Synthesis of 11-carbon terpenoids in yeast using protein and metabolic engineering

Synthesis of 11-carbon terpenoids in yeast

Engineering the biosynthesis of a non-canonical building block expands the chemical space accessed by terpenoids

Codruta Ignea[1,2,§], Marianna Pontini[2,3,§], Mohammed S. Motawia[1], Massimo E. Maffei[3], Antonios M. Makris[4] and Sotirios C. Kampranis[1,2,*]

[1]Department of Plant and Environmental Sciences, University of Copenhagen, Denmark

[2]Mediterranean Agronomic Institute of Chania, Chania, Greece

[3]Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy

[4]Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece

[§]Equal author contribution

[*]Corresponding author. Email: soka@plen.ku.dk

Abstract:

One application of synthetic biology is the re-design of existing biological systems to acquire new functions. In this context, expanding the chemical code underlying key biosynthetic pathways will lead to the synthesis of compounds with new structures and potentially new biological activities. Terpenoids are the largest group of specialized metabolites with numerous applications. Yet, being synthesized from 5-carbon units, they are restricted to distinct classes that differ by 5 carbon atoms (C_{10}, C_{15}, C_{20}, etc.). To expand the diversity of terpenoid structures, we engineered yeast cells to synthesize a non-canonical building block with 11 carbons, and produced 40 C_{11} terpene scaffolds that can form the basis for an entire terpenoid class. By identifying a single-residue switch that converts C_{10} plant monoterpene synthases to C_{11}-specific enzymes, we engineered dedicated synthases for C_{11} terpene production. This approach will enable the systematic expansion of the chemical space accessed by terpenoids.

For centuries, specialized metabolites, the small molecules made by plants and microorganisms to protect themselves and to communicate with their environment, provide humankind with pharmaceuticals, flavors, fragrances and colorants. Despite the immense diversity of specialized metabolites found in nature, the pathways responsible for their biosynthesis are based on a handful of basic building blocks derived from primary metabolism [1]. As a result, the chemical space accessed by specialized metabolites is restricted to the structural complexity that can be obtained using the specific building blocks. This is particularly evident in the case of terpenoids, which are synthesized from prenyl diphosphate precursors involved in the synthesis of sterols and other membrane lipids. The core terpenoid biosynthetic pathway is built up by the successive addition of 5-carbon (C_5) units, consequently giving rise to a small set of building blocks that differ by 5 carbon atoms. In turn, each of these building blocks generates a distinct class of terpenoid compounds, i.e. monoterpenoids (C_{10}), sesquiterpenoids (C_{15}), diterpenoids (C_{20}), etc [2,3]. The use of a small number of building blocks, in combination with the modular structure of the downstream pathways [4], establishes a chemical code for terpene biosynthesis (Fig. 1). According to this code, biosynthetic activities involving initially terpene synthases and subsequently decorating enzymes (cytochrome P450s, dehydrogenases, O-methyltransferases, O-acetyltransferases, etc.) act in a sequential manner on prenyl diphosphate building blocks of defined size to yield a plethora of structurally complex molecules (Fig. 1). This general scheme is very strict and only few exceptions have been reported [5-9]. Expanding the chemical code of terpenoid biosynthesis will provide numerous structures with unexplored properties and potentially new pharmacological or industrial applications.

To introduce additional layers in the stratified terpenoid chemical space, we aimed to bypass the strict dependence of terpenoid biosynthesis on building blocks of specific size. Previous studies using recombinant enzymes suggested that it may be possible for a terpene synthase to convert non-standard substrates [10,11]. To enable a systematic approach, we set out to develop a complete biosynthetic system for novel terpenoids by engineering the yeast *Saccharomyces cerevisiae* to produce alternative terpene scaffolds. To this end, we searched for enzymatic activities that can be used to synthesize non-canonical prenyl diphosphates. Certain bacteria produce a unique volatile, 2-methylisoborneol (2MIB, **1**), which contributes to the musty-earthy fragrant notes of Brie and Camembert cheeses [12]. 2MIB is an 11-carbon molecule synthesized by a unique mechanism in which the universal monoterpene precursor geranyl diphosphate (GPP) becomes methylated at C-2 by a dedicated methyltransferase and subsequently the 2-methyl-GPP (2meGPP) substrate is cyclized by a specific synthase to yield 2MIB (inset in right panel of Fig. 1 and references [13,14]). This mechanism appears to be specific for the biosynthesis of 2MIB and the related compound 2-methylenebornane (2MB, **2**), since no other C_{11}

compound has been reported as a major metabolite in bacteria [15-17]. In addition, all gene clusters containing the combination of a methyltransferase with a terpene synthase characterized to date have been found to produce either 2MIB or 2MB as the main product [13,14,18-20]. Thus, we explored the possibility to use 2meGPP as an alternative substrate for the systematic synthesis of non-canonical terpenes in yeast.

Results

Establishing efficient 2meGPP synthesis. To engineer yeast cells to produce 2meGPP, we introduced the GPP methyltransferase from the cyanobacterium *Pseudanabaena limnetica* (*Pl*GPPMT) [19] into a strain engineered for terpenoid production (strain AM94; Table S1). To monitor synthesis of 2meGPP, we also introduced the second enzyme in the pathway, *P. limnetica* 2MIB synthase (*Pl*MIBS) [19], and applied solid-phase micro extraction (SPME) of the head-space above yeast liquid cultures to analyze the volatile compounds produced by the engineered cells. The 2MIB degradation products 2MB, 1-methylcamphene (1MC, **3**) and 2-methyl-2-bornene (2M2B, **4**) were clearly detected in the head-space above cells co-expressing *Pl*GPPMT and *Pl*MIBS, confirming synthesis of 2meGPP and production of 2MIB (Fig. 1, 2a, and Table S2). 2MIB is unstable and readily converts to an approximately 1:1 ratio of 2MB:1MC and traces of 2M2B [21]. The identity of these compounds was confirmed by comparison with authentic 2MIB and its degradation products, while control experiments confirmed that these compounds were not produced by a yeast enzyme or any of the two *P. limnetica* enzymes alone (Fig. 2a and Fig. S1).

Establishing 2meGPP synthesis in yeast opened the possibility to use this alternative building block to produce non-canonical terpenoids. To facilitate this effort, we first aimed to improve 2meGPP production. In yeast, GPP is synthesized by Erg20p, which is a bi-functional enzyme that subsequently converts it to farnesyl diphosphate (FPP). The mechanism of Erg20p is distributive, allowing GPP to be released and taken up by *Pl*GPPMT to synthesize 2meGPP. Thus, as shown in the scheme in Fig. 2b, Erg20p competes with *Pl*GPPMT for GPP. To facilitate *Pl*GPPMT in obtaining access to GPP, a series of interventions were implemented and the *Pl*MIBS products were used as readout to evaluate system performance in each step. Protein fusions between enzymes catalyzing successive steps have previously been shown to improve heterologous terpenoid production [22]. Thus, N-terminal and C-terminal fusions of *Pl*GPPMT with Erg20p were constructed and tested. Expression of the Erg20p-*Pl*GPPMT fusion resulted in a 15-fold increase in *Pl*MIBS-produced C_{11} compounds compared to expression of the non-fused Erg20p and *Pl*GPPMT proteins (Fig. 2c). A lower 6.1-fold increase was observed when the reverse fusion was used (Fig. 2c). Evaluation of the *Pl*GPPMT protein levels by

western blotting revealed that the observed improvements in C_{11} production could be attributed to the higher levels of the *Pl*GPPMT protein when fused to Erg20p (Fig. S2). Subsequently, to further decrease competition for GPP by Erg20p, we introduced a mutant form of Erg20p that functions as a GPP synthase. Erg20p(N127W) had previously been developed in our lab as a variant that specifically synthesizes GPP and at the same time inhibits FPP synthesis by wild-type Erg20p [23-25]. By incorporating the N127W mutation into the Erg20p-*Pl*GPPMT fusion an additional 4.5-fold improvement in C_{11} production was obtained (Fig. 2d). To further improve C_{11} synthesis, the level of wild-type Erg20p was reduced by heterozygous deletion of *ERG20* [26]. Strain MIC1 (Table S1), generated by replacing one of the two *ERG20* alleles with the gene encoding *Pl*GPPMT, was 5.3-fold more efficient than the parent strain, AM94, when *Pl*MIBS and the Erg20p(N127W)-*Pl*GPPMT fusion were overexpressed (Fig. 2e; Table S3).

