Systematic mining of fungal chimeric terpene synthases using an efficient precursor-providing yeast chassis

Rong Chen (陈蓉)a, Qidong Jiab, Xin Muc, Ben Hu (胡奔)c, Xiang Sunb, Zixin Denga,d, Feng Chenb,d,1, Guangkai Biana (卞光凯)a,1, and Tiangang Liua (刘天罡)a,c,1,8

aKey Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education, Wuhan University School of Pharmaceutical Sciences, Wuhan 430071, People’s Republic of China; bGenome Science and Technology Graduate Program, University of Tennessee, Knoxville, TN 37996; cHubei Engineering Research Center of Materia Medica Extraction and Processing, School of Pharmaceutical Sciences, Huazhong University of Science and Technology, Wuhan 430070, People’s Republic of China; and dDepartment of Plant Sciences, University of Tennessee, Knoxville, TN 37996

Edited by Rodney B. Croteau, Washington State University, Pullman, WA, and approved June 07, 2021 (received for review December 10, 2020)

Chimeric terpene synthases, which consist of C-terminal prenyltransferase (PT) and N-terminal class I terpene synthase (TS) domains (termed PTTSs here), is unique to fungi and produces structurally diverse di- and sesterterpenes. Prior to this study, 20 PTTSs had been functionally characterized. Our understanding of the origin and functional evolution of PTTS genes is limited. Our systematic search of sequenced fungal genomes among diverse taxa revealed that PTTS genes were restricted to Dikarya. Phylogenetic findings indicated different potential models of the origin and evolution of PTTS genes. One was that PTTS genes originated in the common Dikarya ancestor and then underwent frequent gene loss among various subsequent lineages. To understand their functional evolution, we selected 74 PTTS genes for biochemical characterization in an efficient precursor-providing yeast system employing chassis-based, robot-assisted, high-throughput automatic assembly. We found 34 PTTS genes that encoded active enzymes and collectively produced 24 di- and sesterterpenes. About half of these di- and sesterterpenes were also the products of the 20 known PTTSs, indicating functional conservation, whereas the PTTS products included the previously unknown sesterterpenes, sesterevins (1), and sesterbiculene (2), suggesting that a diversity of PTTS products awaits discovery. Separating functional PTTSs into two monophyletic groups implied that an early gene duplication event occurred during the evolution of the PTTS family followed by functional divergence with the characteristics of distinct cyclization mechanisms.

Significance

Chimeric terpene synthases, termed PTTSs, are a unique family of enzymes occurring only in fungi. Characterizing PTTSs is challenging due to the complex reactions they catalyze and the structural complexity of their products. Here, by devising an efficient precursor-providing yeast chassis and incorporating a high-throughput automated platform, we identified 34 active PTTSs, which was considerably more than the number of known functional PTTSs. This effective and rapid pipeline can be employed for the characterization of other PTTSs or related terpenoid biosynthetic enzymes. By systematically analyzing the presence/absence of PTTS genes together with phylogenetic analysis, the ancestral PTTS gene was inferred to have undergone duplication and functional divergence, which led to the development of two distinct cyclization mechanisms.
5/5/6/5 pentacyclic quinannulatene (19), and 5/6/5/5 tetracyclic homopsene (20) (Fig. 1). According to the initial carbocation formation strategy, the cyclization mechanisms of PTTSs can be classified into a type A (C1-IV-V) mode in which sequential cyclization is initiated between the C1-C15/C14-C18 of geranylfernaryl diphosphate (GFPP) to yield a five to 15 ring system and a type B (C1-III-IV) mode in which a five to 11 ring system is uniformly generated by cyclization at the C1-C11/C10-C14 of GFPP/geranylgeranyl diphosphate (GGPP) (14, 21).

The accumulating number of PTTSs requiring functional characterization has generated several fundamental questions that await answers, such as whether PTTS genes of interest are widely distributed among fungi or restricted to specific groups and which mechanisms govern their functional evolution. Numerous PTTSs need to be characterized to answer these questions, and this requires an efficient high-throughput system, although Aspergillus oryzae is widely used for heterologous expression and it has facilitated the characterization of a series of PTTSs from filamentous fungi (18, 22). However, rapidly characterizing the functions of numerous unknown PTTSs would be impossible using this system. The development of robust terpene precursor-providing chassis in Escherichia coli and Saccharomyces cerevisiae has provided efficient approaches for the overproduction of farnesene, taxadiene, and artemisinic acid (23–25) and a scalable strategy for rapidly characterizing the functions of TSs, accelerating the process of mining novel sesqui-, di-, and sesterterpenes with unusual skeletons (26–29).

