1   **Releasing potential power of terpene synthases by a robust precursor supply platform**

2   **Authors:** Guangkai Bian[1,4], Yichao Han[2,4], Anwei Hou[1], Yujie Yuan[1], Xinhua Liu[1], Zixin

3   Deng[1,3], Tiangang Liu[1,3]*

4   **Affiliations:**

5   [1]Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education and

6   Wuhan University School of Pharmaceutical Sciences, Wuhan 430071, P. R. China.

7   [2]J1 Biotech Co., Ltd., Wuhan 430075, P. R. China.

8   [3]Hubei Engineering Laboratory for Synthetic Microbiology, Wuhan Institute of Biotechnology,

9   Wuhan 430075, P. R. China.

10   [4]These authors contributed equally to this work.

11   *e-mail: liutg@whu.edu.cn

12   **Abstract**: Approximately 76,000 discovered makes terpenoids the largest family of natural

13   products in nature with widespread applications. The wide-spectrum of structural diversity of the

14   terpenoids were largely due to the variable skeletons generated by terpene synthases. The

15   number of terpene skeletons found in nature, however, were so much more than those

16   conceivably generated from known terpene synthases and the limited characterized terpene

17   synthases also make no chance for some useful terpenoids overproduction in microbe. Here, we

18   first demonstrated that the promiscuous synthases *in vivo* can produce more variable terpenoid

19   products by converting precursors of different lengths (C_{10}, C_{15}, C_{20}, C_{25}). This discovery was

20   prompted by the development of an efficient *in vivo* platform by combining the two promiscuous

21   terpene synthases and three prenyltransferases to generate 50 terpenoids, at least 3 ring systems

22   of which were completely new. Furthermore, protein engineering was further integrated to

23   enhance product diversity. Clearly, the work is expected to dramatically reshape the terpenoid

24   research by widening the flexibility of the terpene synthases for the fresh discovery or creation of

25   the new terpenoid compounds by skeleton reframing.

26   **Keywords:** terpenoids; terpene synthases; promiscuous; protein engineering; precursor supply;

27   Combinatorial biosynthesis; skeleton reframing

28

29 **1. Introduction**

30   Nature is an excellent synthetic chemist, utilising two simple processes, i.e., mutations and

31 combinations of biosynthetic genes, to create diverse natural products (Cane et al., 1998). By

32 combining sets of genes, domains, and modules identified from different biosynthetic pathways

33 or created by scientists in various ways, the concept of combinatorial biosynthesis has emerged

34 to accelerate the discovery and expand the diversity of products (Kim et al., 2015; Tsoi and

35 Khosla, 1995). This concept has been applied extensively to two large families of complex

36 natural products, namely, polyketides and nonribosomal peptides (Khosla et al., 2014; Sieber and

37 Marahiel, 2005; Weissman and Leadlay, 2005). Compared with these two families, very few

38 reports have described the use of this strategy for terpenoids, the largest and most structurally

39 diverse family of natural products. Recently, combinatorial biosynthesis has been proposed to

40 generate diverse labdane-based diterpenoids (Andersen-Ranberg et al., 2016; Jia et al., 2016;

41 Mafu et al., 2016). However, many other skeletons in the big terpenoid family remain to be

42 explored.

43   The diversity of terpenoids arises from three steps (Fig. S1): (i) prenyltransferases (PTs)

44 assemble the C_5 precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate

45 (DMAPP) into polyisoprenoid diphosphates with various lengths, such as C_{10} geranyl

46 diphosphate (GPP), C_{15} farnesyl diphosphate (FPP), C_{20} geranylgeranyl diphosphate (GGPP), and

47 C_{25} geranylfarnesyl diphosphate (GFPP) (Kellogg and Poulter, 1997); (ii) terpene synthases

48 cyclise these polyisoprenoid diphosphates to generate terpene products with a single ring or

49 intricate multiple rings (Christianson, 2006); (iii) tailoring enzymes, including oxygenases,

50 methyltransferases, acetyltransferases, and glycosyltransferases, add functional groups at

51 different positions, further enhancing structural diversity. The possibility of obtaining linear

52   products from the first step is very limited for the fixed length of C_5 precursors, and the

53   flexibility of the final products from the third step strongly depends on the skeletons from the

54   second step. Thus, variability in the skeleton is mainly introduced through the cyclisation step.

55   Biosynthesis by terpene synthases in this step has inherent advantages over chemical synthesis

56   (Maimone and Baran, 2007); however, the biosynthetic repertoire identified to date only covers a

57   small proportion of the terpene skeletons discovered, which only covers a small proportion of

58   potential terpene skeletons through theoretical calculations (Tian et al., 2016). Additionally, most

59   traditional methods to investigate terpene synthases only focus on their major functions under

60   native conditions, ignoring their potential for different applications.

61   Combinatorial biosynthesis of terpenoid skeletons will be an effective solution when two

62   prerequisites are satisfied: identification of terpene synthases that tolerate different lengths of

63   isoprene-diphosphate as substrates, and generation of suitable polyisoprenoid diphosphate

64   substrates for these enzymes. In this study, we successfully identified promiscuous terpene

65   synthases in nature and developed a terpene-overproducing chassis to satisfy these two criteria.

66   Six *Escherichia coli* variants, obtained by combining two terpene synthases and three PTs,

67   generated 50 terpenoids. We also demonstrated that terpene synthases with substrate promiscuity

68   are easily available in nature, and protein engineering of theses enzymes can be further

69   incorporated in our strategy.

70   **2. Materials and Methods**

71   **2.1 Materials**

72   Linalool (51782, > 99%; Sigma-Aldrich, St. Louis, MO, USA), α-terpineol (30627, ≥ 90%;

73   Sigma-Aldrich), pyrophosphate reagent (P7275; Sigma-Aldrich), and α-farnesene (F102425;

74   J&K Scientific, Beijing, China) were used in this study. Kinetic assays were recorded on a

75   Multimode Plate Reader Enspire (PerkinElmer, MA, USA). Proteins were purified using a

76   Biologic DuoFlow Chromatography System (Bio-Rad, Hercules, CA, USA). Terpenoids were

77   separated using a Dionex Ultimate 3000 UHPLC system (Thermo Scientific, Waltham, MA,

78   USA). Optical rotations were measured on a PerkinElmer Model 341 polarimeter (PerkinElmer,

79   USA). Single-crystal data were measured on an Bruker Kappa APEX-Duo CCD diffractometer

80   (Bruker, MA, USA). ^{1}H NMR and ^{13}C NMR were performed on an Agilent 400 MHz or 600

81   MHz instrument (DirectDrive2; CA, USA).

2.2 Identification of class I terpene synthases

83   Class I terpene synthases were identified by screening all predicted proteins against a set of

84   hidden Markov models constructed for each terpene synthase family. The model trained from the

85   non-plant terpene cyclase alignment profile (cd00687 from the Conserved Domain Database)

86   showed the best prediction coverage in our preliminary test data. The final identified proteins

87   were manually curated on the basis of protein alignment and transcript assembly.