We subsequently explored whether the performance of *Pl*GPPMT could be improved. We constructed a model of *Pl*GPPMT based on the structure of the closely related GPPMT from *Streptomyces coelicolor* (PDB id: 3VC2; reference [27]) (Fig. S3) and searched for residue substitutions that could facilitate the association of the GPP substrate. A total of 16 residues were selected (Y36, M160, Y161, I202, Y206, C208, I210, Y240, W241, R244, V250, V253, F257, Y261, F266, and Y268; Fig. S3) and a library of 44 different mutants of the non-fused form of the *Pl*GPPMT enzyme was constructed (Table S5). Screening this library in yeast using the *Pl*MIBS products as readout revealed two substitutions, V250A and F266H, which improved performance by approximately 2- and 3-fold, respectively (Fig. 2f and Table S4). The most efficient substitution, F266H, was subsequently introduced into the Erg20p(N127W)-*Pl*GPPMT fusion and a 2-fold improvement in production was obtained (Table S3). The combination of all the above interventions amounted in an overall 760-fold improvement in C_{11} yield from the starting strain (Table S3 and Fig. 2g), reaching 23.5 mg/L. Consistent with these findings, analysis of the intracellular prenyl diphosphates revealed a 1.94±0.29:1 ratio of 2meGPP:GPP in the most efficient strain.

The new building block does not interfere with yeast metabolism. First, we confirmed that 2meGPP was not taken up by Erg20p to produce larger prenyl diphosphates. We set up *in vitro* reactions using bacterially produced Erg20p (Fig. S4) and found that although the recombinant enzyme could readily produce FPP from GPP or DMAPP (and IPP), it was not active with 2meGPP (online methods). We also analyzed the extracts of Erg20p(N127W)-*Pl*GPPMT-expressing yeast cells and found no additional Erg20p-derived product beyond the regular sterol, squalene or sesquiterpenol compounds (Fig. S5). We were also unable to detect any 2meGPP-derived products of Erg20p in yeast cell extracts supplemented with 2meGPP and IPP (online methods). We also confirmed that *Pl*GPPMT

could not methylate other diphosphate substrates than GPP. For this, we tested recombinant *Pl*GPPMT (Fig. S4) in *in vitro* reactions with other prenyl diphosphate substrates (DMAPP, IPP, FPP and GGPP) and were not able to identify any new products (online methods). In combination with the overall observation that yeast growth properties were unaffected by *Pl*GPPMT expression (Fig. S6), we concluded that the system developed here is orthogonal to the yeast metabolism.

The 2meGPP building block can provide several C_{11} terpenes beyond 2MIB and 2MB. Having established efficient synthesis of 2meGPP, we questioned whether this building block can yield other C_{11} terpene scaffolds. Initially, we evaluated the range of products obtained by other microbial enzymes by expressing the 2MB synthases from *Micromonospora olivasterospora* (*Mo*MBS) [13] and *Pseudomonas fluorescens* (*Pf*MBS) [18]. The volatile profiles produced by *Mo*MBS or *Pf*MBS were clearly dominated by 2MB, 1MC and 2M2B which in combination reached 93.8% and 67.1% of total C_{11} terpene products, respectively (Table S2). However, several other C_{11} terpenes could also be detected in lower amounts (Fig. S7). These included 2-methylmyrcene (**5**), 2-methyllimonene (**6**), 2-methyllinalool (**7**), 2-methyl-α-terpineol (**8**), 2-methylgeraniol (**9**) (Fig. 1 and Fig. S8), and 17 additional C_{11} products (Fig. S7, Tables S2 and S6) identified by characteristic molecular fragment ions in mass-spectrometry (m/z 150 for C_{11} terpene hydrocarbons or 168 for C_{11} terpene alcohols). The structure of the 17 minor products could not be characterized at this stage because of their low quantities and the complexity of the mixture. Re-examination of the *Pl*MIBS product profile using the improved yeast platform revealed that some of these molecules could also be detected as minor products of *Pl*MIBS (Tables S2 and S6). Several of the minor C_{11} compounds (2-methylmyrcene, 2-methyllimonene, 2-methyllinalool, 2-methyl-α-terpineol) had previously been identified in *in vitro* assays using the studied enzymes [18,28] or in minute amounts in the headspace of bacterial cultures [16,29], where, in agreement with present findings, they were considered to be side-products of a promiscuous terpene synthase rather than the main product of a specific enzyme. Our engineered yeast platform also produces GPP, and we observed that all three 2MIB/2MB synthases were also able to accept GPP as a substrate and produce C_{10} terpenes (Fig. 3 and S7). *Mo*MBS was the least efficient with GPP, producing only 8.9% of C_{10} terpenes, while *Pl*MIBS produced 10.9% C_{10} compounds (Fig. 3). *Pf*MBS was the most promiscuous enzyme of the three, producing an almost equal amount of the two product groups (56.2% C_{11} and 43.8% C_{10} terpenes; Fig. 3). These results strongly suggested that 2meGPP can give several other products beyond 2MIB/2MB and that there is likely a broad chemical diversity of C_{11} structures that can be obtained from this building block, which could be similar in complexity to that obtained by the canonical monoterpene substrate GPP (Fig. 1). However, since the bacterial synthases identified so far are restricted to the 2MIB/2MB bouquet (2MIB, 2MB, 2M2B, 1MC) [13],

tapping into the unexplored C_{11} diversity would require mining or engineering suitable terpene synthases.

Expanding the diversity of C_{11} terpenes. The ability of the 2MIB and 2MB synthases to accept both GPP and 2meGPP, albeit with varying efficiency, suggested that these two building blocks could serve as alternative substrates for the same enzyme. Thus, to access the C_{11} diversity, we investigated whether enzymes characterized as canonical terpene synthases could accept 2meGPP to yield C_{11} products. To this end, we turned to the rich resource of plant monoterpene synthases and selected 7 different enzymes with varied product specificity. Expression of the 1,8-cineole synthase from *Salvia fruticosa* (*Sf*CinS1) [30] together with the Erg20p(N127W)-*Pl*GPPMT(F266H) fusion resulted in the production of 19 C_{11} compounds, alongside the characteristic C_{10} monoterpenes produced by this enzyme from GPP. The main C_{11} products were 2MB (58.5% of C_{11}), 2-methylmyrcene (19.4% of C_{11}), a yet unknown C_{11} alcohol (11.2% of C_{11}), and 2-methyllimone (2.1% of C_{11}) (Fig. 4a, Tables S2 and S6). Still, wild-type *Sf*CinS1 produced predominantly C_{10} terpenes (73.9% of total C_{10} and C_{11} production) in the yeast system (Fig. 3). The second plant monoterpene synthase tested, pinene synthase from *Pinus taeda* (*Pt*PinS) [31], converted 2meGPP to 29 different C_{11} compounds accounting for almost 26% of the product blend from both substrates (Fig. 3 and 4b, Tables S2 and S5). 2MB, 2-methyllinalool, 1MC and 2-methylmyrcene were among the most abundant C_{11} products of *Pt*PinS (Table S2). Sabinene synthase from *S. pomifera* (*Sp*SabS) [30] produced a complex blend of 17 C_{11} volatiles with 2-methylmyrcene as the main C_{11} product (45% of C_{11}) (Fig. 3 and 4c, Tables S2 and S6). Myrcene synthase from *Ocimum basilicum* (*Ob*MyrS) [32] converted 2meGPP mainly to 2-methylmyrcene (59.7% of C_{11}) and 2-methyllinalool (24.1% of C_{11}) (Fig. 3 and Tables S2 and S6). Geraniol synthase from *O. basilicum* (*Ob*GerS) [33] produced 19 different C_{11} terpenes (Fig. 3 and Tables S2, S6), three of which were made only by this enzyme. *Ob*GerS produced 2% C_{11} compounds in the yeast system. The highly specific limonene synthase from *Citrus limon* (*Cl*LimS) [34] produced a C_{11} blend dominated by 2-methyllimonene (70% of C_{11}; Fig. 3, Tables S2 and S6). *Solanum elaeagnifolium* camphene synthase (*Se*CamS) [35] produced 14 C_{11} compounds, including 2-methylmyrcene (55.2% of total C_{11}), 2M2B, 1MC, 2MB, traces of 2-methyllimonene and 9 additional C_{11} molecules, amounting to 6% of total terpenes produced (Fig. 3 and 4d, Tables S2 and S6).