In this study, we systematically collected PTTS genes from various fungal species and propose an evolutionary model in which PTTS genes have a common origin in the Dikarya ancestor. Gene duplication and frequent gene loss are among the inferred mechanisms governing PTTS gene evolution. We then determined the catalytic functions of 34 PTTSs using a robot-based automatic high-throughput assembly platform and an efficient precursor-providing S. cerevisiae chassis. Thus, our findings provide profound insights into the origin and functional evolution of PTTSs, a class of TS genes that are unique to fungi.

Results
Mining and Phylogenetic Analysis of PTTSs. We exhaustively searched the sequenced genomes of 477 fungi as well as genes deposited in the National Center for Biotechnology Information (NCBI) and
UniProt databases. We identified 227 unique PTTS genes in 139 fungal species (SI Appendix, Fig. S2), including 20 that have already been characterized (SI Appendix, Table S1). Among these, 224 unique PTTS genes were identified in Ascomycota, and three were identified in Basidiomycota. However, genome sequencing revealed that PTTS genes were not ubiquitous in any fungal lineage (SI Appendix, Fig. S2). We identified PTTS genes in all Ascomycota lineages except Pezizomycetes (nine species), Saccharomyces (40 species), and Taphrinomycotina (seven species). Within Basidiomycota, we detected PTTS genes in only one of the 138 species of Agaricomycotina, which contained three PTTS genes, but not in Ustilaginomycotina (16 species) or Pucciniomycotina (22 species). No PTTS genes were detected in Zygomycota (17 species) or any other basal fungal lineages (15 species).

To understand their evolutionary relatedness, a phylogenetic analysis of all 227 PTTS genes was performed. Prior to this study, a few studies had investigated the phylogeny of small datasets of PTTS genes, which led to the classification of six subfamilies (A to F) (17, 18, 21, 30). Our phylogenetic analysis using a larger and more comprehensive dataset substantiated these findings (Fig. 2), and we designated them as PTTS subfamilies A to F. Our findings also supported bifurcation of the PTTS genes with subfamilies A, E, and F forming Clade I and those from the remaining subfamilies forming Clade II. Known PTTSs in Clades I and II catalyze type A (C1-IV-V) and B (C1-III-IV) cyclization (4–21), respectively. Clade II notably contains PTTS genes from seven lineages of fungi containing PTTSs, whereas Clade I contains PTTS genes from five lineages (Fig. 2). Within each of the six subfamilies, >90% of the PTTS genes remain functionally unknown, attesting to the large gap in knowledge regarding the catalytic functions of PTTSs.

High-Throughput Mining of Active PTTSs Using an Efficient Precursor-Providing Yeast Chassis. Compared with other classes of TSs, PTTSs are far more challenging to characterize due to the complexity of the reactions they catalyze and the structural complexity of their products. We implemented a metabolic engineering strategy to create a yeast sesterterpenoid overproduction chassis by overexpressing GFPP synthase (GFPPS) and PTTS assemblies in engineered S. cerevisiae YZL141 (13). To functionally characterize numerous PTTSs, we facilitated structural characterization by fusing the PTTSs to GFPPS using a linker peptide GSTGS (31) that decreases the distance between them and enhanced the metabolic flux of GFPP toward sesterterpene synthases. Introducing GFPPS can also unlock the GFPP utilization ability of chimeric diterpene synthases to generate diverse sesterterpenes. In selecting candidate PTTSs for functional characterization, we considered amino acid (aa) sequence length (~700 aa), sequence similarity (<80%), and the characteristics of the conserved motifs (DDXXD/E and NSE/DTE for TS and DDXXD/N for PT domains). As a result, we selected 74 candidate PTTSs. To characterize the functions of PTTSs in S. cerevisiae, the 74 candidates were codon optimized, synthesized, and cloned into pYJ117 to produce corresponding plasmids that were transformed using the robot-based automated high-throughput platform to obtain yeast mutants harboring PTTS genes (SI Appendix, Fig. S3). The 74 engineered strains were obtained from gene fragments within only 10 d, which greatly accelerated TS mining. The engineered S. cerevisiae strains were cultivated in 24-well plates for 72 h. Gas chromatography–mass spectrometry (GC-MS) results indicated that 34 of the 74 analyzed PTTS genes encoded active TSs, including six with only diterpene synthase, 24 with only sesterterpene, and four with bifunctional di- and sesterterpene synthase activities (SI Appendix, Table S3).