2.3 Phylogenetic analysis

89   Multiple sequence alignment was performed with MAFFT software (Katoh et al., 2002)

90   using the local pair iterative refinement method. The following evolutionary analyses were

91   conducted in MEGA7 (Kumar et al., 2016). The phylogenetic tree was inferred using the

92   maximum likelihood method based on the Jones-Taylor-Thornton model. All positions

93   containing gaps or missing data were eliminated. The final dataset included 211 positions.

2.4 Motif analysis

95   The active-site cavity was mapped according to existing crystal structures of fungal terpene

96   synthases (PaFS, AtARS, and FsTDS) and refined alignment of all characterised fungal terpene

97   synthases. The binding pocket of terpene synthases consisted of six α-helixes, i.e., C, D, F, G, H,

98 and J. Strictly conserved and catalysis-related sites were excluded. Sequence logos of

99 nonconserved residues in the active-site cavity were generated using the WebLogo server

100 (http://weblogo.berkeley.edu/).

2.5 Molecular modeling

102 The three-dimensional model of the TC domain of FgMS was built using the SWISS-

103 MODEL web server (https://www.swissmodel.expasy.org/). The crystal structure of the TC

104 domain of PaFS from *Phomopsis amygdali* was selected as the template. The resulting homology

105 structures were visualised using UCSF Chimera.

2.6 *In vitro* assays and kinetic measurements

107 To test the *in vitro* activities of FgMS, its mutants, and FgGS, reactions were conducted

108 using 10 μM purified proteins, 100 μM substrates (GPP, FPP, GGPP, or GFPP), and 2 mM Mg^{2+}

109 in 200 μL of 50 mM PB buffer (pH 7.6) with 10% glycerol at 30°C overnight. The sample was

110 extracted with hexane (200 μL) and then analysed by GC-MS. For steady-state kinetics, 100-μL

111 scale reactions were carried out in 50 mM Tris-HCl buffer (pH 7.6) with 10% glycerol and 50 μL

112 pyrophosphate reagent. The concentration of the enzyme (FgMS or FgGS) were kept constant at

113 1 mg/mL with 2 mM Mg^{2+} as the only metal ion, while substrates (GPP, FPP, GGPP, and GFPP)

114 were added at difference concentration ranging from 1 to 200 mM. Product assays were carried

115 out by measuring the release of pyrophosphate (PPi), which recorded on an Enspire Multimode

116 Plate Reader, using a previously described method (Agger et al., 2008).

2.7 Fermentation and purification of terpenoids

118 *E. coli* T7 (harbouring pMH1, pFZ81, and pGB310), T8 (harbouring pMH1, pFZ81, and

119 pGB311), T9 (harbouring pMH1, pFZ81, and pGB312), T10 (harbouring pMH1, pFZ81, and

120 pGB313), and T11 (harbouring pMH1, pFZ81, and pGB314) were cultivated in 2-L flasks

121   containing 1 L LB medium at 37°C with 100 mg/L ampicillin (AMP), 50 mg/L KAN, and 34

122   mg/L chloramphenicol (CM). When the OD_{600} reached 0.6–0.8, 0.1 mM IPTG was added to the

123   cultures, protein expression was induced at 16°C for 18 h, and the strains were then cultivated at

124   28°C for an additional 3 days. The cell cultures were extracted twice with an equal volume of

125   hexane. The organic layer was concentrated using a rotary evaporator, and terpenoids were

126   purified using a Dionex Ultimate 3000 UHPLC system.

127   **2.8 Detection of terpenoids**

128   Terpenoids were extracted by hexane and detected by GC-MS (Thermo TRACE GC ULTRA

129   combined with a TSQ QUANTUM XLS MS). The samples were injected into a TRACE TR-

130   5MS (30 m × 0.25 mm × 0.25 um). The oven temperature was set at 80°C for 1 min, increased to

131   220°C at a rate of 10°C/min, and held at 220°C for 15 min. The injector and transfer lines were

132   maintained at 230°C and 240°C, respectively. Product structures were determined by comparison

133   with authentic standards, comparison with mass spectra data from the NIST library, and NMR

134   analysis.

135   **3. Results**

136   **3.1 Terpene synthases with substrate promiscuity**

137   To predict promiscuous terpene synthases from sequences, we collected sequences of terpene

138   synthases in fungi, a major source of terpenoids that has not yet been fully explored, and

139   constructed a phylogenic tree. Fifty-one genetically or biochemically characterised terpene

140   synthases (Additional Data Table S1), including mono-, sesqui-, di-, and sesterterpene synthases,

141   were mainly divided into five clades (Fig. 1a). One of them, clade III, was the most interesting.

142   Notably, as of the writing of this manuscript (August 2016), all characterised di- and

143   sesterterpene synthases in fungi were enriched in clade III rather than widely distributed, while

144 diterpene synthases were dispersed in the phylogenic trees of bacterial terpene synthases

145 (Dickschat, 2016; Yamada et al., 2015). Thus, enzymes in this clade are more likely to generate

146 large complex terpene skeletons (Fig. 1b). More importantly, in this clade, three enzymes are

147 able to catalyse the cyclisation of multiple polyisoprenoid diphosphate substrates *in vitro* (Chen

148 et al., 2016; Chiba et al., 2013; Qin et al., 2016). Although the substrate specificities of other

149 enzymes have not been investigated to date, we speculated that the broad substrate range could

150 be a common property of this type of terpene synthase. For these two reasons, we selected

151 enzymes in clade III as candidates for combinatorial biosynthesis.

152 From the collection of genomes we sequenced in our lab, we found two clade III terpene

153 synthases in an endophytic fungus *Fusarium graminearum* J1-012, a well-known pathogen

154 causing Fusarium head blight. These two enzymes (i.e., FgJ07578 and FgJ07623 in our

155 annotated genome) were selected as candidates because no similar clade III enzymes of them

156 have been characterised before. FgJ07578 is a chimeric protein consisting of an N-terminal

157 terpene cyclase (TC) domain and a C-terminal PT domain; in contrast, FgJ07623 only contains a

158 TC domain of 316 amino acids. Notably, no PT exists near FgJ07623, and the biosynthetic genes

159 in its cluster are not typically related to terpenoid synthesis (Fig. S2).

160 **3.2 Validating broad substrate specificities**

161 These two enzymes (FgJ07578 and FgJ07623) were overexpressed in *E. coli*, purified, and

162 characterised biochemically *in vitro*. Surprisingly, in the absence of IPP, both FgJ07578 and

163 FgJ07623 could directly catalyse the conversion GPP, FPP, GGPP, and GFPP to generate

164 multiple products *in vitro*, as detected by gas chromatography mass spectrometry (GC-MS),

165 exhibiting unprecedentedly broad substrate specificities for polyisoprenoid diphosphates (Fig. 2).