Overall, the 7 plant monoterpene synthases tested produced 36 different C_{11} terpenes beyond the 2MIB/2MB blend (2MIB, 2MB, 2M2B and 1MC; Tables S2 and S5). Although all plant enzymes exhibited relative preference for GPP, their efficiency with 2meGPP in the yeast system ranged from approximately 25% C_{11} in the case of *Sf*CinS and *Pt*PinS to 1% in *Cl*LimS1 (Fig. 3). Many C_{11} compounds were synthesized exclusively by the plant terpene synthases. These results suggested that it

would be possible to expand the terpene universe in a predictable and systematic manner using an alternative substrate and terpene synthases from plants or other organisms. But since all the plant enzymes tested here still produced preferentially the canonical monoterpene products in the yeast system, we set out to develop dedicated C_{11} synthases by protein engineering.

Developing C_{11}-specific terpene synthases. Due to the availability of a crystal structure [30], *Sf*CinS1 was selected as the starting point to generate variants with improved selectivity towards 2meGPP. We examined the structure of *Sf*CinS1 in conjunction with that of the closely related limonene synthase from *Mentha spicata* in the presence of substrate analogues [36]. The early steps of the monoterpene synthase mechanism involve the binding and ionization of the GPP substrate, followed by the *syn*-migration of the diphosphate moiety to C-3 to yield linalyl diphosphate (LPP) in the transoid conformation (Fig. 5a). Subsequent rotation of the C-2,3 bond converts LPP to the cisoid conformer, which brings C-6 and C-1 at a position competent for cyclization (closed conformation). Another ionization event promotes cyclization to generate the α-terpinyl cation, which further isomerizes to the final products. Recent evidence from experiments with 2-fluoro-GPP and 2-fluoro-LPP suggested that initial substrate binding occurs at an extended conformation, as shown in Fig. 5c left panel, and not in the catalytically competent closed conformation (Fig. 5d right panel) [36]. This has been considered to be an essential requirement for the transoid to cisoid transition, because the required rotation around the C-2,3 bond would be forbidden by the position of the 6,7-double bond in the closed conformer of LPP [36]. As a result, the closed form is believed to be adopted only subsequently to the conversion of the transoid to the cisoid conformer.

Since the bulkier 2-propenyl substituent of 2-methyl-LPP would pose an even greater difficulty in the rotation of the C-2,3 bond, it is possible that canonical C_{10} terpene synthases could be less efficient with 2meGPP because the shape of their active site would be too narrow to fit an adequately extended configuration of 2meGPP. Thus, we selected two sites at the bottom of the active site cavity of *Sf*CinS1, namely N338 and I451 (Fig. 5c, d), which if substituted could provide the additional space required to facilitate the binding of the extended substrate conformation. A series of site-directed mutants were constructed to test this hypothesis (Table S7). Among the N338 variants, *Sf*CinS1(N338S) resulted in an approximately 66% improvement in total C_{11} levels and a shift to 2-methylmyrcene as the main product. Another N338 variant, *Sf*CinS1(N338A), was equally efficient with wild-type in overall C_{11} synthesis but was three times more specific for 2-methylmyrcene compared to other products (Table S7). Substitution of I451 by A, V or S resulted in improved or comparable C_{11} production and a shift in specificity towards 2-methylmyrcene in all cases (Table S7). In one of the I451 variants, *Sf*CinS1(I451A), 2-methylmyrcene reached 72% of the total C_{11} production. The two

most successful of these substitutions were combined and the double variant *Sf*CinS1(N388S-I451A) showed significant improvement in C_{11} yield and increased C_{11}-product specificity, producing almost exclusively (88% of C_{11}) 2-methylmyrcene. As a result, in *Sf*CinS1(N388S-I451A) 2-methylmyrcene became the main product among both C_{10} and C_{11} compounds (Fig. 6a). Kinetic analysis of recombinant *Sf*CinS1(N388S-I451A) suggested that the observed increase in C_{11} conversion by this enzyme was the result of a significant improvement in 2meGPP binding affinity, manifested by an 8-fold decrease in the apparent K_M (7.07 μM for the mutant vs. 57.7 μM for the wild-type; Table S8). This was in agreement with the hypothesis that the specific mutation would facilitate binding of the bulkier 2meGPP substrate. However, the overall catalytic efficiency of *Sf*CinS1(N388S-I451A) was significantly compromised (Tables S8, S9), prompting us to explore alternative approaches for the development of synthases suitable for C_{11} production in yeast.

Identification of a C_{11}-specific single-residue switch. Further examination of the available structural information highlighted one residue in *Sf*CinS1, F571, which could be involved in the conversion from the open to the closed substrate conformation. As shown in Fig. 5b, F571 is positioned in the trajectory of the transoid to cisoid transition and at a short distance from C-1 of LPP in the closed form and could play a stabilizing or regulatory role in the extensive conformational changes undergone by the substrate [36]. We substituted F571 in *Sf*CinS1 with different side chains and found that variants F571H and F571Y shifted production towards 2meGPP-derived compounds in yeast (Fig. 6a, 6d). In *Sf*CinS1(F571H), this substitution improved specificity for 2MB, which comprised 84.7% of C_{11} products (Fig. 6a; Table S10), while in *Sf*CinS1(F571Y) it shifted product specificity to 2-methyl-α-terpineol (58% of total C_{11}; Table S10). To understand the molecular basis of these changes, we carried out kinetic analysis of recombinant *Sf*CinS1(F571Y) and found that the substitution resulted in a 6-fold improvement in the affinity for 2meGPP and a significant decrease in the k_{cat}^{app} for GPP by 17-fold. Although there was also an 8-fold decrease in 2meGPP turnover, the apparent catalytic efficiency (k_{cat}^{app}/K_M^{app}) for 2meGPP surpassed that for GPP (Table S8).

Based on this finding, we examined an amino acid sequence alignment of *Sf*CinS1 with the other plant terpene synthases included in this study and found that residues analogous to F571 were typically an F or an H in the other enzymes (Fig. S9). We targeted these residues for site-directed mutagenesis (Table S5). Three limonene synthase variants, *Cl*LimS(H570V), *Cl*LimS(H570L) and *Cl*LimS(H570I) became 2-methyllimonene synthases producing 70-95% C_{11} compounds in yeast, of which 79-88% was 2-methyllimonene (Fig. 6c, 6d and Table S10). In *Pt*PinS, all variants tested switched specificity to become C_{11} synthases (between 60-88% of C_{11} products). *Pt*PinS(F607I) synthesized mainly 2MB, while *Pt*PinS(F607L) produced predominantly 2-methyllinalool (Fig. 6b, 6d

and Table S10). Moreover, in *Ob*MyrS, two variants, F579V and F579I, switched their preference to 2meGPP and produced 55% C_{11} terpenes giving rise to specific 2-methyllinalool synthases (Fig. 6d and Table S10). In *Sp*SabS, variants H561Y, V or L became promiscuous C_{11} synthases (60-70% C_{11} production) synthesizing 2MB as the main compound (Fig. 6d and Table S10). *Se*CamS(H583L) was also selective for 2meGPP producing mainly 1MC (Fig. 6d and Table S10).