Terpene Products of 34 Active PTTSs and Function–Phylogenetic Relationships of Functional PTTSs. The high-throughput assays showed that 34 PTTSs encoded active TSs to produce a total of 24 di- and sesterterpenes, for which titers were adequate in 18 of them. We then determined their structures via spectroscopic analysis (Fig. 3). The structures of the remaining six compounds were proposed by comparing their GC-MS spectra with reported data (SI Appendix, Fig. S6). Among the 24 terpene products [designated as products (1)–(24)], 20 were cyclic, four were acyclic, and 11 were the products of known PTTSs (Fig. 3). Among these, seven known sesterterpenes were produced by 12 new PTTSs. The known diterpenes were produced by eight new PTTSs, and the remaining 13 di- or sesterterpenes were generated by novel PTTS with previously unknown catalytic activities. Two of the cyclic terpenes identified as new PTTS products were novel sesterterpenes and designated as sesitervesene (1) and sesterboriculene (2).

Among the four acyclic products, β-geranylfernesene (5) (32) and geranylfernesol (6) (33, 34) were sesterterpenes and geranylgeraniol (24) (35, 36) and geranylialool (23) (37, 38) were diterpenes (Fig. 3). These acyclic products could have resulted from ionization of the allylic diphosphate and either deprotonation of the initial carbocation or trapping with water. To exclude the possibility that these products arose from altered/misfolded PTTSs linked with GFPPS, their respective genes, PTTS013, 021, 051, 052, and 066, were overexpressed in S. cerevisiae as GFPPS fusions and as individual proteins. The results showed that PTTS-GFPPS fusion did not influence the production of these compounds. In fact, this strategy helped to determine the ability of the diterpene synthases PTTS013, 052, and 066 to use GFPP for geranylfernesol production (SI Appendix, Fig. S7).

Some of the sesterterpenes have previously been shown to be produced by plants or bacteria. Previous studies have reported that plants can produce fungal-type sesterterpenes (33). Several fungal PTTSs can also produce plant and bacterial terpenes. The GC-MS data showed that PTTS051 can produce variculatriene A (3), which is usually derived from plants, providing an additional source for the high-yield production of variculatriene A and its bioactive derivative variculanol (33, 39). Each of PTTS040, 050, 067, 086, and 106 produced the known sesterterpene (-)-variculatriene B (7), which is also produced by the ArtIPS19Y260 variant of the plant Arabidopsis thaliana (33). Likewise, PTTS036 produced fusaproliferene (10), another product of the ArtIPS19Y260 variant (33). PTTS125 produced brassitriene A (8) and brassitriene B (9), which are unified 15/3 bicyclic sesterterpene intermediates of three plant diterpene synthases (Cr089 from Capsella rubella, ArtIPS17 from A. thaliana, and Br580 from Brassica napus) (40). We thus observed eight PTTSs that can catalyze the synthesis of plant-derived variculatriene A (3), (-)-variculatriene B (7), brassitriene A (8), brassitriene B (9), and fusaproliferene (10). The diterpene synthase PTTS017 produced spiroviolene (21), which is also produced by spiroviolene synthase in bacterium Streptomyces violenc (41). As introduced earlier, cyclic di- and sesterterpene products of PTTSs are formed through either type A (C1-IV-V) or type B (C1-III-IV) cyclization. The known enzymes catalyzing types A and B cyclization belong to two separate clades (Fig. 2). To gain further insights into this function–phylogeny relationship, we created a phylogenetic tree with the 34 new PTTSs and the 20 known PTTSs. Consistent with previous results and our findings from a larger dataset (Fig. 2), all active PTTSs were grouped into the six subfamilies (A to F; Fig. 3). Among the novel active PTTSs, 19 that located with subfamilies A, E, and F in Clade I were characterized by the C1-IV-V cyclization mode (type A), the remaining 15 PTTSs that located with subfamilies B to D in Clade II were characterized by the C1-III-IV cyclization mode (type B; Fig. 3). A number of observations were made from these results. First, the cyclic products of PTTSs from clades I and II were respectively produced through type A and type B cyclization, which was consistent with previous findings (14). Second, most of the 34 active PTTSs were single-product enzymes. Third, some closely related PTTSs had the same catalytic function.
which was not surprising. For example, PTTS040, 050, 067, 086, and 106 that clustered within subfamily A, all appeared to be (-)-variculatriene B synthases. Fourth, PTTSs that shared a high identity generated different products, and those with a low identity generated similar products. For example, PTTS042 and 090, which respectively shared 36% and 94% identity with PvPS, produced the sesterterpene skeleton preasperterpenoid A (17). PTTS023, 107, 110, and 119 shared 40 to 64% identity with EvVS and produced the same sesterterpene skeleton, variediene (22) (Fig. 3 and SI Appendix, Table S3) (42). These data suggested that only a few amino acid residues near the active pocket regulate the catalytic activity of the TS domain. The last observation was that PTTSs in Clades I and II produced acyclic products. Although such PTTSs in Clade I produced only acyclic sesterterpenes, those in Clade II produced acyclic sesterterpenes and acyclic diterpenes (Fig. 3).