166   Because isoprenoid diphosphate substrates could be elongated by the PT domain of FgJ07578

167   itself, we also conducted *in vitro* experiments with the addition of IPP for FgJ07578. When

168   DMAPP and IPP were used as substrates, mono-, sesqui-, di-, and sesterterpenes were all

169   detected through *in vitro* reactions, indicating that the PT domain of FgJ07578 could generate

170   multiple polyisoprenoid diphosphates from C_{10} to C_{25} *in vitro* (Fig. S3). In addition, when GPP,

171   FPP, or GGPP was incubated with the addition of IPP as substrates (Fig. S3), both terpenes with

172   corresponding lengths and terpenes with longer lengths were generated, suggesting that

173   polyisoprenoid diphosphate substrates could be not only directly cyclised by the TC domain but

174   also first elongated by the PT domain to form longer substrates.

175   We then performed kinetic analyses to understand the turnover rates and specificities

176   quantitatively. When substrates reaching their saturating concentrations, FgJ07578 exhibited

177   higher turnover rates for GGPP and GFPP (almost equally), and FgJ07623 exhibited the highest

178   turnover rate for GGPP. However, both FgJ07578 and FgJ07623 will prefer GPP as the better

179   substrate if different substrates were given, because of their highest specificity constants (k_{cat}/K_m)

180   for GPP. Notably, the specificity constants towards FPP, GGPP, and GFPP were similar for these

181   two enzymes, with their differences being less than 1.5-fold change (Table 2). This similarity

182   provided a basis for engineering to exploit substrate promiscuity by redistributing substrates *in*

183   *vivo*.

184   **3.3 Combinatorial biosynthesis of terpenoid skeletons**

185   To realise the full potential of these promiscuous terpene synthases, we developed a platform

186   consisting of three parts: a terpene-overproducing chassis, a PT library, and the promiscuous

187   terpene synthases of interest. First, we built an *E. coli* strain as a terpene-overproducing chassis

188   based on previous design principles (Zhu et al., 2014). Specifically, on the basis of data from *in*

189   *vitro* reconstitution, proteomics, and metabolomics, the entire mevalonate pathway (AtoB,

190   ERG13, tHMG1, ERG13, ERG8, MVD1, and Idi) was appropriately constructed to provide

191   sufficient C5 precursors (Fig. 3a). This strategy has successfully been applied to overproduction

192   of farnesene (Zhu et al., 2014), lycopene (Zhu et al., 2015), astaxanthin (Ma et al., 2016), and

193   taxadiene (unpublished data). Second, efficient PTs, including FPPS, GGPPS, and GFPPS from

194   previous reports (Ajikumar et al., 2010; Furubayashi et al., 2015; Zhu et al., 2014), were selected

195   and codon-optimised to construct a library. GPPS was not included because of the relatively

196   limited possible conformations of monoterpenes. Third, we combined the two terpene synthases

197   with these three PTs to form six variants from T7 to T12 (Fig. 3b). In *E. coli* T8 and T9, the PT

198   domain of FgJ07578 was inactivated by replacing a key catalytic residue, aspartic acid, with

199   alanine (D510A) to block substrate flux towards GFPP.

200   GC-MS analysis revealed that 50 different terpenoids were produced by these six strains,

201   including nine sesterterpenoids, 24 diterpenoids, and 17 sesquiterpenoids (Fig. 3c, Figs. S4a and

202   S5). Twelve products (**1** to **12**) were separated followed by nuclear magnetic resonance (NMR)

203   analyses to determine their chemical structures.

204   Three sesterterpenoids were subjected to NMR analyses, two of which were new precursors

205   of natural products with pharmaceutical activities. Specifically, two sesterterpenoid products (**1**

206   and **2**) in *E. coli* T7 were identified as a 5-5-6-5 tetracyclic compound and an 11-6 bicyclic

207   compound, respectively. Compound **1** is the sesterterpene precursor of mangicols, of which

208   mangicol A and mangicol B exhibit anti-inflammatory activity (Renner et al., 2000). Compound

209   **2** is the sesterterpene precursor of variecolin, which has been shown to inhibit angiotensin II

210   receptor (Hensens et al., 1991) and act as an immunosuppressant (Fujimoto et al., 2000). The

211   identification of these skeletons could fill gaps between biosynthetic repertoire and known

212 natural products. The cyclisation mechanisms of these two compounds were summarised in Fig.

213 S6a according to previous classification (Ye et al., 2015). Besides these two compounds, the GC-

214 MS data (Fig. 3c and Fig. S5) of another six sesterterpenoids (**37** to **42**) detected from *E. coli* T7

215 cannot be matched with existing sesterterpenoids data in the literatures, suggesting that new

216 structures could be discovered. The only sesterterpenoid product (**3**) in *E. coli* T10 was

217 determined to be 2*E*-alpha-cericerene; this compound is also a sesterterpene product of the EvVS

218 variant (Qin et al., 2016), which shares only 40% sequence identity with FgJ07623.

219   From structural elucidation of diterpene products, three entirely new ring systems were

220 identified. The diterpenoid main product (**4**) in *E. coli* T8 was found to be cembrene A, a simple

221 skeleton with a large 14-membered ring that has been reported to be a byproduct of the bacterial

222 diterpene synthase DtcycB (Meguro et al., 2013). In addition, some of its side diterpene products,

223 such as **18** and **19**, were predicted to have the same skeleton as cembrene A on the basis of mass

224 spectra data from the NIST library (Figs. S4a and S5). To the best of our knowledge, this was the

225 first fungal gene found to generate this type of C_{20} single-large-ring skeleton. The skeletons of

226 diterpenoids **5** to **10** in *E. coli* T11 all possessed two five-membered rings, which were formed in

227 the first two ring-closure steps (Fig. S6b). Compound **5** was identified as a 5-5-5-5 membered

228 tetracyclic diterpene. Compounds **6** and **8** were identified as 5-5-9 membered tricyclic diterpenes,

229 which differed in the positions of double bonds. The relative configuration of compound **6** was

230 matched with variediene (Qin et al., 2016). Compound **7** was identified as a 5-5-7-4 tetracyclic

231 diterpene. Compounds **9** and **10** were identified as 5-5-7-4 and 5-5-6-5 tetracyclic diterpene

232 alcohols without any double bonds, respectively. To the best of our knowledge, the 5-5-5-5, 5-5-

233 6-5, and 5-5-7-4 tetracyclic skeletons are unprecedented. The absolute configuration of **10** was

234 then determined by single-crystal X-ray diffraction analysis (Additional File).

235   Sesquiterpenoids identified were often linear products or alcohols. Two sesquiterpenoid main

236   products (**11** and **12**) in *E. coli* T9 were validated as *trans*-nerolidol and 2*E*, 6*E*-farnesol, two

237   linear alcohols. *E. coli* T12 also produced *trans*-nerolidol as the main product, which is one of

238   the few overlapping products of these two terpene synthases (Fig. S7).

239   According to existing natural products mangicol and variecolin, the new products **1** and **2**

240   were designated mangicdiene and variecoltetraene respectively. Because no reference could be

241   given to new products **5** and **7** to **10**, they were designated GJ1012 A to E, respectively; thus,

242   these two terpene synthases (FgJ07578 and FgJ07623) were designated *F. graminearum*

243   mangicdiene synthase (FgMS) and *F. graminearum* GJ1012 synthase (FgGS), respectively.

