences), *M. lasiocarpa* is often cultivated as an ornamental plant and is known as "Di Yong Jin Lian" in Chinese. In addition to its ornamental value, the flowers and bracts of *M. lasiocarpa* are used in traditional folk medicine to stop bleeding and to counteract inflammation [11]. A recent phytochemical investigation of *M. lasiocarpa* showed the plant contained various PhPNs and also linear diarylheptanoids (LDHs), which are considered biosynthetic precursors of PhPNs [13]. Our previous research showed that LDHs and PhPNs accumulate in the seed coats of *M. lasiocarpa*

during the middle and late stages of seed development [14], and thus the seeds may be a useful model system for the study of PhPN biosynthesis. In the present work we focused on the *O*-methylation of PhPNs catalysed by *O*-methyltransferases (OMTs). During *O*-methylation, the transfer of a methyl group from *S*-adenosyl-L-methionine (SAM) to a hydroxyl group of an appropriate substrate is achieved. In plants, *O*-methylation occurs as part of the biosynthesis of many types of secondary metabolites, including terpenoids, alkaloids, flavonoids, and also PhPNs [7,15-17]. *O*-methylation of PhPNs contributes significantly to the bioactivity of these metabolites by influencing their diffusion into biological membranes [18]. Previous studies found that methylation of phenolic hydroxyl groups can improve PhPNs' activity against the pathogen *Mycosphaerella fijiensis* [19].

In this work, we assembled a high-quality genome of *M. lasiocarpa*. Using the genomic data as the fundament, we integrated multi-omics data to screen candidate enzymatic genes involved in the methylation of PhPNs, and along with functional experiments to verify the real *O*-methyltransferase genes and elucidate the diversified functional roles. The present study provides valuable information for the use of genetic resources found in the wild banana relative *M. lasiocarpa* and offers important insights into the function of *O*-methyltransferases in PhPN biosynthesis.

Results

Nearly Telomere-to-Telomere Gap-less Genome Assembly and Annotation of *M. lasiocarpa*

We sequenced and assembled a telomere-to-telomere gap-less genome of *M. lasiocarpa*, based on approximately 35.49 Gb (> 65 X) of HiFi data, 26 Gb (> 48 X) Hi-C data. The K-mer distribution analysis revealed a genome size of 535 Mb with 1.02% heterozygosity and 51.12% repetition (Figure S1), fitting the genome size based on flow cytometry analysis and k-mer analysis (Table S1). The preliminary assembly by HiFi data generated 440 contigs with a total length of 509 Mb and a contig N50 value of 56.62 Mb (Table 1). The completeness of the *M. lasiocarpa*'s genome assembly was evaluated by BUSCO [20] and CEGMA (https://github.com/marbl/merqury), respectively, in which 98.50% complete and 96.77% coverage of the complete matches were identified in the assembly. Then, the *M. lasiocarpa* contigs were further assembled to scaffolds using Hi-C data. Approximately 470 Mb (14 contigs, 92.32%) of the total assembly were successfully anchored into nine pseudochromosomes, corresponding to the nine haplotype chromosomes of *M. lasiocarpa* (Figure S2A and Figure 1C).

The identification of centromeres and telomeres indicates that our assembly reached a high completeness and continuity, with all nine centromeres and almost all (16/18) telomeres detected. *M. lasiocarpa* telomeres consist of tandem repeats of TTTAGGG and are located at both ends of chromosomes, except Chromosome 7 and 9, which only possess an intact telomere at one end,

respectively. The lengths of the identified *M. lasiocarpa* telomeres vary greatly, ranging from 261 bp to 5,674 bp, and most are above 1 Kb, with only two below that (Table S2). The nine *M. lasiocarpa* centromeres consist of tandem repeats of different sequences and varying lengths, and the lengths range from 198 Kb to 2,271 Kb (Table S3). These identified telomeres and centromeres make our assembly a telomere-to-telomere gap-less genome, with only nine gaps present and one gap for each chromosome. All nine gaps are all found to be located within the telomere regions, so that are unlikely to encode any protein-coding genes, which means our current assembly and annotation have covered a complete gene repertoire.

We then performed the genome annotation in *M. lasiocarpa* assembly. A total of 34,361 protein-coding genes were predicted, with an average sequence length of 4,379.73 bp per gene, similar to those reported in other plants from the same family (Table S4) [21-23]. On average, each predicted gene contains 4.92 exons with 245.10 bp in length for each exon. Functional annotation captured 97.67% of the protein-coding genes by similarity searches against protein domains and homologous sequences (Table S5). Moreover, we identified noncoding RNA (ncRNA) genes in *M. lasiocarpa* assembly, including 12,929 rRNA, 412 miRNA, 380 snRNA, and 3,015 tRNA genes (Table S6).

Phylogenetic Position and Genome evolution of *M. lasiocarpa*

To resolve the phylogenetic position of *M. lasiocarpa*, and the relationship with other Musaceae species, a concatenated dataset comprising 1,371 single copy orthologous genes from 12 species was constructed and phylogenomic analysis based on the dataset was performed using the maximum likelihood (ML) method. The analysis resolved *M. lasiocarpa* as the sister to *Ensete glaucum*, and they together were recovered to be sister to the *Musa* genus. Molecular clock analysis estimated the origin of Musaceae to be around 56.3 million years ago (MYA), and the divergence between *M. lasiocarpa* and *Ensete glaucum* to be around 37.6 MYA (Figure 2A). Intergenomic co-linearity analysis (Figure S2B) showed an almost one-to-one syntenic relationship at the chromosome level between *M. lasiocarpa* and *E. glaucum*, suggesting the well preserved genomic structure for the two species. More substantial rearrangement between *M. lasiocarpa* and *M. acuminata* were also detected. For instance, Chromosome 2 of *M. acuminata* was only a part of that of *M. lasiocarpa*, suggesting either a chromosome break event in *M. acuminata*, or a chromosome fusion event in *M. lasiocarpa*. By calculating the synonymous mutation rates (Ks) of anchored paralogous gene pairs, we were able to identified potential whole genome duplications (WGDs) in these species (Figure 2B). The Ks density distribution of *M. acuminata* showed two peaks that indicated multiple WGD events, which is in line with *E. glaucum* and banana in the same family [24,25]. Previous studies reported either three or four rounds of WGDs in the evolution of Musaceae plants, and proposed that the most recent two

rounds of WGDs (α and β) occurred consecutively at a similar period around 65 MYA, so that only display one single peak was displayed representing α and β WGDs [24-26]. Our intragenomic synteny analysis did found that, in the three Musacaeae species most paralogous gene clusters shared relationships with three other clusters, with similar Ks values in *M. lasiocarpa*, *E. glaucum*, and *M. acuminata* (Figure S3), supporting the continuously occurring WGDs.