To further understand the observed shifts in specificity, we analyzed the kinetic parameters of *Cl*LimS variants H570L, V and I, and *Se*CamS variant H583L, and found that the H570/H583 substitutions resulted in strong decreases in the k_{cat}^{app} for GPP but only small changes in the k_{cat}^{app} for 2meGPP (Table S8). As a result, the catalytic efficiency of the mutants shifted in favor of 2meGPP. To further corroborate these findings, we carried out *in vitro* substrate competition experiments using recombinant *Sf*CinS1(F571Y) and *Cl*LimS(H570L). We confirmed that at equal GPP and 2meGPP concentrations, both variants produced higher amounts of C_{11} than C_{10} products (Fig. S10). At a 2:1 2meGPP:GPP ratio, corresponding to the determined ratio in our engineered yeast cells, both *Cl*LimS(H570L) and *Sf*CinS1(F571Y) produced >75% C_{11} products (Fig. S10), explaining the dominance of C_{11} products in the product blend of yeast cells expressing these variants (Fig. 6d). These findings suggest that F571 in *Sf*CinS1, and related residues in other terpene synthases, likely perform a critical function in the reaction of canonical terpene synthases with GPP, which is, however, not equally critical for 2meGPP turnover. Thus, this residue can serve as a single-residue switch to change the substrate specificity of existing synthases and derive C_{11}-specific enzymes.

Overall, we obtained several dedicated C_{11} synthases specific for products other than those made by the bacterial 2MIB/2MB synthases. These include the 2-methyllimonene synthases *Cl*LimS(H570V, L or I), the 2-methyl-α-terpineol synthase *Sf*CinS1(F571Y), the 2-methyllinalool synthases *Ob*MyrS(F579V) and *Pt*PinS(F607L), and the 2-methylmyrcene synthase *Sf*CinS1(N388S-I451A). Although some of these variants were less efficient in C_{11} synthesis than their wild-type counterparts (yeast titer ranging from 56% to 14% of the corresponding wild-type enzyme, Table S9), having now established the principles for substrate selectivity, further protein engineering will enable the development of optimized C_{11} synthases for the highly efficient and specific synthesis of products desirable for industrial applications.

Discussion

The biosynthetic pathways of specialized metabolism are based on a handful of basic building blocks (e.g. aromatic amino acid-derived C_6-C_3 or C_6-C_1 blocks, acetyl CoA-derived C_2 blocks, isoprenoid C_5 blocks, etc.). This restricts the chemical space that can be accessed by specialized metabolites to the

complexity that can be generated using these basic scaffolds. Expanding specialized biosynthesis using additional building blocks can help explore uncharted areas of the chemical space and provide molecules with new or improved properties. To address this challenge, we focused on terpenoids because they are the most numerous group of plant natural products and have wide-spread applications as pharmaceuticals (e.g. taxol, artemisinin), flavors and fragrances (e.g. ambroxan), sweeteners (e.g. steviosides), neutraceuticals and colorants. To establish proof-of-principle, we started from the simplest group, the monoterpenoids, and hijacked a biosynthetic mechanism dedicated to the synthesis of specific microbial compounds (2MB/2MIB) to produce a new terpenoid precursor in yeast cells. We established a modular platform to harvest the chemical diversity that can be produced from this new building block in a systematic and predictable manner by acquiring parts from canonical monoterpene biosynthesis and engineering them to become specific for the new substrate. This chemical diversity can now be expanded further by introducing the subsequent modules of terpene biosynthesis, such as P450-driven oxidations, reductions, acetylations etc., to yield a whole class of non-canonical terpenoids. The efficient yeast production system established here will facilitate analysis of the properties of the new compounds and enable the development of new industrial products. Furthermore, the parts and know-how developed in yeast can be transferred to other organisms, such as bacteria or plants, to help investigate *in viv*o the potential biological functions of the novel compounds. The monoterpene building block GPP is also involved in the biosynthesis of other natural products, such as iridoids, monoterpene indole alkaloids, cannabinoids, and other prenylated aromatic compounds. Thus, establishing a yeast platform that synthesizes 2meGPP also opens up the possibility to engineer the synthesis of non-canonical variants of many other GPP-containing specialized metabolites, several of which belong to pharmacologically important groups (e.g. the monoterpene indole alkaloid anticancer agents vinblastine and vincristine).

This effort makes the first step towards creating a far broader diversity of non-canonical terpenoid structures. Larger terpenoids, such as sesqui-, di- and tri-terpenoids, have higher inherent complexity due to their larger size. Having established proof-of-concept with the production of C_{11} terpenes, the same tools and overall approach can now be extended to the other terpenoid classes to achieve the production of larger-size non-canonical terpenes. This can be achieved by applying gene mining and protein engineering to establish methylation of larger building blocks in the terpene biosynthetic ladder, such as FPP and GGPP. In combination with the increased potential for complex decoration of the larger scaffolds, this approach will result in an expansion of novel structures.

These findings also raise important questions regarding the extent of C_{11} terpene biosynthesis in bacteria or other organisms and the process of evolution of specialized metabolic pathways. Although

our results suggest that it is possible to produce a large number of C_{11} scaffolds from 2meGPP and to develop specific synthases for several C_{11} compounds from existing parts, such synthases seem to be absent from bacteria while the volatile bouquet of all microbial cultures studied to date is strictly dominated by 2MIB or 2MB. It is possible that the present situation is a snapshot of evolution taken at an early stage of the development of a new pathway. It would be very informative to analyze the continuously expanding microbial genomic information for clusters containing combinations of terpene synthase and methyltransferase genes and to study their evolution.

Our work used terpenoid biosynthesis to show that synthetic biology approaches can be successful in expanding the diversity of natural biosynthetic pathways. The same concept can now be applied to other groups of specialized metabolites, beyond terpenoids.

Online methods

Chemicals and enzymes. Standards used include: 2-methylisoborneol (Sigma, M3933-1ML), 1,8-cineole (Aldrich, C8,060-1), α-pinene (Aldrich, P-7408), γ-terpinene (Aldrich, T2134), β-myrcene (M-0382) and a 70% sabinene solution (kindly donated by VIORYL S.A., Athens, Greece). In-house standards of 2-methyllimonene, 2-methylmyrcene, 2-methyl-α-terpineol were obtained by extraction and partial purification of yeast cultures expressing *Pf*MBS. 2-methyllinalool and 2-methylgeraniol were obtained by hydrolysis under acidic conditions of the products of *in vitro* reactions of recombinant *Pl*GPPMT protein with SAM (A2408-Sigma) and GPP (G6772-Sigma). Phusion High-Fidelity DNA Polymerase (New England BioLabs, M0530S) and MyTaq DNA polymerase (BIO-21105, Bioline) were used in PCR amplifications. QIAquick Gel Extraction Kit (#28704, Qiagen) was used for gel extraction and DNA purification. NucleoSpin Plasmid Kit (740588.250, Macherey-Nagel) was used for plasmid DNA purification. Anti-c-Myc antibody from Sigma-Aldrich (cat no: M4439).

Yeast media. D-(+)-glucose monohydrate (16301, Sigma); D-(+)-galactose (G0625, Sigma); raffinose pentahydrate (R1030, US Biological); Yeast Nitrogen Base w/o AA (Y2025, US Biologicals); Complete Minimal (CM) medium was composed of 0.13% (w/v) dropout powder (all essential amino acids), 0.67% (w/v) yeast nitrogen base w/o AA, 2% glucose; For galactose based medium, glucose was substituted with 2% galactose, 1% raffinose.

Gene cloning and expression in yeast. All primers used for cloning purposes are listed in the Table S5. *Cloning of the P. limnetica GPPMT in an inducible yeast expression:* The codon optimized version of the GPP methyltransferase from *P. limnetica* (GenBank HQ630882.1) [19] was kindly provided by Professor David Cane (Brown University, USA) in bacterial expression plasmid pET28a(+) and was subcloned into an appropriate yeast expression vector. An inducible expression plasmid (pYESmyc)

was employed in order to minimize any adverse effects in yeast growth due to the activity of the methyltransferase. The pYESmyc vector allows for the galactose-inducible expression of an N-terminally myc-tagged version of the inserted gene. The *Pl*GPPMT insert was PCR amplified using pET28a(+)/*Pl*GPPMT as a template. PCR primers (GPPMT5-MfeNdeI and GPPMT3-NotXhoI) were designed to incorporate *Mfe*I and *Nde*I restriction sites immediately upstream of the ATG and *Not*I and *Xho*I sites right after the stop codon to allow for further subcloning of the amplified product. The PCR product was inserted into the pCRII-TOPO vector using the TOPO cloning methodology (Invitrogen Inc.) and insert-containing plasmids were selected by the release of an 800 bp fragment upon *Eco*RI restriction digestion. The pCRII-TOPO/GPPMT plasmid was then digested with *Mfe*I and *Xho*I restriction enzymes to release the GPPMT (~800 bp) insert, creating compatible ends to allow subcloning into the *Eco*RI and *Xho*I digested yeast expression vector pYESmyc. Following ligation, the construct was confirmed by restriction digestion and sequencing.