**Detailed Characterization of PTTSs with Previously Unknown Sesterterpene Products.** Bioinformatic data revealed that the chimeric enzyme PTTS010 (ZbSS), isolated from *Zymoseptoria brevis* and shared 46% identity with EvQS, was grouped in Clade I-F (Fig. 3). Sequence alignment indicated that the highly conserved Asp-rich motifs 95DDYYD and 230NDCHSWPKE were located in the TS domain, and the Asp-rich motifs 487DDIED and 614DDYQN were located in the PT domain. PTTS010 produced a new compound (1) with a characteristic sesterterpene molecular ion at m/z 340. Its molecular formula was confirmed as C_{25}H_{40} based on a high-resolution electron ionization mass spectrometry (HR-EI-MS) ion peak [M]+ at m/z 340.3126 (calculated for 340.3130). The 13C NMR spectrum revealed 25 signals, including four olefinic carbons, suggesting a tetracyclic skeleton. The 1H and 13C NMR data showed the same planar structure of compound 1 (SI Appendix, Table S5) and aspergildiene A (43). The 13C NMR chemical shifts of C-6, C-17, and C-24 of 1 differed from those of aspergildiene A, indicating that the relative configuration differs between these compounds. The relative configuration of 1 was determined by NOESY correlation analysis and NMR chemical shift calculations. The NOESY correlations of H-12/H-6, H-6/H-21, H-12/H-14, H-23/H-2, and H-2/H-7 confirmed the relative configurations of C-2, C-6, C-7, C-12, C-14, and C-15 (SI Appendix, Fig. S8). The relative configuration of C-18 was determined by 13C NMR and 1H NMR chemical shift calculations (SI Appendix, Tables S15 and S16). Its absolute configuration was assigned as 2'S, 6'S, 7'R, 12'S, 14'S, 15'R, and 18'R through Electronic Circular Dichroism (ECD) calculation (SI Appendix, Fig. S9). Thus, compound 1 was identified as a previously unknown 5/8/6/5 tetracyclic sesterterpene, named sesterevisene (1).
and the sesterterpene synthase was designated as Z. brevis sester-
evisene synthase (ZbSS). We proposed the following cyclization
mechanism for sestertervisene (1) based on those of 5/8/6/5 scaffolds
(Fig. 5) as follows. Initiated by the elimination of a pyrophosphate
group from GFPP, two successive cyclizations at C1 to C15 and C14
to C18 yield the bicyclic cation A-2+ with a fused ring system. A subsequent 1,5-H shift from C12 to C19 affords cation A-3+.
A second C6 to C10 cyclization yields cation A-4+ with a tricyclic
system. Two sequential 1,2-H shifts, followed by a third C2 to C12
cyclization, form the tetracyclic cation A-7+, producing 1 (Fig. 5).

The phylogenetic findings showed that PTTS037 (CoSS)
clustered and shared 70 and 77% identity with PTTS009 and 054,
respectively (SI Appendix, Table S3). The GC-MS data showed
that PTTS009, 037, and 054 produced an identical sesterterpene
hydrocarbon 2 with the formula C25H40 as determined by an
HR-EI-MS [M]+ ion at m/z 340.3137 (calculated for 340.3130).
Its ultraviolet (UV) absorption at 241 nm implied the presence of
a conjugated double bond. The structure of 2 was subsequently
elucidated by NMR spectra (SI Appendix, Table S6). A compari-
son of its 1H and 13C spectra showed that 2 was closely related to
Bm3 (21). The only difference between the two structures was that
2 possessed an additional degree of unsaturation and one less
hydroxyl group than Bm3, which was confirmed by heteronuclear
multiple bond correlations (HMBCs) from H-9 (δH 5.45, doublet

Fig. 3. Phylogenetic analysis of characterized PTTSs and resulting products. (Left) Phylogenetic tree grouped 34 novel and 20 known PTTSs into six sub-
families to form Clades I (A, E, and F) and II (B to D). Accession numbers for active PTTSs are provided in SI Appendix, Table S3. (Right) Structure and cyclization
models for novel products. Blue triangles represent characterized enzymes. Red squares represent enzymes assessed herein. The white triangle represents the
plant-derived AtTPS25 outgroup. The ‘#’ indicates structures of compounds proposed by comparing GC-MS spectra with literature. The ‘**’ indicates structures of compounds produced by previously reported PTTSs and the newly characterized PTTSs.