MlOMT genes transcriptional expression in _M. lasiocarpa_ seeds during different developmental stages

Methyltransferase, a subclass of the transferase family, plays a vital role in the formation of secondary metabolites in plants. Methyltransferases usually have conserved substrate binding domains and methyl donor binding domains. We searched for candidate methyltransferase genes (*MlOMT*) by combining homology-based BLAST and conserved domains (PF01596 or PF00891) in our *M. lasiocarpa* genome database [27-29]. In total, 30 *MlOMT* genes were predicted in the *M. lasiocarpa* genome (Figure S4). Next, we compared and analyzed the PhPN components in *Musa* and their related plant species. Results showed that PhPNs were more abundant in *M. lasiocarpa* than in other banana species. Especially in mature seeds of *M. lasiocarpa* the content of PhPNs was high. We then analyzed the content of PhPNs in seeds of *M. lasiocarpa* at three developmental stages (S2, yellow seed; S4, brown seed and S6, block seed) by HPLC/Q-TOF MS. The results showed that the PhPN content increased gradually in the order S2 <S4 <S6 (Figure S5). The results showed that the distribution and accumulation of PhPNs were time-specific in different stage seeds. This was in accordance with previously published results [14]. We investigated the expression of *MlOMT* genes in *M. lasiocarpa* seeds of different developmental stages. Most MlOMTs were unexpressed or expressed only at low levels. Only three MlOMT genes (Ml01G0494, Ml04G2958, and Ml08G0855, FPKM values ≥ 100) were higher expressed (Figure 3A). Furthermore, Ml01G0494, Ml04G2958, and Ml08G0855 were stage-specifically expressed. The expressional level of Ml01G0494 was 5.98-fold higher in S2 than in S6, while Ml04G2958 was expressed 2.21-fold lower in S2 than in S6 (Table S7). Therefore, Ml01G0494, Ml04G2958, and Ml08G0855 were selected as candidate genes which were probably involved in the biosynthesis of methoxylated PhPN.

We gathered all OMTs identified from 12 plant genomes (the same species used for species evolutionary analysis in Figure 2A) and from other plants previously characterized, and performed phylogenetic analysis for the OMTs. The results turned out that all OMTs should be classified into two subfamilies, namely Methyltransf_24 (subfamily I) and Methyltransf_2 (subfamily II) (Figure 3B and Figure S6). In fact, the two subfamilies show very low sequence identity. To be precise, they should be considered as two gene families with functional convergence. Of our screened three genes,

Ml01G0494 and Ml04G2958 fell into Methyltransf_2, and Ml08G0855 fell into Methyltransf_24, suggesting the convergent evolution of their potential functions.

Functional characterization and subcellular localization analysis of MlOMTs

The three candidate *MlOMT* genes, Ml01G0494, Ml04G2958, and Ml08G0855 were cloned into pMAL-c4x vectors with MBP tags and subsequently expressed in *Escherichia coli* BL21 (DE3) strains, respectively. The purified bands of the candidate MlOMTs proteins were analyzed by SDS-PAGE and found to be in accordance with theoretical molecular weights (Figure S7). To characterize the catalytic activity of the putative MlOMTs Ml01G0494, Ml04G2958, and Ml08G0855 *in vitro*, PhPN substrates with variable degrees of hydroxylation in rings A and/or D were used for methylation assays (Figure 4). We examined three types of PhPNs, 4-PhPNs, 9-PhPNs, and a dimeric 4-PhPN (Figure 1B). The crude MlOMT proteins were used for the assays, and the reaction products were analyzed by HPLC/Q-TOF-MS. The results revealed that the MlOMTs exhibited differential catalytic activity on the A or D ring of the PhPNs with regioselectivity (Figure 4, Figure S8, and Table S8). When the enzymes were tested with 9-PhPNs, the *O*-methylation of position 2 of ring A (Figure 4A, 4B, 4D and 4E) was observed. When the 9-PhPN was methoxylated in position 2, and hydroxylation of position 4' and 5' of ring D was present, a mono-methylation of position 5' was observed (Figure 4C). Ml08G0855 was capable to further methylate the reaction products MLT2 (Figure 4B) and Methoxy-MLT1 (Figure 4C). The dimeric compound $(4\text{-PhPN})_2$ (Figure 1B) could not be methylated by any of the enzymes. Compounds of this type seem too bulky to be suitable substrates for the enzymes.

In the next series of assays 4-PhPNs were examined. The methylation of position 2 could be observed by all three enzymes when a substrate as depicted in Figure 4F, 4G and 4H. Again, Ml08G0855 was able to methylate hydroxy functions in ring D. When monomethoxy-MLT10 was used as substrate, it was translated into dimethoxy-MLT10 by this enzyme (Figure 4F). We used the observed regiospecificities in an experiment that resulted in the sequential methylation of all hydroxy groups in a 4-PhPN substrate. In the first step, a substrate methoxylated in position 2 was recovered from the assay when Ml01G0494 or Ml08G0855 were used with a triply hydroxylated (positions 2, 4', and 5') substrate (Figure 4H and 4I). The reaction product was then further incubated with Ml01G0494 or Ml08G0855 to achieve mono-methoxylation of position 5'. The triple-methoxylation of position 4' was only produced by Ml08G0855. Results of this experiment show that Ml08G0855 is capable of catalyzing the methylation of a wide array of hydroxylated PhPNs.

In order to optimize the optimal reaction conditions of MlOMTs recombinase, the pH and temperature for the *in vitro* catalytic reaction of MlOMTs recombinase were examined using MLT4 as a substrate. Results showed that MlOMTs displayed optimal activity in 50 mM Tris-HCl buffer

(pH 8.0) at 45 °C (Figure S9). *Km* and *Kcat* values were calculated by nonlinear curve fitting the Michaelis-Menten model (Figure 5). As illustrated in Table S9, the apparent *Km* value of Ml01G0494 for MLT9 (970.60 μM) was the highest; however, the *Kcat/Km* value of Ml01G0494 for MLT9 (2.19 μM^{-1} s^{-1}) was the lowest. The apparent *Km* and *Kcat/Km* values of Ml08G0855 for MLT9 were 81.93 μM and 12.24 μM^{-1} s^{-1}. Above results indicated that the catalytic efficiency of Ml08G0855 for substrate MLT9 was better than that of Ml01G0494. The apparent *Kcat/Km* values of Ml01G0494, Ml04G2958, and Ml08G0855 for MLT3 were 25.58, 7.14, and 8.88 μM^{-1} s^{-1}, respectively. Similarly, the apparent *Kcat/Km* values of Ml01G0494, Ml04G2958, and Ml08G0855 for MLT4 were 84.24, 63.24, and 7.42 μM^{-1} s^{-1}, respectively. Overall, Ml01G0494 was the more efficient enzyme for the methylation of MLT3 and MLT4 compared with Ml04G2958 and Ml08G0855. While, for substrate MLT9, Ml08G0855 was more efficient than Ml01G0494.