*Fusion of the Pl*GPPMT *protein with ERG20:* Digestion of pYESmyc/ERG20-GS [23] with *Eco*RI and *Xho*I restriction enzymes generated a linerized fragment. The *Pl*GPPMT gene was released from the pCRII-TOPO-*Pl*GPPMT plasmid through restriction digestion with compatible enzymes, *Mfe*I and *Xho*I. The obtained fragments of pYESmyc/ERG20-GS and *Pl*GPPMT were ligated to generate pYESmyc/ERG20-*Pl*GPPMT construct.

Cloning of the terpene synthases in compatible yeast and bacteria expression vectors: Yeast codon optimized version of *Citrus limon* (+)-S-limonene synthase (*Cl*LimS) was kindly provided as pCEV-G2-Ph/*Cl*LimS (GenBank AF514287.1) by Dr. Claudia E. Vickers, University of Queensland, Australia. The ORF was PCR amplified using primers LimS-BGL and LimS-XHOstop to introduce *Bgl*II and *Xho*I restriction sites at the 5' and 3' ends, respectively. The PCR product was initially introduced into vector pCRII-TOPO by TOPO cloning. The generated construct, pCRII-TOPO/*Cl*LimS was then digested using *Bgl*II and *Xho*I enzymes and the *Cl*LimS gene was excised and subcloned into vector pESC-TRP (Agilent Technologies, Cat. #217453) restricted with *Bam*HI and *Xho*I. In this construct, *Cl*LimS was expressed under the control of the P_{GAL1} promoter. Yeast codon optimized versions of *P. limnetica* 2-methylisoborneol syntahse (*Pl*MIBS) (GenBank ADU79148.1), *M. olivasterospora* 2-metylenebornene synthase (*Mo*MBS) (GenBank BAK26793.1), *P. fluorescens* pfl_1841 encoding for a 2-metylenebornene synthase (*Pf*MBS) (GenBank WP_011333305), *O. basilicum* myrcene synthase (*Ob*MyrS) (GenBank Q5SBP1.1), *P. taeda* α-pinene synthase (*Pt*PinS) (GenBank Q84KL3.1), *O. basilicum* geraniol synthase (*Ob*GerS) (GenBank Q6USK1.1) were obtained by gene synthesis bearing flanking regions containing specific restriction sites and a generic sequence compatible for USER cloning. Using the generic primes USER-Gen-FP and USER-Gen-RP, the above

genes were amplified and cloned by USER cloning into a generic backbone introducing *Bam*HI and *Eco*RI sites at the 5' end, and *Not*I, *Bgl*II and *Xho*I sites at the 3' end. The constructs where confirmed by sequencing. Subsequently, the genes of interest were excised from the generic USER vector by *Bam*HI and *Xho*I digestion and ligated into pESC-TRP vector (Agilent Technologies, Cat. #217453) linearized with *Bam*HI and *Xho*I restriction enzymes. This approach enabled subcloning of the genes of interest under P_{GAL1} promoter and resulted into construction of the following plasmids: pESC-TRP/*Pl*MIBS, pESC-TRP/*Mo*MBS, pESC-TRP/*Pf*MBS, pESC-TRP/*Ob*MyrS, pESC-TRP/*Pt*PinS, pESC-TRP/*Ob*GerS. Constructs of *S. fruticosa* 1,8-cineole synthase (*Sf*CinS1) (GenBank ABH07677.1), *S. pomifera* sabinene synthase (*Sp*SabS) (GenBank DQ785794.1) and *Se*CamS in appropriate yeast vectors were available from previous work [23,30,35,37].

For expression in bacteria the genes encoding *Pl*MIBS, *Pt*PinS, *Cl*LimS, *Ob*MyrS and *Se*CamS were cloned into pRSETa (Invitrogen). *Pl*MIBS, *Ob*MyrS and *Pt*PinS were digested from pESC-TRP/*Pl*MIBS, pESC-TRP/*Ob*MyrS and pESC-TRP/*Pt*PinS with *Bam*HI and *Xho*I. *Cl*LimS was then digested using *Bgl*II and *Xho*I enzymes from pCRII-TOPO/*Cl*LimS. *Se*CamS was amplified from pCRII-TOPO/*Se*CamS [35] using primers CamS-BamHI FP and CamS-SalI RP to introduce *Bam*HI and *Sal*I restriction sites. Following TOPO cloning, *Se*CamS insert was excised with *Bam*HI and *Sal*I restriction enzymes. The following inserts were produced: *Pl*MIBS (*Bam*HI/*Xho*I), *Ob*MyrS (*Bam*HI/*Xho*I), *Pt*PinS (*Bam*HI/*Xho*I), *Cl*LimS (*Bgl*II/*Xho*I) and *Se*CamS (*Bam*HI/*Sal*I). The resulted inserts were ligated into pRSETa vector digested with *Bam*HI and *Xho*I to generate the corresponding bacterial expression constructs: pRSETa/*Pl*MIBS, pRSETa/*Ob*MyrS, pRSETa/*Pt*PinS, pRSETa/*Cl*LimS, pRSETa/*Se*CamS. Selective mutations were introduced by USER mutagenesis [38] using pRSETa/*Cl*LimS (H570V, H570L and H570I) and pRSETa/*Se*CamS (H583L) as templates. The constructs pRSETa/*Sf*CinS1, pRSETa/*Sf*CinS1(N388S-I451A), pRSETa/*Sf*CinS1(F571Y) and pRSETa/Erg20p were already available from previous work [23,30].

Strain AM94. In *S. cerevisiae*, heterozygous deletions lead to 50% decrease in the level of the corresponding protein in the majority of yeast genes tested [26]. Strain AM94 (see Table S1), the basis strain used in this study, is a diploid strain that contains a heterozygous deletion in the squalene synthase gene ERG9, which is essential for sterol biosynthesis. This deletion helps reduce the drain of isoprenoid substrates and boost the heterologous production of terpenoids [39].

Development of the MIC1 yeast strain. To integrate *Pl*GPPMT in one of the two *ERG20* alleles of the diploid yeast strain AM94 the protocol previously described [37] was used. The *Pl*GPPMT gene was digested from the pCRII-TOPO/*Pl*GPPMT plasmid with *Mfe*I and *Xho*I restriction enzymes and cloned into *Eco*RI-*Xho*I restriction sites of the plasmid construct COD7 (GPPMT-CYC1t, LoxP-*HIS5*-LoxP)

[39]. The COD7/*Pl*GPPMT construct was PCR amplified using primers ERG20F and ERG20OL3, which incorporate flanking sequences complementary to the 5' and 3' end of the *ERG20* gene respectively. Following selection marker excision, strain MIC1 was obtained.

Yeast growth assay. Selected engineered yeast strains were grown in corresponding glucose-based media until OD_{600} ~0.5-0.7. The obtained cells were collected and washed twice with sterile water followed by dilution (in water) to OD_{600} ~0.1. Subsequently, the obtained cell suspensions were used to serial dilution of 1:10, 1:100 and 1:1000. The obtained dilutions of selected strains were plated (20 µL) on glucose-based or galactose/raffinose-based agar media and incubated at 30 °C for 3 days (Fig. S6a). Similarly, yeast cells grown in glucose media were washed twice with sterile water and resuspended in galactose/raffinose containing media to induce protein expression. Subsequently, the obtained cultures were diluted to OD_{600} ~0.1 (considered time 0) and 25 mL were incubated at 30 °C with 150 rpm shaking. Measurements of OD_{600} were taken every 2 h to monitor the curve of yeast growth (Fig. S6b). For OD_{600} values higher than 1, samples were diluted accordingly.