Chen et al. PNAS | 5 of 9
Systematic mining of fungal chimeric terpene synthases using an efficient precursor-
providing yeast chassis
https://doi.org/10.1073/pnas.2023247118
of triplets, 11.4, 8.3) to C-11 (δ_C 129.7), H-10 (δ_C 5.87, d, 11.4) to C-22 (δ_C 17.0), and 1H-1H COSY correlations of H2-8 (δ_H 2.76, m) / H-9 / H-10. We determined the planar structure and relative configuration of 2 by extensive two-dimensional NMR spectral analysis together with the above data (SI Appendix, Fig. S8). The absolute configuration of 2 was determined by the chemical elimination of Bm3 mixed with p-TsOH/toluene at room temperature for 10 h and the high-performance liquid chromatography (HPLC) and GC-MS detection of the conversion to 2 (SI Appendix, Fig. S10). Therefore, 2 was designated as a previously unknown sesterterpene and named sesterorbiculene, and the sesterterpene synthase was designated Colletotrichum orbiculare sesterorbiculene synthase (CoSS). Based on the cyclization mechanism of Bm3, we propose that the cyclization of 2 was initiated by sequential cyclization between the C1-cation, C14-C15 olefin (IV), and C18-C19 olefin (V) of GFPP [type A: C1-IV-V], which was consistent with the phylogenetic results (Fig. 5).

**Discussion**

In this study, we systematically investigated the origin and functional evolution of PTTSs, a family of TSs that is unique to fungi. We found PTTS genes in Dikarya and not in any basal fungi lineages (SI Appendix, Fig. S2). Within Dikarya, PTTS genes were found in most lineages of Ascomycota but in only one of 174
species of Basidiomycota. One explanation for this is that the PTTS gene originated in the common ancestor of dikarya, but genes were lost during diversification, particularly in Basidiomycota. An alternative scenario is that the single species of Basidiomycota containing PTTSs acquired them through other mechanisms, such as horizontal gene transfer. In the latter scenario, PTTS genes would have originated in Ascomycota. PTTS genes of Ascomycota exhibited a pattern of bifurcation, occurring in both Clade I and Clade II, except for a single gene from Lecanoromycetes, which only occurred in Clade I, and all five genes from Orbiliomycetes, which only occurred in Clade II. The absence of Clade I PTTS genes in Orbiliomycetes and Clade II genes in Lecanoromycetes could be caused by gene loss. As Clade II contained PTTS genes from a wider range of lineages than Clade I, the Clade II genes might have evolved first and an ancient duplication event led to the Clade I genes. These different models have important implications for the evolution of the catalytic mechanisms of PTTS enzymes, as they indicate that type B cyclization catalyzed by Clade II PTTSs evolved first, and type A cyclization evolved through gene duplication and functional divergence. Continued identification of PTTS genes from a wider range of fungal taxa and their functional characterization would provide additional evidence regarding the origin and functional evolution of PTTS genes.

We used a high-throughput automated platform together with an efficient yeast chassis to generate precursors and characterized 34 active fungal PTTSs that produced two new sesterterpenes and 22 known terpenes. We inferred from a comparison of their catalytic activities in the contexts of a sequence–function and phylogeny–function relationship that several mechanisms were responsible for the functional evolution of PTTSs. One such mechanism was through changes in protein sequences leading to functional divergence as evidenced by 37 different terpene products produced by 54 active PTTSs. More evidence to support this mechanism is from the identification of PTTSs producing acyclic products. These PTTSs might have evolved from a cyclic terpene-producing progenitor. A similar scenario has been proposed for typical plant TSs. The ancestral plant TS was probably a cyclase from which acyclic terpene producing enzymes evolved (44). An other possibility is that convergent evolution could have been responsible for distantly related PTTSs producing the same product. Therefore, our study not only adds to the classification of cyclization patterns of PTTSs but also shows the cyclic nature of GGPP and GFPP to produce diverse di- and sesterterpenes but also reveals diverse mechanisms underlying their functional evolution. The precursor-providing yeast platform efficiently excavated the functions of chimeric sesterterpenes, enabling these observations. That 40 of the selected PTTS did not produce di-/sesterterpenes using the engineered yeast chassis is notable. Nonfunctional PTTSs might have been a consequence of incorrect exon/intron predictions, inclusion of body formation in S. cerevisiae, mutations in the conserved catalytic motif, or the incomplete evolution of functional genes (45–47). This requires clarification, and further investigation is needed to link the catalytic activity of the 34 active PTTSs including the acyclic terpene producers to the production of specific terpenes in the source fungus. Once such links are established, it will also be interesting to study the biological functions of these PTTSs and the sesterterpenes or diterpenes that they produce.

Interestingly, six of the terpene products of fungal PTTSs are also produced by plants or bacteria. They are produced by typical plant TSs, which are only distantly related to microbial TSs (33). The plant sesterterpene synthases are newly derived members within the TPS-a subfamily, the general function of which is sesterterpene biosynthesis (48). This presents a case of neofunctionalization. Bacterial TSs are more closely related to PTTSs, particularly the PTTS TS domain. The observation that identical terpenes can be produced by PTTSs in fungi and by classic TSs in plants/bacteria provides clear evidence of functional convergent evolution, a mechanism frequently observed in the evolution of plant secondary metabolism (49). The discovery of these PTTSs has expanded the known sources of these terpenes and provides a solid foundation for the high-yield production of terpene skeletons and their derivatives with diverse pharmacological properties.