To analyze the subcellular localization of MlOMTs, the recombinant Ml01G0494, Ml04G2958, and Ml08G0855 plasmids fused with GFP were transiently expressed in *N. benthamiana* leaves, respectively (the primers are listed in Table S10). As shown in Figure S10, the fluorescent signals of the three MlOMTs fusion GFP proteins distributed throughout the cytoplasm and the nucleus. Such a pattern indicated that the cytoplasm might be the subcellular site for the PhPN *O*-methylation.

Antifungal activity of PhPNs against *Fusarium oxysporum*

To investigate their antifungal activity, PhPNs were tested against the banana pathogen *F. oxysporum* f. sp. *cubense*, Foc 4, the cause of the devastating Panama disease. MLT1, MLT3, MLT4, MLT6, MLT7, MLT9, MLT10, and MLT11 exhibited significant inhibitory activity against the pathogen (Figure S11 and Figure S12). In particular, methylated products from the assays shown in Figure 4A, Figure 4H and the substrate from Figure 4C, Figure 4H showed antifungal activity that outcompeted the commercial fungicide, thiophanate methyl (TM). Generally, when 2-hydroxyl groups were derivatized by methylation in either 4- or 9-PhPNs, antifungal activities were enhanced. Clearly, *O*-methylated modifications of PhPNs possess effective antifungal activities.

Discussion

Bananas are a staple food for millions of people in the tropics and subtropics, but yield and quality of the fruits are increasingly affected by diseases as the global warming progresses. However, effective approaches to control these diseases other than the increased use of pesticides are so far lacking [30]. Conventional pesticides may cause environmental contamination and also affect food safety. Because of their low concentrations in cultivated varieties of sweet banana, PhPNs cannot act as endogenous defensive substances. Mainly found in the plant families Musaceae, Strelitziaceae, and Pontederiaceae [14,31,32], PhPNs have recently been isolated and characterized. Once the biosynthetic pathway leading to PhPNs was identified, modern methods of plant breeding by genetic engineering

may be used to increase the concentration of PhPNs in commercial banana crops. The wild banana relative we chose for our study (*M. lasiocarpa*) has a higher PhPN content than that of commercially cultivated banana. To further explore the biosynthetic pathway of PhPNs, we assembled a chromosomal-level genome assembly of *M. lasiocarpa* obtained by combining PacBio, Bionano, Hi-C, and Illumina sequencing technology. At present, genome sequencing has been conducted on a variety of plants of the Musaceae family, such as *M. acuminata*, *M. balbisiana*, *E. glaucum*, and *M. beccarii* [21,22,25,33]. However, the genome data of high PhPN content species was missing. We sequenced the only member of the genus *Musella*, which laid the foundation for mining PhPNs biosynthetic genes. We then analyzed the PhPNs' metabolism in seeds of *M. lasiocarpa* at developmental stages and combined our results with data from the transcriptome analysis. This led us to the discovery of several OMT genes putatively involved in the biosynthesis of PhPNs. After heterologous expression, the resulting enzymes were tested regarding their catalytic activity *in vitro* assays and the reasons for the diversity of methylated PhPNs were identified. Our OMTs show a strong substrate specificity leading to specific methoxylation patterns. This finding firmly corroborates earlier studies that identified *O*-methylation occurring at the end of PhPN biosynthesis [18]. Methylation is an important modification of plant secondary metabolites, which can change their physical and chemical properties, including stability and solubility [34]. Previous studies have shown the 4'-hydroxymethylated of PhPNs could improve the inhibition activity against *Fusarium oxysporum*, the pathogen causing banana fusarium wilt [35]. Furthermore, when the 2,4-dimethoxyphenyl group at the 6-position of 6-position of phenalenone skeleton could increase the antifungal activity against *F. omysporum* [36]. Finally, we verified the *in vitro* activity of PhPNs and found the methylated products synthesized by OMTs possess significantly increased their antifungal activity. Our results are consistent with previous studies, suggesting that methylated PhPN could significantly affect their biological activities. The present study confirms the beneficial function of the identified genes. By introducing them into existing banana breeds they can be used to fight diseases.

Materials and methods

Plant materials

For genomic studies, yellow-bracted *M. lasiocarpa* were collected in Nanhua County, Yunnan Province, China (118°50'38"E, 32°3'44"N) and transplanted in the experimental field, Nanjing, China (101°1'2"E, 25°9'54"N). For RNA sequencing, the leaves, stems, and seeds at three developmental stages were collected (S2, S4, and S6) from the same individual plant.

Whole genome sequencing

Young leaves of *M. lasiocarpa* were used for extraction genomic DNA by the Plant DNA kit (TIANGEN, China). DNA quality was assessed using NanoDrop 2100 spectrophotometry (Agilent, USA) and agarose gel electrophoresis, followed by Qubit fluorometry (Thermo Fisher Scientific, USA). An Illumina HiSeq X Ten platform was used to construct and sequence the short paired-end libraries. For long reads sequencing, the PacBio library was constructed and sequenced on PacBio Sequel II platform. The 15-kb preparation solutions was applied to construct a SMRTbell target size library. The obtained genomic DNA was cross-linked, digested by the restriction enzyme and labelled via biotinylated residues for Hi-C sequencing. Biotinylated constructs were enriched, sheared, and sequenced by the Illumina HiSeq X Ten platform. K-mers analysis is widely used in genome size evaluation. K-mers with 17–31 bp were counted via Jellyfish (version 2.2.7), and GenomeScope website, was used for estimating the genome size and heterozygosity according to the k-mer frequency.

RNA sequencing

M. lasiocarpa seeds from three developmental stages (S2, S4, and S6) were selected for transcriptome sequencing. These samples were collected from *M. lasiocarpa* and frozen in the liquid nitrogen after incubation in RNAlater. The total RNA of each sample was extracted via the plant RNA isolation kit (RC411, Vazyme, Nanjing, China). The mRNA sequencing library was constructed and then sequenced by Illumina Novaseq 6000 platform. For full-length transcriptome sequencing, the mixed RNA library from leaves, seeds and roots of *M. lasiocarpa* was sequenced by PacBio Sequel II.