Terpene quantification and extraction from yeast cells. Selected *S. cerevisiae* strains were cultivated in 2 mL liquid media using 20 ml glass vials with magnetic screw cap (Mikrolab Aarhus A/S, Denmark). Solid Phase Microextraction (SPME) was applied for measuring the volatile terpenes produced in yeast cells, using 2 cm-50/30um DVB/Carboxen™/PDMS StableFlex™ Fiber followed by GC-TQ and GC-qTOF analysis as previously described [37]. For quantification, terpene extraction was performed using 1% (w/v) Diaion HP20 (Supelco, Bellefonte, PA) as adsorbent resin or dodecane extraction following the protocols previously described [23,40]. Samples were analyzed in triplicates and the error calculated as the mean absolute deviation (MAD) around the mean.

Protein expression in bacteria. Erg20p, *Pl*GPPMT and *Pl*MIBS were expressed in *E. cloni* EXPRESS DUOs BL21 [DE3] (Lucigen, Cat. #60401-2) cells. *Pt*PinS, *Cl*LimS, *Sf*CinS1 and *Se*CamS were expressed in *E. coli* BL21 [DE3, pLANT(3)/RIL [41]]. Expression and purification of Erg20p was performed as described in [23]. *Pl*GPPMT and *Pl*MIBS was expressed and purified following the protocol described in [19]. Bacterial expression of *Sf*CinS1, *Pt*PinS, and *Cl*LimS was performed as described in [30] with the following modifications. Bacterial cells carrying the corresponding plasmids were grown at 23 °C until OD_{600} reached 0.5-0.7, then induced with 0.1 mM isopropylthio-β-galactoside (IPTG), after which cell growth was continued for 20 h at 19 °C. The His-tagged proteins were then purified by Ni^{2+} affinity chromatography as described in [30]. The resulting protein fractions are shown in Fig. S4. Production of SpSabS1, *Ob*MyrS and their mutants resulted in the production of insoluble protein aggregates at all conditions tested and was not pursued further. Despite obtaining

soluble *Pt*PinS, the resulting protein preparations showed very low specific activity that precluded kinetic evaluation.

Total yeast extracts. Yeast cells expressing Erg20p(N127W)-*Pl*GPPMT were resuspended in enzymatic reaction buffer (10 mM MOPS (pH 7.0), 5 mM MgCl2, 1 mM DTT, 0.1 mg/mL BSA) supplemented with 1 mM PMSF, protease inhibitor cocktail (Roche) and lysozyme (0.1 mg/mL) and disrupted by sonication. The lysate was centrifuged at 13,000*g* at 4 °C for 20 min and the resulted supernatant was collected and stored at -80 °C until further use.

Enzymatic assays and kinetic analysis. Erg20p activity was assayed in 0.2 mL reactions as described in [23], using as co-substrates 10 µM IPP with 70 µM DMAPP, GPP or 2meGPP. Reactions also included 50 ng purified Erg20p or 50 µL of extract of yeast cells expressing Erg20p(N127W)-*Pl*GPPMT. Reactions were incubated for 1 h at 30 °C and then terminated by the addition of an equal volume (0.2 mL) of 2N HCl in 83% ethanol and 0.2 mL hexane. Following 20 min incubation at 37 °C to allow diphosphate hydrolysis, reactions were neutralized with 0.2 mL of 10% NaOH and the hexane phase was analyzed by GC-MS using the conditions described below.

Enzymatic activity of *Pl*GPPMT was assayed in 0.2 mL reactions as previously described [14,19] using as co-substrates 120 µM SAM and 70 µM of GPP, DMAPP, IPP, FPP or GGPP. The reactions were incubated 18 h at 30 °C and terminated by addition of 0.2 mL 2N HCl in 83% ethanol and 0.2 mL hexane. Acid hydrolysis of the diphosphates was carried for 20 min at 37 °C and the reactions were neutralized and analyzed as above.

The kinetic parameters of different terpene synthases (wild-type and mutants) were determined as previously described [30] using 20 mM MOPS, 20 mM $MgCl_2$, 0.02 mM $MnCl_2$, 1 mM DTT, 0.01 % (v/v) BSA, 50 ng enzyme and varying concentrations (0-150 µM) of substrates, GPP or 2meGPP. The reaction were overlaid with hexane containing 1 ppm nonane and 1 ppm dodecane as internal standards and incubated for 1 h with mild shaking at 30 °C. Extraction and GC-MS analysis were performed as previously described [30]. Kinetic analysis was performed using WinCurveFit 1.1.8 (Kevin Raner software).

All enzymatic assays and kinetic analysis experiments were carried out in triplicates.

Analysis of yeast sterols and squalene. Yeast cells resulting from 4 mL cultures induced to express Erg20p(N127W)-*Pl*GPPMT (or containing the empty pYESmyc vector) were treated with butylated hydroxytoluene (BHT) in a ratio of 2:1 per pellet weight, resupended in 500 µL 10% KOH in 80% EtOH solution and incubated at 70 °C for 1-2 h. Samples were cooled at room temperature and overlaid with 500 µL of hexane followed by vigorous vortexing prior to collecting the hexane phase. This step was repeated twice. The pooled hexane (~1.5 mL) was washed with water (equal volume) three times.

The collected hexane phase was evaporated down to 100 µL and derivatized with 150 µL trimethylsilyl cyanide (Sigma, Cat. #212849) for 30 min at 60 °C prior to GC-MS analysis [42].

Determination of intracellular prenyl diphosphates. The ratio of intracellular GPP and 2meGPP was determined with the method described in reference [23]. Briefly, 50 mL cultures of the selected MIC1 cells expressing the Erg20p(N127W)-*Pl*GPPMT(F266H) fusion were grown to mid-log phase and the cells were harvested washed twice with 10 mL H_2O. Following resuspension in 1 mL of H2O and disruption by glass beads, the cell extract was centrifuged at 13 000 for 10 min and 0.5 mL of the supernatant was mixed with 0.5 mL 2 N HCl in 83% ethanol and overlaid with 1 mL of hexane. The reactions were incubated at 37 °C for 20 min and neutralized by adding 0.35 mL of 10% NaOH. The mixtures were extracted with hexane and analyzed by GC-MS. The ratio of 2meGPP:GPP was determined by quantification of the linalool and 2-methyllinalool peaks that result from their hydrolysis.

GCMS analysis conditions. GCMS analysis was carried on a DB-5 column using hydrogen as a carrier gas with a constant velocity of 30 cm/sec. Samples resulting from incubation of the SPME fiber for 30 min over the head-space of 10 mL yeast cultures producing monoterpenes, C_{11} terpenes or sesquiterpenols or 2 µL of the hexane phase overlaying the in vitro enzymatic reactions were analyzed using the following temperature program: initial temperature 40 °C, ramp to 80 °C with a rate of 3 °C/min, ramp to 110 °C with a rate of 30 °C/min, ramp to 130 °C with a rate of 3 °C/min 280 °C with a rate of 30 °C/min, hold for 3 min. Sample resulting from non-saponifiable lipid extraction were analyzed for squalene and sterols using the following temperature program: initial temperature 60 °C, ramp to 300 °C with a rate of 10 °C/min, ramp to 320 °C with a rate of 3 °C/min, hold for 3.5 min.

Chemical synthesis of 2meGPP. 2meGPP was chemically synthesized by the general procedure of Wang and Cane [14], but including the following modifications (synthesis scheme shown in Fig. S11).

General chemical synthesis procedures. All reactions were monitored by TLC on aluminium sheets coated with silica gel 60F254 (0.2 mm thickness, Merck) and the components present were detected by charring with 10% H_2SO_4 in MeOH. Column chromatography was carried out using silica gel 60 (particle size 0.040-0.063 mm, 230-400 mesh ASTM, Merck). Solvent extracts were dried with anhydrous $MgSO_4$ unless otherwise specified. The ^1H and ^{13}C NMR spectra were recorded on a Bruker Avance 400 spectrometer at 400 and 101 MHz, respectively. $CDCl_3$ was used as solvent (unless otherwise indicated), δH values are relative to internal TMS and δC values are referenced to the solvent [δC ($CDCl_3$) = 77.0].