From a methodological perspective, we developed an efficient approach for the heterologous expression of synthetic PTTS genes in an efficient precursor-providing yeast chassis. Numerous yeast mutants were generated from gene fragments within ~10 d using a high-throughput automated platform. This avoided manual labor and accelerated research progress. Our findings showed that the combination of an efficient precursor-providing yeast chassis and high-throughput automation platform is an effective and rapid pipeline for characterizing PTTSs. This system can also be used to characterize other enzyme families that produce terpenes.

### Materials and Methods

#### Strains and Media

The primers used in this study are listed in SI Appendix, Table S23, and details on the plasmids and strains used are provided in SI Appendix, Tables S25 and S32. The E. coli strain DH5α was used to transform plasmids and propagate all plasmids. S. cerevisiae YL141 was cultivated in yeast extract peptone dextrose (YPD) medium and used as a platform for the heterologous expression of PTTSs, providing the precursors IPP and DMAPP, and facilitating the overproduction of terpene products.

#### Identification and Phylogenetic Analysis of Fungal PTTSs

Two approaches were applied to identify candidate PTTS genes. The first was to search for putative PTTS genes in the 519 sequenced fungal genomes in the Joint Genome Institute (https://genome.jgi.doe.gov/projects/fungi/index.jsf) as of October 2015. Putative TSs were identified by searching all predicted amino acid sequences containing HMMER against the Pfam database (http://pfam.xfam.org, version 28.0). Sequences containing both a terpene synthase (C domain (PF00936) and a polyterpene synthase (PF00348) were identified as putative PTTS genes. The second approach was to search for PTTS genes in the NCBI and UniProt databases. Putative TSs of the appropriate amino acid sequence length (~700) with conserved motifs in their PT and TS domains were selected as candidates for functional characterization. These two sets were then combined, and all PTTSs with similarity <80% were retained. In total, the final set included 74 candidate PTTSs (SI Appendix, Table S2). Multiple sequence alignments were produced in Mafft using a highly accurate setting (L-INS-i) and 1,000 iterations of improvement. Maximum likelihood phylogenetic trees were inferred produced in MAFFT using a highly accurate setting (L-INS-i) and 1,000 iterations of improvement. Maximum likelihood phylogenetic trees were inferred using FastTree (version 2.1.7) using a high accuracy setting (−spr 4 -mlacc 2 -smp 100). All trees were rendered using FigTree (version 1.4.2). To evaluate the diversity of PTTSs in filamentous fungi, a multiple sequence alignment was performed using CLUSTAL W 2.0.12. Evolutionary analyses were subsequently conducted using MEGA7. Finally, a phylogenetic tree was generated using the Jones–Taylor–Thornton matrix-based model.

#### Construction of Plasmids and Mutants

All strains and plasmids are summarized in SI Appendix, Tables S26 and S27. To generate a plasmid to express sufficient levels of the precursors IPP and DMAPP in S. cerevisiae, pY2L141 was constructed as previously described (29). To confirm the functions of the PTTSs in S. cerevisiae, pRC301- pRC383 were constructed. With the help of an automated high-throughput platform (Biomek FXP Laboratory Automation Workstation), the plasmids were linearized and inserted into the HIS3 site of S. cerevisiae YL141 to generate S. cerevisiae mRC310-mRC383. For high-throughput plasmid assembly, the LiAc/SS carrier DNA/polyethylene glycol (PEG) yeast transformation protocol was modified from a previously reported protocol (50).

For plasmid assembly, 300 ng each coding sequence fragment was combined with 300 ng linearized expression vector. Using a Biomek FX® Laboratory Automation Workstation equipped with a MP200 96-Tip Tool for liquid handling operations, the DNA mix was transferred into the CEN.PK2-1D yeast strain using the modified LiAc/PEG protocol. Detailed procedures are presented in SI Appendix, Fig. S3.