3.4 Other promiscuous terpene synthases in this clade

245   To test whether the substrate promiscuity is a widespread feature of this kind of terpene

246   synthases, we performed large scale genome mining. In UniProt database, 2088 proteins from

247   208 fungal genomes were predicted to be class I terpene synthases. According to our phylogenic

248   analysis, 321 (15.4%) of them could be further classified as clade III terpene synthases in 109

249   different species (Additional Data Table S2). The result indicates clade III terpene synthases are

250   easily available for application of our strategy.

3.5 Incorporating protein mutagenesis

252   To further enhance product diversity, we performed site-direct mutations of the promiscuous

253   FgMS as an example to integrate protein engineering into our framework. We first investigated

254   how the diverse sequence features of natural terpene synthases led to functional differentiation in

255   evolution. At least 13 noncatalytic sites within the pocket were diverse in fungi (Table S1). For

256   convenience, we named these specificity-determining sites (SDSs) 1–13, corresponding to

257   residue numbers 65, 69, 85, 88, 89, 159, 168, 191, 219, 222, 307, 311, and 314 in FgMS. In the

258   terpene cyclisation reaction, the aromatic residues phenylalanine (F), tryptophan (W), and

259   tyrosine (Y) not only stabilise carbocation centres through cation-π interactions but also provide

260   steric hindrance, restricting the active-site contour (Li et al., 2014); therefore, we highlighted the

261   SDSs where aromatic residues occurred in the sequence logo for each clade of the phylogenetic

262   tree (Fig. 4a). Aromatic residues in SDSs 5, 6, and 13 were relatively conserved in all fungal

263   terpene synthases; however, other SDSs with abundant aromatic residues seemed specific to

264   distinct clades, e.g., they were enriched in SDSs 1, 8, and 9 in clade III and in SDSs 2, 11, or 12

265   in other clades. Thus, based on our observations of natural evolution, our protein engineering

266   strategy focused on the interchange between aromatic and nonaromatic residues for each SDS in

267   FgMS (Fig. 4b).

268   Our preliminary cell-free incubation experiments showed that when nonaromatic residues of

269   SDSs in FgMS were replaced by F, W, or Y, the mutants generated the same products, but with

270   different product ratios, i.e., substitutions of different aromatic residues did not alter the product

271   compositions of FgMS. Therefore, we only used the mutants replaced by F to represent the

272   interchange from a nonaromatic residue to an aromatic residue. All mutants of interchange

273   between aromatic and nonaromatic residues for SDSs maintained their catalytic activity, at least

274   for FPP (Additional Data Table S3 and Fig. 4c). Nine of the mutants exhibited narrower

275   substrate ranges, and two of them even lost catalytic activity for GGPP (Fig. 4c). These

276   phenomena also implied that interchange between aromatic and nonaromatic residues could be

277   utilised to broaden the substrate ranges of other terpene synthases with narrow ones. From a

278   homolog model view, when the nonaromatic hydrophobic residues were replaced by aromatic

279   hydrophobic residues in the middle or bottom of the binding pocket, the overall pocket volume

280   will be greatly reduced disabling the catalysis of longer substrates. Position A88 was an

281   exception because it did not point to the center of the binding pocket (Fig. 4b). Comparing the

282   sequences of characterised diterpene synthases and sesterterpene synthases could also offer

283   similar explanations. PaFS cannot accept GFPP and it possesses more aromatic residues in their

284   binding pockets (at SDS 1, 5, 6, 8, 9, and 13) (Table S1).

285   For the majority, the product profiles were redistributed, thus we could exploit different

286   mutants to enhance production of certain products (Additional Data Table S3). More importantly,

287   new products could be generated in this way to further enhance the terpene skeleton diversity.

288   For examples, six new sesterterpenoids (**53** to **58**) emerged from mutants F65L and F159G (Fig.

289   4d and Fig. S5). Compound **56** was isolated and identified as a new 5-8-6-6 tetracyclic

290   sesterterpene (Fig. 4e), which has the same skeleton as asperterpenol(Xiao et al., 2013), a fungal

291   metabolite showing inhibition on acetylcholinesterase (acetylcholinesterase inhibition is one of

292   current strategies to Alzheimer's disease). In summary, incorporating protein engineering

293   methodology boosted the performance of the combinatorial biosynthesis strategy.

294

295 **4. Discussion**

296   Our findings suggested that a relative small set of building blocks was sufficient to generate

297 skeletally diverse terpenoids, demonstrating the elegance of terpene biosynthesis in nature.

298 Although we did not test the physiological roles of the promiscuous terpene synthases (because

299 they were silenced in the native fungi under our laboratory conditions), the implement of their

300 multiple biosynthetic functions in living cells suggests the promiscuity of these enzymes could

301 serve as regulation of cell metabolism. Similar examples exist in nature. Plants have utilized

302 promiscuous terpene synthases to regulate emission of different flavors: controlled by targeting

303 signals, nearly identical terpene synthases that can convert both GPP and FPP are delivered to

304 different subcellular locations where different substrates are available, to produce C_{10} linalool

305 and C_{15} nerolidol (Aharoni et al., 2004; Nagegowda et al., 2008), respectively. It is plausible that

306 organisms preserve their promiscuous activity through evolution to increase fitness for survival

307 because the more products that are generated, the more likely the organism is to produce

308 chemicals with useful biological activity (Fischbach and Clardy, 2007).

309   Our discoveries and combinatorial biosynthesis strategy accelerated the entire process for

310 discovery and creation of terpenoids. We demonstrated that terpene synthases with substrate

311 promiscuity were widespread. Hundreds of protein candidates with similar property existed in

312 current database. In addition, engineering the residues in the binding pocket can be further

313 incorporated to access more scaffolds. Notably, the substrate promiscuity was not unique in

314 di/sesterterpene synthases, which were concentrated in clade III fungal terpene synthases. We

315 focused on this kind of enzymes because they could be utilized to generate relatively complex

316 compounds with one or more rings. The polycyclic skeletons extensive in di- and

317 sesterterpenoids offer a variety of diverse structures, promoting numerous pharmaceutical

318 applications (Stockdale and Williams, 2015), e.g. diterpenoids Taxol (paclitaxel) used in cancer

319 therapy. Many skeletons, e.g. compound **1**, **2**, and **56**, we produced in this study are precursors of

320 natural products with important pharmaceutical activities (Fig. S4b). The metabolic engineering

321 platform helps us obtain a sufficient amount of these scaffolds, not only enabling the elucidation

322 of their chemical structures, but also being a start for industrial application.

323 Our work greatly expanded the biosynthetic repertoire of terpene skeletons, i.e., the

324 possibilities of upstream in the entire biosynthetic pathway. Combinatorial biosynthesis can also

325 be applied to downstream modifications via promiscuous tailoring enzymes (Mafu et al., 2016).

326 To further incorporating downstream tailoring enzymes, we recommended to transfer our

327 strategy to engineered terpene-producing *Saccharomyces cerevisiae*. This is because

328 heterologous expression of downstream decorating pathway in *E. coli* is still challenging,

329 although some problems are progressively understood and solved (Biggs et al., 2016). Given that

330 very limited terpenoids have been commercially generated through biosynthesis (Leavell et al.,

331 2016), once multistep combinations are achieved, the fermentative production will cover much

332 more desired terpenoids.