Genome assembly and annotation

To construct the *M. lasiocarpa* genome, Hifiasm (v_0.16.1) was used to generate the assembly contigs with the HiFi reads [37]. Then, the draft genome was assembled into scaffolds with Hi-C data by the 3D-DNA pipeline tool [38]. These scaffolds were roughly split via Juicebox tool and another round of scaffolding. The completeness of genome assembly was assessed by BUSCOs [20] and transcriptome data. Homology-based and *ab initio* prediction approaches were used for repeat analysis, and RepeatMasker was applied to identify homologous sequences according to the RepBase (v_21.12) library [39]. The data from *ab initio*, homology-based, and transcriptome data evidence approaches were then combined for gene structure annotation. For the RNA sequencing used in the genome annotation, RNA sequencing reads were mapped to the genome using HISAT2 (v_2.1) program, and the transcripts were assembled via Cufflinks software [40]. The high-confidence gene models for the *M. lasiocarpa* genome were predicted by the MAKER pipeline tool. The protein functional annotation was performed by eggNOG-mapper and compared using BLASTP with data

stored in KEGG, DOG, GO, NR, and SwissProt databases. Potential telomeres and centromeres were identified using TeloExplorer and CentroMiner integrated into quarTeT [41].

Gene family and genome evolution analysis

Protein sequences in the genome are filtered by retaining the longest isomers and discarding sequences with fewer than 50 amino acids. Then, an all-against-all comparison via BLASTp with an E-value cutoff of 1e10-5 was performed [42], and the OrthoMCL (http://orthomcl.org/orthomcl/) was applied to cluster genes from these different species into gene families [43]. Expansions and contractions of orthologous groups were identified with Cafe´ (v 4.2, http://sourceforge.net/projects/cafehahnlab/). MCMC tree from PAML (v4.9j) was used for estimating the species divergence times. In order to identify genome synteny, the synteny blocks within the *M. lasiocarpa* genome and other species were identified by MCscanX. WGD analysis conducted using wgd [44].

HPLC/Q-TOF MS analysis of PhPNs

Samples of *M. lasiocarpa* seeds of different development stages were ground to powder prior to extraction. The 50 mg of the ground samples added 1 mL methanol and extracted for 35 min in an ultrasonic bath. The mixture was centrifuged for 10 min at 12,000 rpm. After centrifugation (12,000 rpm, 10 min) the supernatant was subjected to analysis that was accomplished with HPLC/Q-TOF-MS (Agilent, USA). Chromatography was carried out using an Agilent Poroshell C_{18} column, 4.6 mm × 100 mm length, and 2.7 μm pore size. The coulumn oven temperature was set to 35 °C. The parameters of gradient elution were set as follows: phase A (water with 1‰ formic acid), phase B (methanol): 5–100% B at 0–60 min, 100% B at 60–70 min, 100–95% B at 70–71 min, and 95% B at 71–90 min. The injection volume and flow rate were kept at10 μL and 1 mL/min respectively. Mass spectra were acquired using electrospray ionization (ESI) in the positive mode. The parameters of ESI source were set as follows: 10.0 L/min drying gas (N_2) flow; capillary voltage was 4.0 kV; temperature was 350 °C; fragmentation voltage was 170 V, and the nebulizerpressure was set to 50 psig. Mass spectral data were acquired in a scanning range from *m/z* 100 to 2000. For data collection and instrument control, the Qualitative Analysis B.05.00 software was applied.

Identification and characterization of methyltransferases

A combined methodology including Pfam searching and homologous alignment was used for discovering the methyltransferase genes in the genome of *M. lasiocarpa*. Two conserved domains, PF01596 and PF00891, were used for a genome-wide search by HMMER v_3.3 and BLAST by an e-value of 1e-5. The search was performed to compare the presence of homologs with library entries. Recombinant plasmids were constructed similarly to a previously reported method with minor modifications [45]. The Coding DNA Sequence of *MlOMT* candidate genes (Supporting Information)

were cloned into pMAL-c4x (EcoRI / SalI) to yield recombinant plasmids. The recombinant plasmids were further identified by Sanger sequencing, and they were then introduced into BL21 (DE3) for recombinant protein expression. Engineered strans *E. coli* harboring recombinant the plasmids were cultured in 50 mL Lysogeny broth (LB) medium at 37 ℃ for 3 h, then induced by 0.1 mM IPTG, followed by a further incubation at 16 ℃ for 24 h. The cells were harvested via centrifugation (6,000 rpm, 4 ℃ and 3 min) and the pellet resuspended in binding buffer (100 mM Tris-HCl under pH 7.5). The suspension was homogenized in an ultrasound bath for 20 min. Cell debris were subsequently removed by centrifugation at 6,000 rpm for 10 min. The supernatant was collected and purified by ÄKTA protein purification device. The relative activities of the crude recombinant enzymes were then measured. The reaction mixtures (100 μL batches) contained 100 mM Tris-HCl buffer, 10% glycerol, 1 mM β-mercaptoethanol, 2 mM SAM, and the respective crude enzyme. Heat-inactivated enzymes (inactivated at 100 ℃ for 10 minutes) were used as negative controls. The assay products were analyzed using high performance liquid chromatography/quadrupole time-of-flight mass spectrometry (HPLC/Q-TOF-MS). Enzyme assays determining the MlOMT's activities were measured using MTase-Glo™ methyltransferase assay according to previously published reports [46]. For example, a series of PhPNs along with the purified recombinant enzymes and 2 mM SAM were used to conduct kinetic analyses during incubation at 45 ℃ for 30 min. The reaction was quenched by adding 0.5% trifluoroacetic acid (TFA), followed by determination of the produced SAH by means of luminescence measurements. The kinetic parameters of *Km* and *Kcat* were calculated via the Michaelis–Menten tool implemented in GraphPad Prism (v_ 5).

Subcellular localization analysis

MlOMT genes were cloned into the expression vector (pBinPLUS.GFP4, with a CaMV 35S promoter and GFP), and the resulting recombinant plasmid was transferred into *Agrobacterium* MSU440 via the conventional freeze-thaw method. Next, the empty pBin-GFP and pBin-MlOMTs-GFP MSU440 plasmids were suspended in expression buffer (10 mM MES, 100 μM acetosyringone, and 10 mM $MgCl_2$) and subsequently infiltrated into *N. benthamiana* leaves (4-week-old). After the infiltrated *N. benthamiana* plants were kept in darkness for 48 h, leaf samples were collected and examined using a Laser scanning confocal microscope (LSCM).

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant number: 32070360, 32200326, and 22207047); Natural Science Foundation of Jiangsu Province (BK20220752 and BK20220749), and the Jiangsu Institute of Botany Talent Fund (JIBTF202304). We thank Prof. Zefu Wang (Nanjing Forestry University) for critical reading of the manuscript and Emily Wheeler for editorial assistance.