#### Fermentation of PTTSs in S. cerevisiae

For product detection, RC310-RC383 mutants were cultivated in 24 deep-well plates containing YPD medium. Cultures were incubated at 30 °C for 3 d, and then compounds were extracted using hexane/ethyl acetate (4:1) and concentrated for GC-MS detection. For structural characterization, candidate mutants were activated on YPD agar

Chen et al. Systematic mining of fungal chimeric terpene synthases using an efficient precursor-providing yeast chassis

PNAS | 7 of 9

https://doi.org/10.1073/pnas.2023247118
plates and inculcated into 5 mL YPD (2% glucose) medium at 30 °C overnight. Next, 1% of the culture was used to inoculate a 250-mL shaker flask containing 50 mL (2% glucose), which was incubated at 30 °C overnight. Finally, 1% of the culture was transferred to a 2-L shaker flask containing 1 L of YPD (2% glucose and 1% D-(+)-galactose) at 30 °C for 3 d of fermentation. Strains were incubated at 30 °C overnight. Finally, 1% glucose and 1% D-(+)-galactose) at 30 °C for 3 d of fermentation. Strains were incubated

100% acetonitrile from 0 to 50 min and then 100% acetonitrile from 60 to 70 min. Fragments were detected using UV light at 210 nm. The product structure was identified by GC-MS, HR-El-MS, and NMR. Sesterorbiculene (2). Yellowish oil; C25H40; m/z 340.3126 (calculated C25H40 for 340.3130). Infrared radiation (IR) (KBr) νmax 2958, 2925, 2854, 1666, 1463, 1363, 1098, 954, and 802 cm⁻¹. Sesterorbiculene (2).

GC-MS Analysis of Terpenoids. Terpenoids were detected by GC-MS on a Thermo TRACE GC Ultra combined with a TSO Quantum XLS MS. The samples were injected into a TRACE TR-SMS GC column (30 m × 0.25 mm × 0.25 μm). The oven temperature was initially set at 80 °C for 1 min, increased to 220 °C at 10 °C/min, and then held at 220 °C for 15 min. The injector and transfer lines were maintained at 230 and 240 °C, respectively. The compounds were all analyzed in an m/z range of 50 to 500. El mass spectra were compared with the National Institute of Standards and Technology (NIST) mass spectral library.

Isolation and Structural Identification of Compounds. The redissolved extracts were preseparated by silica gel (80 to 100 mesh) column chromatography with petroleum ether/ethyl acetate (100:1 to 1:1). Fractions were collected and analyzed by GC-MS for identification. The final purification was performed by preparative HPLC using an Ultimate 3000 HPLC and a SEP LC-52 with a MWD UV detector (Separation Technology Co Ltd.). The elution gradient was 90 to 100% acetonitrile from 0 to 50 min and then 100% acetonitrile from 60 to 70 min. Fragments were detected using UV light at 210 nm. The product structure was identified by GC-MS, HR-El-MS, and NMR. Sesterorbiculene (2). Yellowish oil; C25H40; m/z 340.3126 (calculated C25H40 for 340.3130). Infrared radiation (IR) (KBr) νmax 2958, 2925, 2854, 1666, 1463, 1363, 1098, 954, and 802 cm⁻¹. Sesterorbiculene (2).

Data Availability. Data supporting the findings of this work are included in the article and/or SI Appendix. The sequences of 34 biochemically characterized PTTs reported in this paper have been deposited in the GenBank database (accession nos. MW796200 to MW798233).

ACKNOWLEDGMENTS. This work was financially supported by the National Key Research and Development Program of China (2018YFA0904000, 2018YFC02120600), the National Natural Science Foundation of China (31971341, 31800032, and 32070603), the Medical Science Advancement Program (Clinical Medicine) of Wuhan University (YCT2018002), and the Fundamental Research Funds for the Central Universities (2042020F0504). We thank Dr. Jianwei Tang (The Hong Kong University of Science and Technology) for his help with chemical calculations.

2. L. Lauterbach, J. Rinkel, J. S. Dickschat, Two bacterial diterpene synthases from Allokerzneria albata produce bonndiene, phomopene, and allokerzneriune. Angew. Chem. Int. Ed. Engl. 57, 8280–8283 (2018).

4. M. Chen, W. K. Chou, T. Toyomasu, D. E. Can, D. W. Christianson, Structure and function of fuscoadiene synthase, a hexameric bifunctional diterpene synthase. ACS Chem. Biol. 11, 889–899 (2016).

11. G. Wei.

26. G. Bian, Z. Deng, T. Liu, Strategies for terpenoid overproduction and new terpenoid synthase in fungi. Biotechnol. Bioeng. 115, 1356–1365 (2018).

27. M. Okada et al., Astellialidine: Structure determination by NMR spectroscopy and crystalline sponge method, and elucidation of its biosynthesis. Angew. Chem. Int. Ed. Engl. 55, 5785–5788 (2016).

28. G. Bian et al., Unearthing a sesterterpene biosynthetic repertoire in the unearthing a sesterterpene biosynthetic repertoire in the conidial wall of Sclerotinia sclerotiorum through genome mining reveals convergent evolution. Proc. Natl. Acad. Sci. U.S.A. 114, E6005–E6014 (2017).