333 **GenBank accession number**

334 The DNA and protein sequences of FgJ07578, FgJ07623 and GFPPS described here are

335 deposited in GenBank under accession numbers KY462789, KY462790 and KY462792.

336

341   (2013ACA011), the 973 Project of National Program on Key Basic Research (2012CB721000),

342   and the Young Talents Program of National High-level Personnel of Special Support Program

343   (The "Ten Thousand Talent Program").

344   **Author contributions:** G. B., Y. H., Z. D., and T. L. conceived of the project, designed the

345   experiments, and wrote the manuscript; G. B., Y. Y., and X. L. performed *in vivo* and *in vitro*

346   experiments; Y. H. performed bioinformatics analyses; A. H. performed NMR analyses; X. L.

347   constructed and characterised the mutant library. Z. D. and T. L. managed the project.

348   **Competing financial interests:** Two patent applications based on the results in this study have

349   been filed by J1 Biotech Co., Ltd.

350 **References**

351 Agger, S. A., Lopez-Gallego, F., Hoye, T. R., Schmidt-Dannert, C., 2008. Identification of

352   sesquiterpene synthases from *Nostoc punctiforme* PCC 73102 and *Nostoc* sp. strain PCC

353   7120. J. Bacteriol. . 190**,** 6084-6096.

354 Aharoni, A., Giri, A. P., Verstappen, F. W., Bertea, C. M., Sevenier, R., Sun, Z., Jongsma, M. A.,

355   Schwab, W., Bouwmeester, H. J., 2004. Gain and loss of fruit flavor compounds

356   produced by wild and cultivated strawberry species. Plant Cell. 16**,** 3110-31.

357 Ajikumar, P. K., Xiao, W.-H., Tyo, K. E. J., Wang, Y., Simeon, F., Leonard, E., Mucha, O.,

358   Phon, T. H., Pfeifer, B., Stephanopoulos, G., 2010. Isoprenoid pathway optimization for

359   Taxol precursor overproduction in *Escherichia coli*. Science. 330**,** 70-74.

360 Andersen-Ranberg, J., Kongstad, K. T., Nielsen, M. T., Jensen, N. B., Pateraki, I., Bach, S. S.,

361   Hamberger, B., Zerbe, P., Staerk, D., Bohlmann, J., Moller, B. L., Hamberger, B., 2016.

362   Expanding the landscape of diterpene structural diversity through stereochemically

363   controlled combinatorial biosynthesis. Angewandte Chemie. 55**,** 2142-6.

364 Biggs, B. W., Lim, C. G., Sagliani, K., Shankar, S., Stephanopoulos, G., De Mey, M., Ajikumar,

365   P. K., 2016. Overcoming heterologous protein interdependency to optimize P450-

366   mediated Taxol precursor synthesis in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A.

367   113**,** 3209-14.

368 Cane, D. E., Walsh, C. T., Khosla, C., 1998. Harnessing the biosynthetic code: combinations,

369   permutations, and mutations. Science. 282**,** 63-8.

370 Chen, M., Chou, W. K. W., Toyomasu, T., Cane, D. E., Christianson, D. W., 2016. Structure and

371   function of fusicoccadiene synthase, a hexameric bifunctional diterpene synthase. ACS

372   Chem. Biol. . 11**,** 889-899.

373   Chiba, R., Minami, A., Gomi, K., Oikawa, H., 2013. Identification of ophiobolin F synthase by a

374   genome mining approach: a sesterterpene synthase from *Aspergillus clavatus*. Org. Lett.

375   15**,** 594-597.

376   Christianson, D. W., 2006. Structural biology and chemistry of the terpenoid cyclases. Chem.

377   Rev. . 106**,** 3412-3442.

378   Dickschat, J. S., 2016. Bacterial terpene cyclases. Nat. Prod. Rep. . 33**,** 87-110.

379   Fischbach, M. A., Clardy, J., 2007. One pathway, many products. Nat. Chem. Biol. . 3**,** 353-355.

380   Fujimoto, H., Nakamura, E., Okuyama, E., Ishibashi, M., 2000. Immunomodulatory constituents

381   from an ascomycete, Emericella aurantio-brunnea. Chem. Pharm. Bull. 48**,** 1436-1441.

382   Furubayashi, M., Ikezumi, M., Takaichi, S., Maoka, T., Hemmi, H., Ogawa, T., Saito, K., Tobias,

383   A. V., Umeno, D., 2015. A highly selective biosynthetic pathway to non-natural C50

384   carotenoids assembled from moderately selective enzymes. Nat. Commun. 6.

385   Hensens, O. D., Zink, D., Williamson, J. M., Lotti, V. J., Chang, R. S. L., Goetz, M. A., 1991.

386   Variecolin, a sesterterpenoid of novel skeleton from *Aspergillus variecolor* MF138. J.

387   Org. Chem. . 56**,** 3399-3403.

388   Jia, M., Potter, K. C., Peters, R. J., 2016. Extreme promiscuity of a bacterial and a plant

389   diterpene synthase enables combinatorial biosynthesis. Metab. Eng. . 37**,** 24-34.

390   Katoh, K., Misawa, K., Kuma, K. i., Miyata, T., 2002. MAFFT: a novel method for rapid

391   multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. . 30**,**

392   3059-3066.

393   Kellogg, B. A., Poulter, C. D., 1997. Chain elongation in the isoprenoid biosynthetic pathway.

394   Curr. Opin. Chem. Biol. . 1**,** 570-578.

395   Khosla, C., Herschlag, D., Cane, D. E., Walsh, C. T., 2014. Assembly line polyketide synthases:

396   mechanistic insights and unsolved problems. Biochemistry. 53**,** 2875-2883.

397   Kim, E., Moore, B. S., Yoon, Y. J., 2015. Reinvigorating natural product combinatorial

398   biosynthesis with synthetic biology. Nat. Chem. Biol. . 11**,** 649-659.

399   Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular evolutionary genetics analysis

400   version 7.0 for bigger datasets. Mol. Biol. Evol.**,** msw054.

401   Leavell, M. D., McPhee, D. J., Paddon, C. J., 2016. Developing fermentative terpenoid

402   production for commercial usage. Curr. Opin. Biotechnol. 37**,** 114-9.

403   Li, R., Chou, W. K. W., Himmelberger, J. A., Litwin, K. M., Harris, G. G., Cane, D. E.,

404   Christianson, D. W., 2014. Reprogramming the chemodiversity of terpenoid cyclization

405   by remolding the active site contour of epi-isozizaene synthase. Biochemistry. 53**,** 1155-

406   1168.

407   Ma, T., Zhou, Y., Li, X., Zhu, F., Cheng, Y., Liu, Y., Deng, Z., Liu, T., 2016. Genome mining of

408   astaxanthin biosynthetic genes from *Sphingomonas* sp. ATCC 55669 for heterologous

409   overproduction in *Escherichia coli*. Biotechnol. J. . 11**,** 228-237.