Author contributions

Y.C., and P.R.L. designed the experiments. W.L.Z. and J.Z.W. performed the experiments. M.T., S.H., Z.W. and G.Y.L. performed genomic analysis. S.X. and B.W. carried out antibacterial assays. L.T., R.Y.W, and B.Y. F. isolated and characterized the compounds. W.L.Z., J.Y.X. and Y.C. wrote the paper. C.P., H.L., and X.F reviewed and revised the paper. All authors contributed to discussion of the manuscript.

Data availability

Reference genome data and transcriptome sequence reads are available in GenBank under the project number PRJNA1009687.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Supplementary data

Supplemental information is available at *Horticulture Research* online.

References

1. Perrier, X. *et al.* Multidisciplinary perspectives on banana (Musa spp.) domestication. *Proc Natl Acad Sci U S A* 2011; **108**, 11311–11318.
2. FAOSTAT. Top 10 country production of bananas 2021. *Food and Agriculture Organization of United Nations* 2023.
3. Zorrilla-Fontanesi, Y. *et al.* Strategies to revise agrosystems and breeding to control Fusarium wilt of banana. *Nat Food* 2020; **1**, 599–604.
4. Bohra, A. *et al.* Reap the crop wild relatives for breeding future crops. *Trends in Biotechnology* 2022; **40**, 412–431.
5. García-Palacios, P. *et al.* Side-effects of plant domestication: ecosystem impacts of changes in litter quality. *New Phytol* 2013; **198**, 504–513.
6. Whitehead, S.R. & Poveda, K. Resource allocation trade-offs and the loss of chemical defences during apple domestication. *Ann Bot* 2019; **123**, 1029–1041.
7. Norman, E.O., Lever, J., Brkljača, R. & Urban, S. Distribution, biosynthesis, and biological activity of phenylphenalenone-type compounds derived from the family of plants, Haemodoraceae. *Nat Prod Rep* 2019; **36**, 753–768.
8. Flors, C. & Nonell, S. Light and singlet oxygen in plant defense against pathogens: phototoxic phenalenone phytoalexins. *Acc Chem Res* 2006; **39**, 293–300.
9. Chen, Y., Paetz, C. & Schneider, B. Precursor-directed biosynthesis of phenylbenzoisoquinolindione alkaloids and the discovery of a phenylphenalenone-based Plant defense mechanism. *J Nat Prod* 2018; **81**, 879–884.
10. Brand, S. *et al.* A type III polyketide synthase from *Wachendorfia thyrsiflora* and its role in diarylheptanoid and phenylphenalenone biosynthesis. *Planta* 2006; **224**, 413–428.
11. Liu, A.-Z., Kress, W.J. & Long, C.-L. The ethnobotany of *Musella lasiocarpa* (Musaceae), an endemic plant of southwest China. *Economic botany* 2003; **57**, 279–281.
12. Ma, H., Wang, D.-X., Li, T.-Q. & Li, Z.-H. Phylogeographic study of *Musella lasiocarpa* (Musaceae): providing insight into the historical river capture events. *Pak. J. Bot* 2019; **51**, 191–199.
13. Dong, L.-B. *et al.* Chemical constituents from the aerial parts of *Musella lasiocarpa*. *Nat. Prod. Bioprospect* 2011; **1**, 41–47.

14. Lyu, H., Chen, Y., Gershenzon, J. & Paetz, C. Diarylheptanoid derivatives (Musellins A-F) and dimeric phenylphenalenones from seed coats of *Musella lasiocarpa*, the Chinese dwarf banana. *J Nat Prod* 2023; **86**, 1571–1583.
15. Cui, M.Y. *et al.* Two types of *O*-methyltransferase are involved in biosynthesis of anticancer methoxylated 4'-deoxyflavones in Scutellaria baicalensis Georgi. *Plant Biotechnol J* 2022; **20**, 129–142.
16. Tan, Y. *et al.* Identification and characterization of two Isatis indigotica *O*-methyltransferases methylating C-glycosylflavonoids. *Hortic Res* 2022; **9**, uhac140.
17. Zhao, W. *et al.* Identification and characterization of methyltransferases involved in benzylisoquinoline alkaloids biosynthesis from *Stephania intermedia*. *Biotechnol Lett* 2020; **42**, 461–469.
18. Otálvaro, F. *et al. O*-Methylation of phenylphenalenone phytoalexins in *Musa acuminata* and *Wachendorfia thyrsiflora*. *Phytochemistry* 2010; **71**, 206–213.
19. Hidalgo, W. *et al.* Structure−activity relationship in the interaction of substituted perinaphthenones with *Mycosphaerella fijiensis*. *J. Agric. Food Chem* 2009; **57**, 7417-7421.
20. Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A. & Zdobnov, E.M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 2021; **38**, 4647–4654.
21. Wang, Z. *et al.* A chromosome-level reference genome of *Ensete glaucum* gives insight into diversity and chromosomal and repetitive sequence evolution in the Musaceae. *Gigascience* 2022; **11**.
22. Wang, Z. *et al.* Musa balbisiana genome reveals subgenome evolution and functional divergence. *Nat Plants* 2019; **5**, 810–821.
23. Lu, R. *et al.* Chromosome-level genome assembly of a fragrant Japonica rice cultivar 'Changxianggeng 1813' provides insights into genomic variations between fragrant and non-fragrant Japonica rice. *Int J Mol Sci* 2022; **23**.
24. D'Hont, A. *et al.* The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 2012; **488**, 213–217.
25. Wang, Z.F., Rouard, M., Droc, G., Heslop-Harrison, P.J.S. & Ge, X.J. Genome assembly of *Musa beccarii* shows extensive chromosomal rearrangements and genome expansion during evolution of Musaceae genomes. *Gigascience* 2022; **12**.
26. Vanneste, K., Baele, G., Maere, S. & Van de Peer, Y. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Res* 2014; **24**, 1334–1347.
27. Cui, G. *et al.* Meliaceae genomes provide insights into wood development and limonoids biosynthesis. *Plant Biotechnol J* 2023; **21**, 574–590.
28. Miao, Y. *et al.* Genome sequencing reveals chromosome fusion and extensive expansion of genes related to secondary metabolism in *Artemisia argyi*. *Plant Biotechnol J* 2022; **20**, 1902–1915.
29. Liu, X. *et al.* Systematic analysis of *O*-methyltransferase gene family and identification of potential members involved in the formation of *O*-methylated flavonoids in Citrus. *Gene* 2016; **575**, 458–472.
30. Pegg, K.G., Coates, L.M., O'Neill, W.T. & Turner, D.W. The Epidemiology of Fusarium Wilt of Banana. *Front Plant Sci* 2019; **10**, 1395.
31. Krishnamurthy, P. *et al.* Phenylphenalenone-type phytoalexins in banana (Musa species): a comprehensive review for new research directions. *Phytochem. Rev.* 2023; **22**, 187–210.
32. Liang, Z., Zhang, P., Xiong, Y., Johnson, S.K. & Fang, Z. Phenolic and carotenoid characterization of the ethanol extract of an Australian native plant *Haemodorum spicatum*. *Food Chemistry* 2023; **399**, 133969.
33. D'Hont, A. *et al.* The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 2012; **488**, 213–217.