29. Y. Matsuda et al., Astellialidine: Structure determination by NMR spectroscopy and crystalline sponge method, and elucidation of its biosynthesis. Angew. Chem. Int. Ed. Engl. 55, 5785–5788 (2016).

30. T. Mitsuhashi, J. Rinkel, M. Okada, I. Abe, Chimeric terpene synthases possessing both terpene cyclization and prenyltransfer activities. ChemBioChem 19, 1106–1114 (2018).

31. T. Mitsuhashi, I. Abe, Chimeric terpene synthases possessing both terpene cyclization and prenyltransfer activities. ChemBioChem 19, 1106–1114 (2018).

32. P. K. Aikjum et al., Isooenzyme pathway optimization for Taxol precursor overproduction in Escherichia coli. Science 330, 70–74 (2010).

33. A. T. Sato et al., Identification of novel sesterterpene/terpene synthase from Bacillus clausii. ChemBioChem 14, 822–825 (2013).

34. A. C. Huang et al., Unearthing a sesterterpene biosynthetic repertoire in the Brassicaeae through genome mining reveals convergent evolution. Proc. Natl. Acad. Sci. U.S.A. 114, E6005–E6014 (2017).

35. A. C. Huang et al., Solid-phase organic synthesis of polysaccharides with cellulose matrix. J. Org. Chem. 73, 7197–7203 (2008).

36. T. Q. Song et al., Engineering Saccharomyces cerevisiae for geranylgeranoid overproduction by combinatorial design. Sci. Rep. 7, 14991 (2017).

37. L. F. You, L. Q. Guo, J. F. Lin, T. Ren, J. R. Wang, Overproduction of geranylgeraniol in Coprinopsis cinerea by the expression of geranylgeranyl diphosphate synthase gene. J. Biotechnol. 182, 241–247 (2014).

38. V. Falara, J. M. Alba, M. R. Kant, R. C. Schuurink, E. Pichersky, Geranylgeraniol synthases in solanaceae and other angiosperms constitute an ancient branch of diterpene synthases involved in the synthesis of defensive compounds. Plant Physiol. 166, 428–441 (2014).

39. T. A. Pemberton et al., Exploring the influence of domain architecture on the catalytic function of diterpene synthases. Biochemistry 56, 2010–2023 (2017).

40. U. W. Hawas, A. A. El-Beih, A. M. El-Halawany, Bioactive anthraquinones from endophytic fungus Aspergillus versicolor isolated from red sea algae. Arch. Pharm. Res. 35, 1479–1486 (2012).

41. A. C. Huang, Y. H. Hong, A. D. Bond, D. J. Tantillo, A. Osbourn, Diverged plant terpene synthases involved in the synthesis of defensive compounds. Angew. Chem. Int. Ed. Engl. 57, 1291–1295 (2018).

42. L. F. You et al., Mechanistic investigations of two bacterial diterpene cyclases: sapphriene synthase and tricosadiene synthase. Angew. Chem. Int. Ed. Engl. 56, 2776–2779 (2017).

8 of 9

PNAS

https://doi.org/10.1073/pnas.2023247118

Chen et al.

Systematic mining of fungal chimeric terpene synthases using an efficient precursor-providing yeast chassis
42. B. Qin et al., An unusual chimeric diterpene synthase from Emericella variecolor and its functional conversion into a sesterterpene synthase by domain swapping. *Angew. Chem. Int. Ed. Engl.* 55, 1658–1661 (2016).

43. J. Guo et al., Genome mining reveals a multiproduct sesterterpenoid biosynthetic gene cluster in *Aspergillus ustus*. *Org. Lett.* 23, 1525–1529 (2021).

44. F. Chen, D. Tholl, J. Bohlmann, E. Pichersky, The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* 66, 212–229 (2011).

45. J. M. Murphy, P. D. Mace, P. A. Eyers, Live and let die: Insights into pseudoenzyme mechanisms from structure. *Curr. Opin. Struct. Biol.* 47, 95–104 (2017).

46. A. Villaverde, M. M. Carrió, Protein aggregation in recombinant bacteria: Biological role of inclusion bodies. *Biotechnol. Lett.* 25, 1385–1395 (2003).

47. F. Rueda et al., Functional inclusion bodies produced in the yeast *Pichia pastoris*. *Microb. Cell Fact.* 15, 166 (2016).

48. Q. Chen et al., Molecular basis for sesterterpene diversity produced by plant Terpene synthases. *Plant Commun.* 1, 100051 (2020).

49. E. Pichersky, E. Lewinsohn, Convergent evolution in plant specialized metabolism. *Annu. Rev. Plant Biol.* 62, 549–566 (2011).

50. R. D. Gietz, R. H. Schiestl, Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2, 38–41 (2007).