Ethyl (*E*/*Z*)-2,3,7-trimethylocta-2,6-dienoate (*43*). (By using the LiCl/DBU System [43] but solvent-free). A neat mixture of LiCl (0.80 g, 18.9 mmol), triethyl 2-phosphopropionate (***41***) (3.0 mL, 14.0 mmol) and DBU (1,8-diazabicyclo[5,4,0]undec-7-ene)(2.3 mL, 15.6 mmol) was stirred at room

temperature for 1.0 h under Ar followed by the addition of 6-methyl-5-hepten-2-one (***42***) (2.3 mL, 15.6 mmol). Stirring was continued at room temperature for 24 h. The reaction was quenched with water (50 mL) and the reaction mixture was extracted with Et_2O (200 mL). The extract was washed with brine and dried ($MgSO_4$). The solvent was evaporated and chromatographed on silica (100 g) with 0-10 % Et_2O in n-pentane to afford compound ***42*** as colorless oil (2.6 g, 88 %) *E*/*Z*-mixture.

Ethyl (*E*)-2,3,7-trimethylocta-2,6-dien-1-ol (*44*). (Using DIBAL-H as the reducing agent [44]). A stirred solution of Ethyl (*E*/*Z*)-2,3,7-trimethylocta-2,6-dienoate (***43***) (2.4 g, 11.4 mmol) in diethyl ether (50 mL) and cooled to -78 °C. DIBAL-H (1.0 M in toluene) (27.7 mL, 27.7 mmol) was added dropwise and the reaction mixture was stirred at -78 °C for 2 h and then quenched by addition of MeOH (5.0 mL). The reaction mixture was warmed gradually to 0 °C and diluted with Et_2O (100 mL) followed by the addition of saturated aqueous solution of Rochelle salt (100 mL), water (100 mL) and vigorously stirred. Once the Et_2O layer became clear, it was separated, dried over $MgSO_4$, filtered, and concentrated on a rotary evaporator under reduced pressure. Chromatographic purification of the resulting oil on silica (100 g) with 20 % Et_2O in n-pentane afforded the separation of the pure (*E*)-isomer ***44*** colorless oil (0.40 g, colorless oil, 21 %). ^1H and spectral data for ***44*** were identical with those reported[14].

(*E*)-2-Methylgeranyl pyrophosphate (2meGPP) [14,45]**.** (The Ph_3P/CBr_4 system [46] was used instead). To a stirred solution of ethyl (*E*)-2,3,7-trimethylocta-2,6-dien-1-ol (***44***) (0.10 g, 0.60 mmol) in dry benzene (5.0 mL) and dry CH_2Cl_2 (1.0 mL) was added CBr_4 (0.41 g, 1.20 mmol) and Ph3P (0.32 g, 1.20 mmol) at 0 °C under a nitrogen atmosphere. After the solution was stirred at the same temperature for 2 h, n-pentane (50 mL) was added to the reaction mixture. The reaction mixture was filtered to remove triphenylphosphine oxide. The filtrate was concentrated under reduced pressure to afford the crude (*E*)-2-methylgeranyl (0.14 g, quant.) as a yellow oil which was immediately subjected to the following pyrophosphorylation reaction without further purification. A solution of the crude bromide in dry CH_3CN (15 mL) was added dropwise to a stirred solution of $[(n\text{-}Bu_4)N]_3HP_2O_7$ (0.81 g , 0.9 mmol, 1.5 eq) in dry CH_3CN (15 mL) at 0 °C under Ar. The mixture was stirred at room temperature for 24 h and concentrated on a rotary evaporator below 35 °C. The obtained residue was transferred to two centrifuge tubes with 10 mL of acetone and concentrated NH_4OH (0.5-1 mL) was added to each. The precipitated ammonium salts, isolated by centrifugation (4000 rpm for 10 min), were washed twice by resuspension in 5-mL aliquots of acetone containing 0.01 N NH_4OH. The combined supernatants of the acetone containing NH_4OH were rota-evaporated at below 35 °C. The resulting crude 2meGPP as its ammonium salt was pure enough to be used for the enzymatic studies. However, a pure sample of

2meGPP was accomplished by preparative TLC-plates using the eluent system MeOH/CH_2Cl_2/0.01 N NH_4OH (1: 1.2: 0.3 v/v). When the synthesized organic diphosphate was subjected to acid hydrolysis, the recovered alcohol was compared with authentic samples by TLC and GC. In each case the recovered alcohol was essentially identical with the corresponding (*E*)-2-methylgeranyl alcohol (***44***).

Molecular graphics were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311) [47].

Mutagenesis. Site-directed mutagenesis of *Pl*GPPMT, *Sf*CinS1, *Sp*SabS, *Se*CamS was performed with the Quickchange method (Stratagene) and mutagenesis of *Ob*MyrS, *Pt*PinS, *Cl*LimS was performed by USER mutagenesis [38] using the primers listed in Table S5.

Construction of the *Pl*GPPMT model. The structural model of *Pl*GPPMT was constructed using the SWISS-MODEL server [48] and 3VC2 [27] as template.

Data availability statement. All data generated or analysed during this study are included in this published article (and its supplementary information files

Acknowledgments

We would like to thank Prof. David Cane and Dr. Wayne Chou (Brown University, USA) for providing the bacterial construct pET28a(+)/*Pl*GPPMT, Dr. Claudia E. Vickers, (University of Queensland, Australia) for providing construct pCEV-G2-Ph/*Cl*LimS, Assist. Prof. Fernando Geu-Flores (University of Copenhagen, Denmark) for critical reading of the manuscript, and Morten Raadam for technical assistance. We are grateful to Dr. David Ian Pattison and Eleni Lazaridi from the PLEN metabolomics platform for their support. This work was supported by the Greek General Secretariat of Research and Technology (GSRT) grant 11ΣΥΝ_3_770 (to AMM and SCK) and the Novo Nordisk Foundation grants NNF16OC0019554 (to CI) and NNF16OC0021760 (to SCK).

Author contributions

SCK conceived the project and designed experiments; CI designed experiments, engineered MIC1 strain, expressed and analyzed C_{10} and C_{11} wild-type and mutant synthases in yeast and bacteria, performed mutagenesis of *Pl*GPPMT, *Pt*PinS, *Sp*SabS, *Ob*MyrS, *Cl*LimS, *Se*CamS, produced the purified proteins, conducted in vitro enzymatic assays for the determination of the kinetic parameters; MP performed expression and analysis of *Pl*GPPMT, *Sf*CinS1 wt and mutants, *Sp*SabS wt and conducted yeast optimization for 2meGPP production, MSM performed the chemical synthesis of

2meGPP, MEM and AMM assisted in data analysis, SCK and CI analyzed the data and drafted the manuscript.

Competing financial interests

The authors declare no competing financial interest.

Supplementary Materials

Table S1 – S10

Fig S1 – S11

Figure legends

Figure 1. Terpenoid biosynthetic code. The biosynthesis of terpenoids is a modular process that follows an established pattern. In module 1, C_5 isoprenoid units are put together by prenyltransferase enzymes to produce universal diphosphate substrates with sizes that differ by 5-carbon increments. In module 2, each of these substrates forms the basis for a different terpenoid class (monoterpenoids, sesquiterpenoids, etc.), being converted to the different terpene backbones by one of the many terpene synthase enzymes. The basic terpene skeletons become decorated in subsequent modules involving different decorating enzymes, i.e. cytochrome P450s, O-acetyltransferase, O-methyltransferases, dehydrogenases, etc. The use of few building blocks and the modularity of the pathway establish a code for terpene biosynthesis. In this work, we hijacked the 2MIB pathway (inset in the right panel) to establish a parallel universe of C_{11} terpene scaffolds (right panel).