410   Mafu, S., Jia, M., Zi, J., Morrone, D., Wu, Y., Xu, M., Hillwig, M. L., Peters, R. J., 2016.

411   Probing the promiscuity of ent-kaurene oxidases via combinatorial biosynthesis. Proc.

412   Natl. Acad. Sci. U.S.A. 113**,** 2526-2531.

413   Maimone, T. J., Baran, P. S., 2007. Modern synthetic efforts toward biologically active terpenes.

414   Nat. Chem. Biol. . 3**,** 396-407.

415   Meguro, A., Tomita, T., Nishiyama, M., Kuzuyama, T., 2013. Identification and characterization

416   of bacterial diterpene cyclases that synthesize the cembrane skeleton. ChemBioChem. 14**,**

417   316-321.

418 Nagegowda, D. A., Gutensohn, M., Wilkerson, C. G., Dudareva, N., 2008. Two nearly identical

419   terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers.

420   Plant J. 55, 224-39.

421 Qin, B., Matsuda, Y., Mori, T., Okada, M., Quan, Z., Mitsuhashi, T., Wakimoto, T., Abe, I.,

422   2016. An unusual chimeric diterpene synthase from *Emericella variecolor* and its

423   functional conversion into a sesterterpene synthase by domain swapping. Angew. Chem.

424   Int. Ed. . 55, 1658-1661.

425 Renner, M. K., Jensen, P. R., Fenical, W., 2000. Mangicols: structures and biosynthesis of a new

426   class of sesterterpene polyols from a marine fungus of the genus *Fusarium*. J. Org.

427   Chem. . 65, 4843-4852.

428 Sieber, S. A., Marahiel, M. A., 2005. Molecular mechanisms underlying nonribosomal peptide

429   synthesis: approaches to new antibiotics. Chem. Rev. . 105, 715-38.

430 Stockdale, T. P., Williams, C. M., 2015. Pharmaceuticals that contain polycyclic hydrocarbon

431   scaffolds. Chem. Soc. Rev. . 44, 7737-63.

432 Tian, B., Poulter, C. D., Jacobson, M. P., 2016. Defining the product chemical space of

433   monoterpenoid synthases. PLoS Comput. Biol. 12, e1005053.

434 Tsoi, C. J., Khosla, C., 1995. Combinatorial biosynthesis of 'unnatural'natural products: the

435   polyketide example. Chem. Biol. 2, 355-362.

436 Weissman, K. J., Leadlay, P. F., 2005. Combinatorial biosynthesis of reduced polyketides. Nat.

437   Rev. Microbiol. . 3, 925-936.

438 Xiao, Z. e., Huang, H., Shao, C., Xia, X., Ma, L., Huang, X., Lu, Y., Lin, Y., Long, Y., She, Z.,

439   2013. Asperterpenols A and B, new sesterterpenoids isolated from a mangrove

440   endophytic fungus *Aspergillus* sp. 085242. Org. Lett. 15, 2522-2525.

441   Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-ya, K., Omura, S., Cane, D. E., Ikeda, H., 2015.

442   Terpene synthases are widely distributed in bacteria. Proc. Natl. Acad. Sci. U.S.A. 112,

443   857-862.

444   Ye, Y., Minami, A., Mandi, A., Liu, C., Taniguchi, T., Kuzuyama, T., Monde, K., Gomi, K.,

445   Oikawa, H., 2015. Genome mining for sesterterpenes using bifunctional terpene

446   synthases reveals a unified intermediate of di/sesterterpenes. J. Am. Chem. Soc. . 137,

447   11846-53.

448   Zhu, F., Lu, L., Fu, S., Zhong, X., Hu, M., Deng, Z., Liu, T., 2015. Targeted engineering and

449   scale up of lycopene overproduction in *Escherichia coli*. Process Biochem. 50, 341-346.

450   Zhu, F., Zhong, X., Hu, M., Lu, L., Deng, Z., Liu, T., 2014. In vitro reconstitution of mevalonate

451   pathway and targeted engineering of farnesene overproduction in *Escherichia coli*.

452   Biotechnol. Bioeng. . 111, 1396-1405.

453

454   Table 1. Strains and plasmids used in this research.

Strains	Relevant genotype			Reference
BL21 (DE3)	*E. coli B F⁻ dcm ompT hsdSB(rB⁻mB⁻)gal*			Invitrogen
C1	*E. coli* BL21:: pMH1, pFZ81			This work
T7	*E. coli* BL21:: pMH1, pFZ81, pGB309			This work
T8	*E. coli* BL21:: pMH1, pFZ81, pGB310			This work
T9	*E. coli* BL21:: pMH1, pFZ81, pGB311			This work
T10	*E. coli* BL21:: pMH1, pFZ81, pGB312			This work
T11	*E. coli* BL21:: pMH1, pFZ81, pGB313			This work
T12	*E. coli* BL21:: pMH1, pFZ81, pGB314			This work
Plasmids	**Origin of replication**	**Description**	**Resistance**	**Reference**
pMH1	p15A	plac: AtoB, ERG13, tHMG1	CM	(Zhu et al., 2014)
pFZ81	pBBR1MCS	plac: ERG12, ERG8, MVD1, Idi	KAN	(Zhu et al., 2014)
pGB301	pBR322	pT7: N-terminal his$_6$-tag FgMS	KAN	This work
pGB302	pBR322	pT7: FgMS-D510A	AMP	This work
pGB303	pBR322	pT7: N-terminal his$_6$-tag FgGS	KAN	This work
pGB304	pBR322	pT7: FgGS	AMP	This work
pGB305	pBR322	pT7: FPPS	AMP	This work
pGB306	pBR322	pT7: Idi	AMP	This work
pGB307	pBR322	pT7: Idi	AMP	This work
pGB308	pBR322	pT7: FPPS, Idi	AMP	This work
pGB309	pBR322	pT7: FgMS, GFPPS; pT7: Idi	AMP	This work
pGB310	pBR322	pT7: FgMS-D510A, GGPPS; pT7: Idi	AMP	This work
pGB311	pBR322	pT7: FgMS-D510A, FPPS, Idi	AMP	This work
pGB312	pBR322	pT7: FgGS, GFPPS; pT7: Idi	AMP	This work
pGB313	pBR322	pT7: FgGS, GGPPS; pT7: Idi	AMP	This work
pGB314	pBR322	pT7: FgGS, FPPS, Idi	AMP	This work

455   Note: The *fpps* and *Idi* genes were of *E. coli* origin (Zhu et al., 2014); the *ggpps* was of *Taxus canadensis*

456   origin and codon optimized (Ajikumar et al., 2010); the *gfpps* was of *Geobacillus stearothermophilus*

457   origin and codon optimized (Furubayashi et al., 2015).