34. Wang, S., Alseekh, S., Fernie, A.R. & Luo, J. The structure and function of major plant metabolite modifications. *Mol Plant* 2019; **12**, 899–919.
35. Lazzaro, A. *et al.* Light- and singlet oxygen-mediated antifungal activity of phenylphenalenone phytoalexins. *Photochem Photobiol Sci* 2004; **3**, 706–710.
36. Song, X. *et al.* Synthesis and biological activities assessment of 4-, 6-, and 9-Phenylphenalenone derivatives. *ChemistrySelect* 2022; **7**, e202203793.
37. Lan, L. *et al.* A high-quality *bougainvillea* genome provides new insights into evolutionary history and pigment biosynthetic pathways in the Caryophyllales. *Hortic Res* 2023; **10**, uhad124.
38. Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* 2017; **356**, 92–95.
39. Flynn, J.M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci* 2020; **117**, 9451–9457.
40. Kim, D., Paggi, J.M., Park, C., Bennett, C. & Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019; **37**, 907–915.
41. Lin, Y. *et al.* quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Hortic Res* 2023; **10**.
42. Liao, B. *et al.* Allele-aware chromosome-level genome assembly of Artemisia annua reveals the correlation between ADS expansion and artemisinin yield. *Mol plant* 2022; **15**, 1310–1328.
43. Li, L., Stoeckert, C.J. & Roos, D.S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003; **13**, 2178–2189.
44. Zwaenepoel, A. & Van de Peer, Y. wgd—simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* 2018; **35**, 2153–2155.
45. Zhao, W. *et al.* Biosynthesis of plant-specific alkaloids tetrahydroprotoberberines in engineered Escherichia coli. *Green Chem* 2021; **23**, 5944–5955.
46. Liu, X. *et al.* Characterization of a caffeoyl-CoA *O*-methyltransferase-like enzyme involved in biosynthesis of polymethoxylated flavones in Citrus reticulata. *J Exp Bot* 2020; **71**, 3066–3079.

Table 1. The main genome assembly features of *M. lasiocarpa*

Assembly feature	Number	Length (Mb)
Total contigs	440	509.15
Contig N50	5	56.62
Contig N90	9	–
Total scaffolds	431	509.15
Scaffold N50	5	56.62
Scaffold N90	9	–
Pseudochromosomes	9	–
Repetitive sequences	51.56%	–
Protein-coding genes	34,361	–

Figures and Legends

Figure 1. Sequencing samples, overview genome assembly of *M. lasiocarpa* and PhPNs structure. Photograph of *M. lasiocarpa* (A), types of PhPNs (B, R^1 and R^2, -OH or -OCH_3), and distribution of *M. lasiocarpa* genomic features (C). The rings from the outside to the inside indicate nine chromosomes, gene density, repeat density, GC contents, and syntenic genomic blocks in section C.

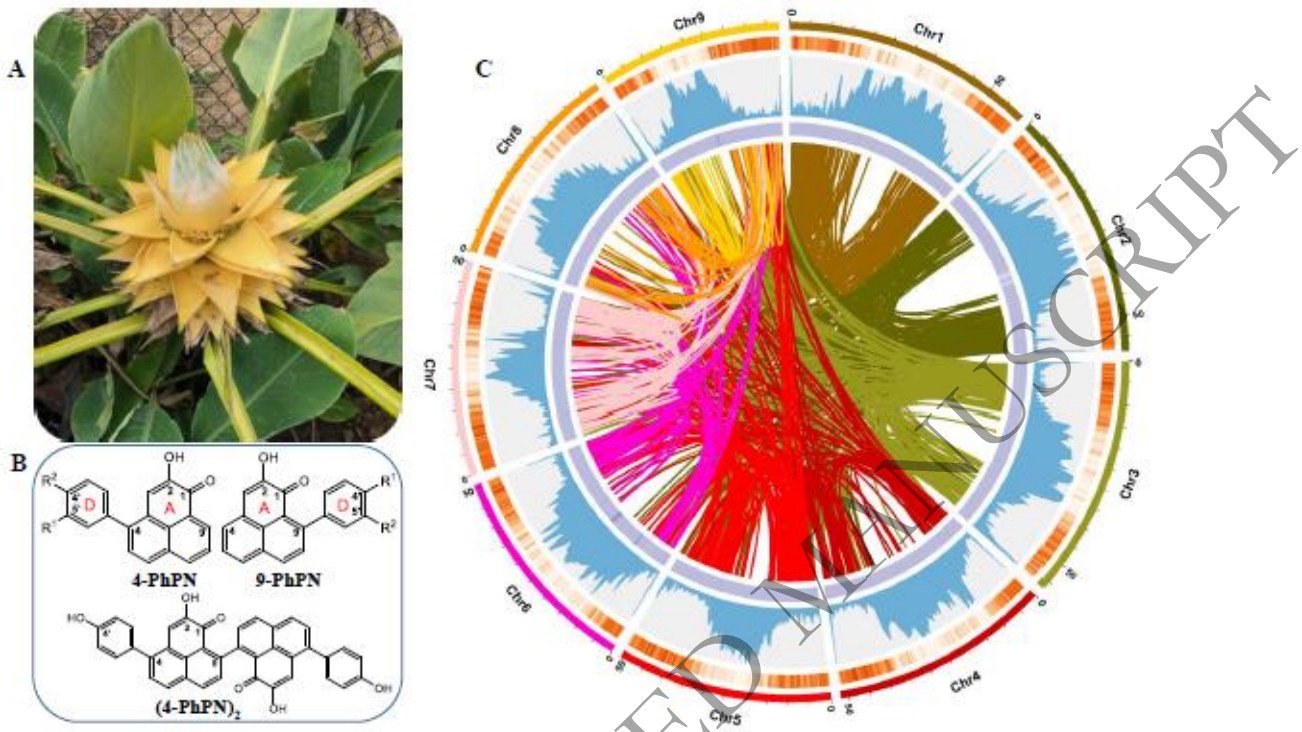
Figure 2. Evolutionary analyses of *M. lasiocarpa*. Phylogeny and divergence time of 12 angiosperms (A). WGD analysis of four species (B).