458   Table 2. Catalytic activity of FgMS and FgGS.

Enzyme	Substrate	K_m(μM)	k_{cat}(s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})
FgJ07578 (FgMS)	GPP	0.24±0.04	0.0071±0.0002	3×10^4
	FPP	3.12±0.98	0.033±0.003	1.06×10^4
	GGPP	6.43±0.87	0.068±0.002	1.05×10^4
	GFPP	8.88±1.91	0.07±0.007	7.8×10^3
FgJ07623 (FgGS)	GPP	0.061±0.026	0.008±0.0005	1.3×10^5
	FPP	2.58±0.38	0.019±0.0007	7.4×10^3
	GGPP	7.26±1.9	0.047±0.0031	6.5×10^3
	GFPP	3.9±0.79	0.023±0.001	5.9×10^3

459

460

461
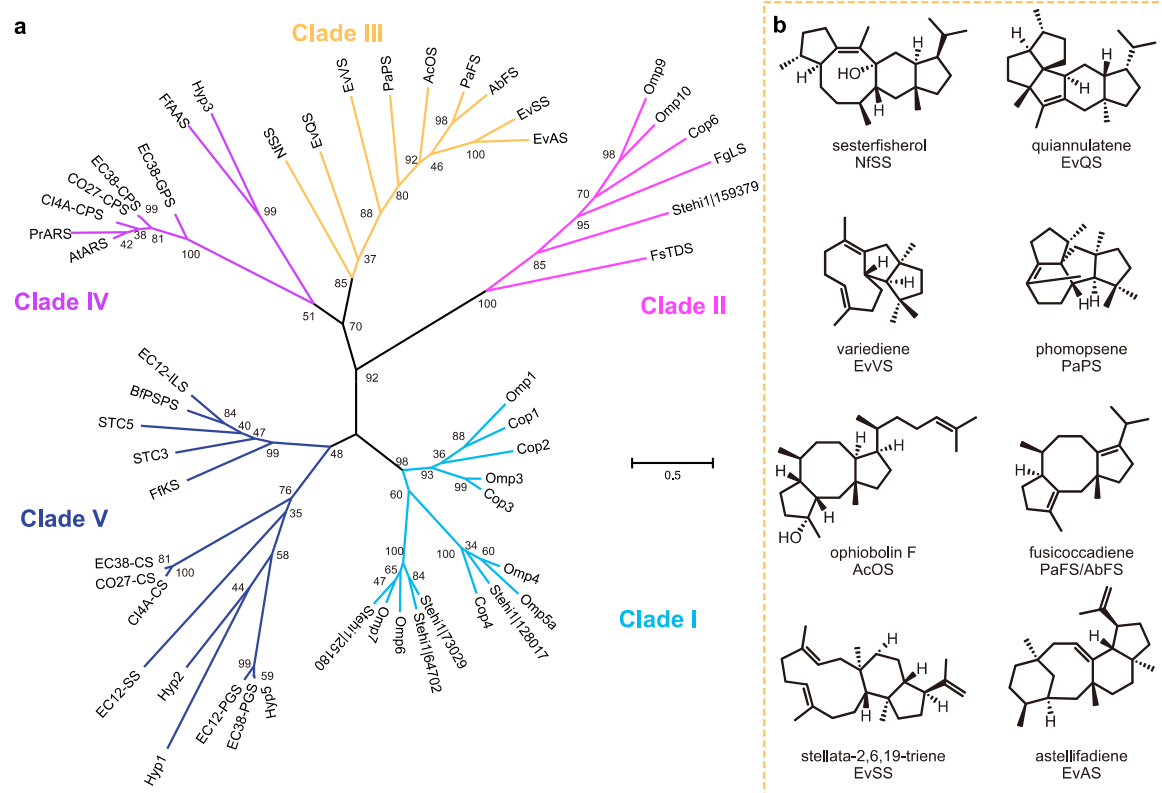
462 **Fig. 1. Clades of characterised fungal class I terpene synthases.** (**a**) Phylogenetic tree of 51

463 experimentally characterised fungal terpene synthases. The enzyme abbreviations are described

464 in the Additional Data Table S1. Branch length indicates the substitutions per site. Branches are

465 labeled with the percentage of 1,000 bootstrap replicates. (**b**) Products of terpene synthases in
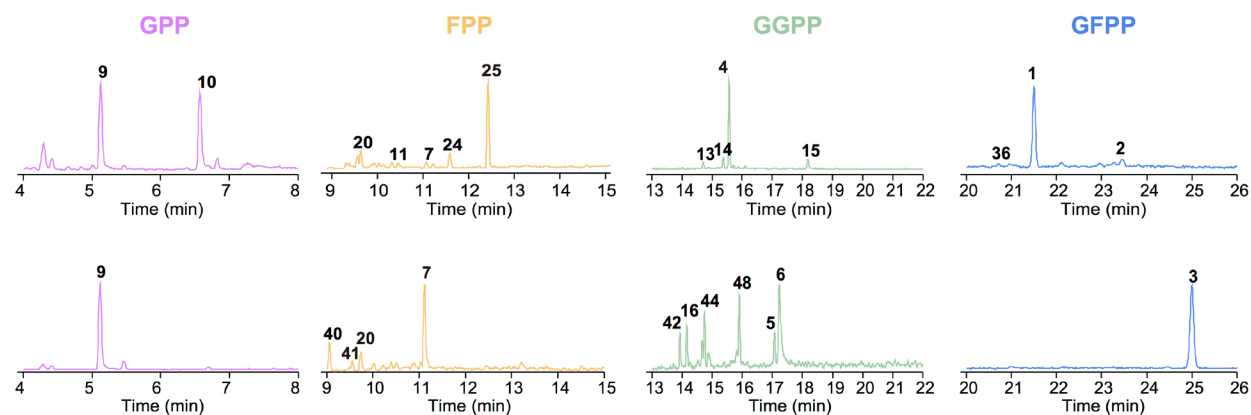
466 clade III in previous studies.

Figure 2. *In vitro* characterisation of FgJ07578 and FgJ07623. Chromatograms showing GC-MS analysis of terpenoid products from *in vitro* assays of FgJ07578 (top) and FgJ07623 (bottom) with GPP, FPP, GGPP, and GFPP, respectively, in the absence of IPP.
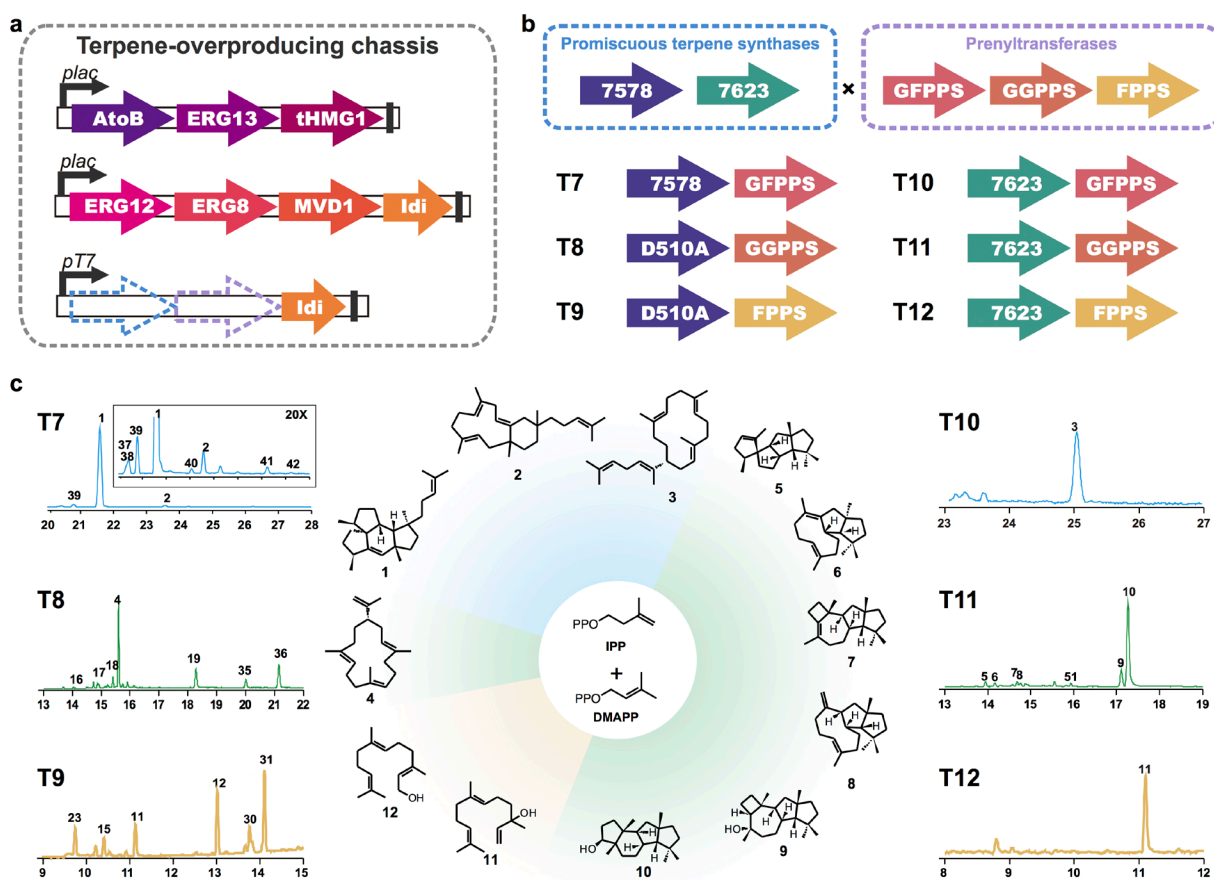
Fig. 3. Accessing diverse complex terpenoid skeletons using combinatorial biosynthesis. (a) A terpene-overproducing chassis in which the entire mevalonate pathway was introduced. **(b)** Combinations of two terpene synthases and three prenyltransferases. 7578: FgJ07578; 7623: FgJ07623; D510A: the D510A mutant of FgJ07578. **(c)** Fermentation results of six *E. coli* strains. All the structures shown were identified by NMR analysis.
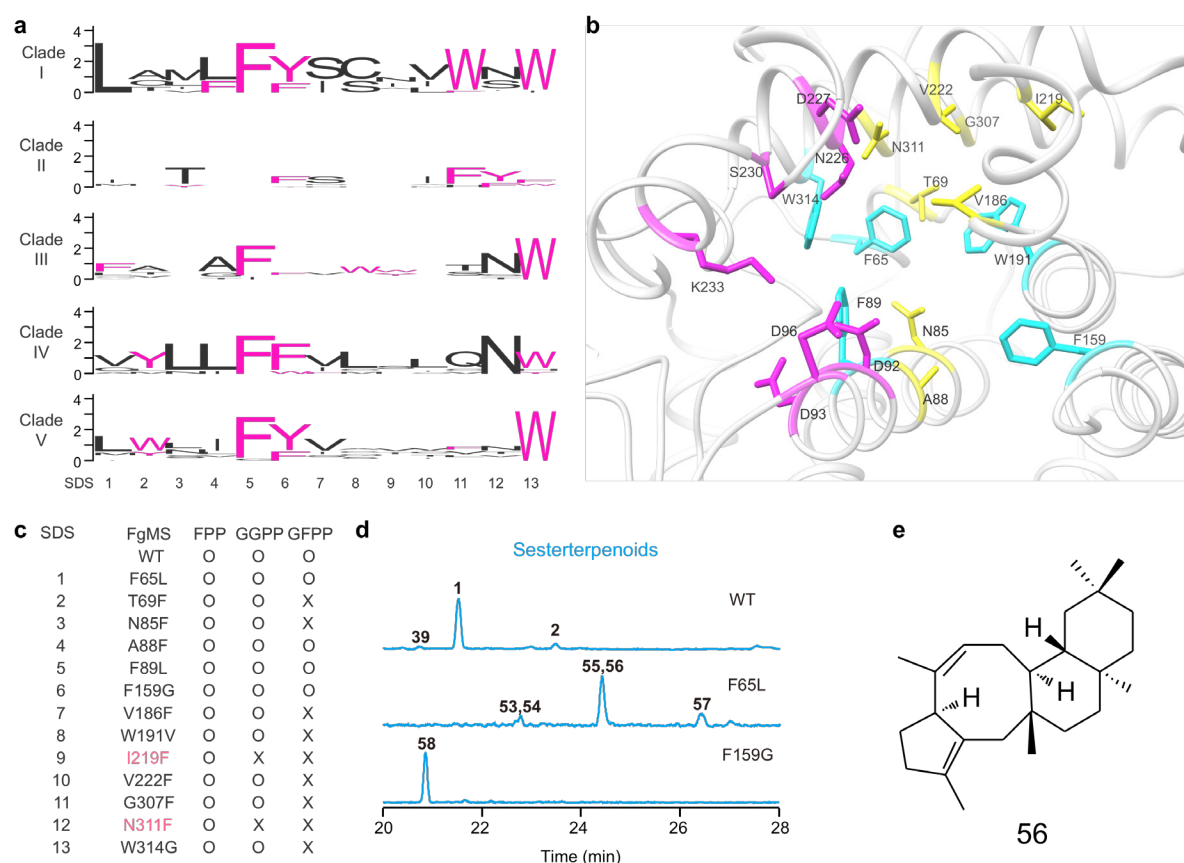
477

Figure 4. Protein engineering by interchange between aromatic and nonaromatic residues.

(**a**) Sequence logos of specificity-determining sites (SDSs) for each clade in the phylogenetic tree. SDSs 1–13 in these logos correspond to residue numbers 65, 69, 85, 88, 89, 159, 168, 191, 219, 222, 307, 311, and 314 in FgMS. Aromatic residues are highlighted. (**b**) Three-dimensional model of the active-site cavity of FgMS. Metal-binding catalytic motifs are indicated in magenta, aromatic SDSs are indicated in cyan, and nonaromatic SDSs are indicated in yellow. (**c**) Substrate ranges of FgMS mutants. O, maintain catalytic activity; X, loss catalytic activity. (**d**) New products can be generated by FgMS mutants. WT, wild-type FgMS. (**e**) The chemical structure of compound **56**.