Figure 3. Expression heatmap and phylogeny of *MlOMT* genes. (A) The expression of 30 MlOMT genes in *M. lasiocarpa* seeds during three developmental stages (yellow seed, S2; brown seed, S4 and block seed, S6) based on the transcriptomic data. The FPKM values of Ml01G0494, Ml04G2958, and Ml08G0855 (marked in red) were greater than 100. The expression level was measured by FPKM. Eight genes (Ml01G2217, Ml01G2223, Ml01G2225, Ml01G2224, Ml02G2158, Ml01G1752, Ml01G1753, and Ml06G1505) that were not expressed in all samples were not listed. (B) Phylogeny of OMTs from 12 genomes and characterized enzymes.

Figure 4. Results of the MlOMT assays with PhPNs. A-D and F-H, Schematic diagram of catalytic reaction between MlOMTs and PhPNs. E and I, the HPLC/Q-TOF MS results of representative substrates MLT4 and MLT9 reacted with Ml01G0494, Ml04G2958, and Ml08G0855, respectively.

Figure 5. Kinetic properties of recombinant MlOMTs with different PhPN substrates. Kinetic parameters were estimated by nonlinear curve fitting using Michaelis–Menten. The concentration of SAH generated during the substrate reaction was assayed.

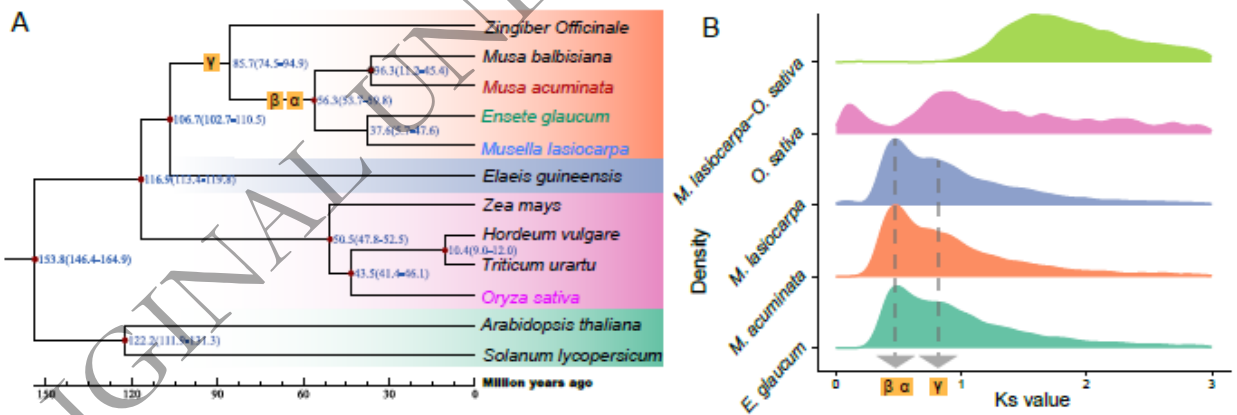
580 **Fig. 1.**



581

582

583 **Fig. 2.**



584

585

586

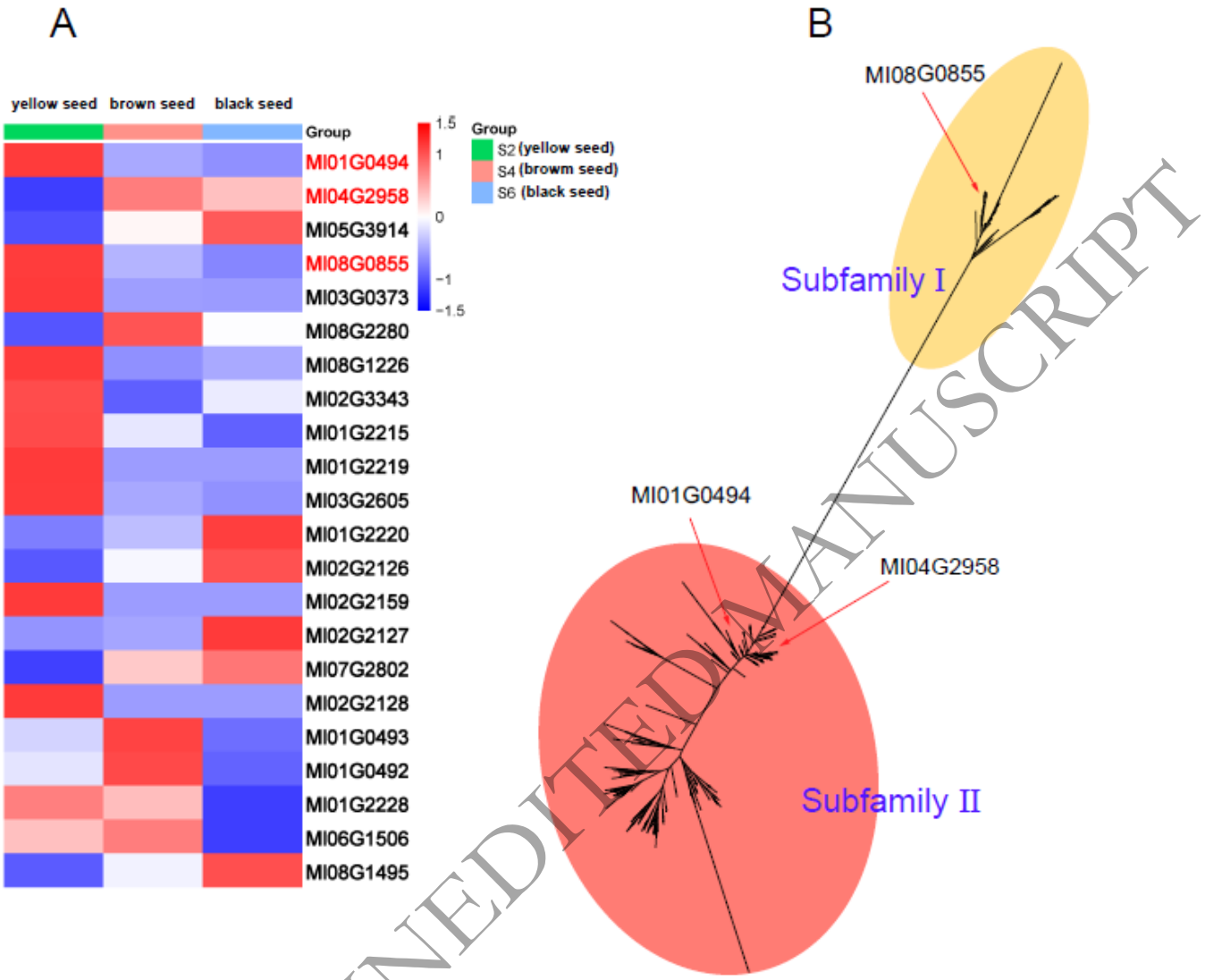
587

588

589

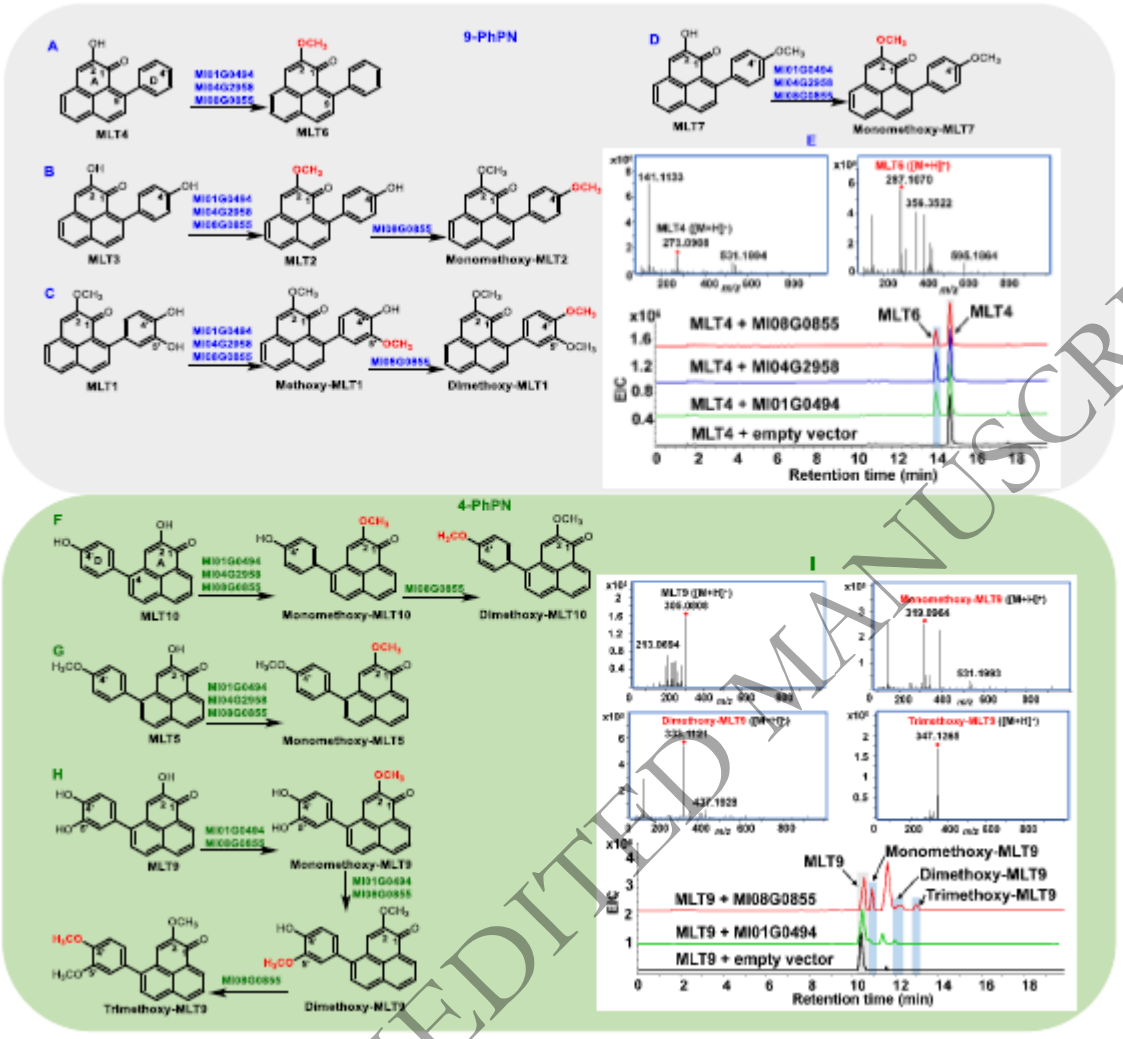
590

591 **Fig. 3.**

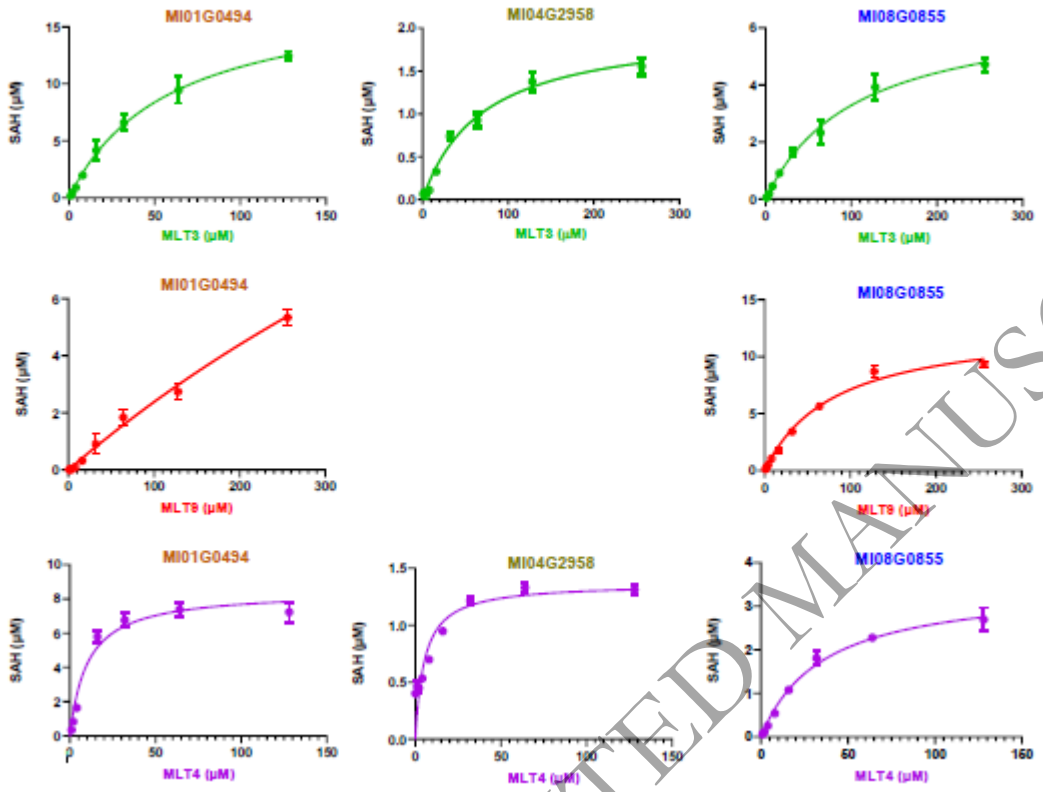


592
593
594
595
596
597
598
599
600
601
602
603
604
605

Fig. 4.



622 **Fig. 5.**



623

624