Figure 2. Production of 2meGPP in yeast. a. Expression of *Pl*GPPMT and *Pl*MIBS in yeast resulted in the production of the 2MIB degradation products 2MB, 1MC and 2M2B (blue), identified by comparison with the authentic 2MIB standard (pink). None of these compounds was produced by any of the two enzymes alone (green, brown) or by yeast cells carrying the two empty vectors (red). **b.** Diagram showing the synthesis of GPP, 2meGPP and derived isoprenoids in yeast. Synthesis of canonical monoterpenes (C_{10}; green) or non-canonical C_{11} terpenes (blue) also require GPP (TPS: terpene synthase). **c.** *Pl*GPPMT was fused with Erg20p. The N-terminal fusion, Erg20p-*Pl*GPPMT, enabled a 15-fold increase. **d.** Introduction of the Erg20p(N127W) variant in the Erg20p-*Pl*GPPMT fusion resulted in a 4.8-fold increase over the previous step. **e.** Strain MIC1 was derived by

chromosomal integration of *Pl*GPPMT in strain AM94. Both strains were engineered to overexpress *Pl*MIBS and the Erg20p(N127W)-*Pl*GPPMT fusion and their performance was compared. **f.** Mutagenesis of *Pl*GPPMT (non-fused form) identified two variants, V250A and F266H, with improved performance when introduced into AM94 together with *Pl*MIBS. **g.** Bar chart showing a summary of the yield improvement achieved in each of the previous steps (shown in this figure in panels c, d, e and f). In the experiments shown in panels c, d, e and f, samples were analyzed in triplicates and the error bars represent the mean absolute deviation (MAD) around the mean.

Figure 3. Percentage of C_{11} and C_{10} terpene scaffolds produced by the different wild-type terpene synthases in yeast. Samples were analyzed in triplicates, yield of each compound was averaged, and the percentages of C_{11} and C_{10} compounds were calculated.

Figure 4. Production of C_{11} terpenoids in yeast by plant monoterpene synthases. a. Co-expression of *Sf*CinS1 with Erg20p(N127W)-*Pl*GPPMT (blue) resulted in the production of a blend of C_{11} terpenoids. Compound ***30***, which accounted for 11.2% of the total C_{11} production, displays a 168 m/z fragment (Table S6) characteristic of C_{11} terpene alcohols. **b.** *Pt*PinS co-expressed with Erg20p(N127W)-*Pl*GPPMT (blue) produced 29 C_{11} molecules representing almost 26% of the total terpene production. Compounds ***31*** – ***35*** are specific to plant enzymes. **c.** In the presence of Erg20p(N127W)-*Pl*GPPMT, *Sp*SabS (blue) produced a blend of C_{11} compounds dominated by 2-methylmyrcene and 2MB. **d.** *Se*CamS (blue) showed preference for 2-methylmyrcene (55%) and the 2-MIB dehydration products. Yeast samples of single transformations expressing the either the corresponding monoterpene synthase alone (green), Erg20p(N127W)-*Pl*GPPMT alone (brown), or the two empty vectors (red) are shown as controls.

Figure 5. a. Proposed GPP cyclization mechanism. Initially, the substrate binds in the extended conformation and the transition from the transoid extended conformation to the catalytically competent cisoid closed conformation is essential for the reaction (according to [36]). **b.** Proposed mechanism of 2meGPP cyclization. **c.** Model of the active site of SfCinS1 with the "extended" form of 2-fluorolinalyl diphosphate (2F-LPP) superimposed. The extended conformation is believed to be achieved by on-enzyme conversion of 2-fluorogeranyl diphosphate (2F-GPP) in the crystal. The structure of the substrate analog in the extended conformation is considered to resemble that of the early steps of the mechanism. **d.** Model of the active site of SfCinS1 with the "closed" structure of 2-fluorolinalyl

diphosphate (2F-LPP) superimposed. This analogue adopts the closed cisoid conformation that is competent for cyclization. Extensive conformational changes are needed for the substrate to modify its structure, and F571 is positioned in the heart of this transition. The residues selected for mutagenesis are indicated. Graphic produced with UCSF Chimera [47].

Figure 6. Engineering dedicated C_{11} terpene synthases. a. Mutagenesis of *Sf*CinS1. The double mutant N338S-I451A lost its ability to synthesize 1,8-cineole and became a dedicated 2-methylmyrcene synthase, while substitution of F571 with H resulted in a 2MB-specific synthase (shown in blue). The product profile of the wild-type enzyme is shown in green. The asterisk (*) denotes a yeast-produced non-terpene compound. **b.** Substitution of H570 in *Cl*LimS with V or L (shown in blue) changed the specificity of this enzyme from limonene to 2-methyllimonene. **c.** Mutagenesis of *Pt*PinS. The product profile of the wild-type enzyme is shown in green. Variant F607L became a dedicated 2-methyllinalool synthase, while variant F607I turned into a 2MB synthase (blue). Peaks marked a star correspond to non-terpenoid compounds. **d.** Bar chart showing the specificity of the different plant monoterpene synthases and their variants for C_{10} (green) or C_{11} terpene production in yeast (blue). In panel D, samples were analyzed in triplicate and the mean value of total C_{10} and C_{11} products is shown. All panels show results from the yeast system.

References

1 Dewick, P. M. Medicinal natural products: a biosynthetic approach. (Wiley, Chichester, UK, 2009).

2 Pichersky, E., Noel, J. P. & Dudareva, N. Biosynthesis of plant volatiles: nature's diversity and ingenuity. *Science* **311**, 808-811, doi:10.1126/science.1118510 (2006).

3 Gershenzon, J. & Dudareva, N. The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **3**, 408-414 (2007).

4 Ignea, C. *et al.* Reconstructing the chemical diversity of labdane-type diterpene biosynthesis in yeast. *Metab. Eng.* **28**, 91-103, doi:10.1016/j.ymben.2014.12.001 (2015).

5 Jiang, J., He, X. & Cane, D. E. Biosynthesis of the earthy odorant geosmin by a bifunctional Streptomyces coelicolor enzyme. *Nat. Chem. Biol.* **3**, 711-715, doi:10.1038/nchembio.2007.29 (2007).

6 Herde, M. *et al.* Identification and Regulation of TPS04/GES, an Arabidopsis Geranyllinalool Synthase Catalyzing the First Step in the Formation of the Insect-Induced Volatile C16-Homoterpene TMTT. *Plant Cell* **20**, 1152-1168 (2008).

7 Lee, S. *et al.* Herbivore-induced and floral homoterpene volatiles are biosynthesized by a single P450 enzyme (CYP82G1) in Arabidopsis. *Proc. Natl Acad. Sci. USA* **107**, 21205-21210, doi:10.1073/pnas.1009975107 (2010).

8 Magnard, J. L. *et al.* PLANT VOLATILES. Biosynthesis of monoterpene scent compounds in roses. *Science* **349**, 81-83, doi:10.1126/science.aab0696 (2015).

9 Richter, A. *et al.* Characterization of Biosynthetic Pathways for the Production of the Volatile Homoterpenes DMNT and TMTT in Zea mays. *Plant Cell* **28**, 2651-2665, doi:10.1105/tpc.15.00919 (2016).

10 Vedula, L. S. *et al.* Exploring biosynthetic diversity with trichodiene synthase. *Arch. Biochem. Biophys.* **466**, 260-266 (2007).

11 Rising, K. A. *et al.* Formation of a Novel Macrocyclic Alkaloid from the Unnatural Farnesyl Diphosphate Analogue Anilinogeranyl Diphosphate by 5-Epi-Aristolochene Synthase. *ACS Chem. Biol.* **10**, 1729-1736, doi:10.1021/acschembio.5b00145 (2015).

12 Karahadian, C., Josephson, D. B. & Lindsay, R. C. Volatile compounds from Penicillium sp. contributing musty-earthy notes to Brie and Camembert cheese flavors. *J. Agric. Food. Chem.* **33**, 339-343 (1985).

13 Komatsu, M., Tsuda, M., Omura, S., Oikawa, H. & Ikeda, H. Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol. *Proc. Natl Acad. Sci. USA* **105**, 7422-7427 (2008).

14 Wang, C. M. & Cane, D. E. Biochemistry and molecular genetics of the biosynthesis of the earthy odorant methylisoborneol in Streptomyces coelicolor. *J. Am. Chem. Soc.* **130**, 8908-8909 (2008).

15 Dickschat, J. S. *et al.* Biosynthesis of the off-flavor 2-methylisoborneol by the myxobacterium Nannocystis exedens. *Angew. Chem. Int. Ed. Engl.* **46**, 8287-8290, doi:10.1002/anie.200702496 (2007).

16 Brock, N. L., Ravella, S. R., Schulz, S. & Dickschat, J. S. A detailed view of 2-methylisoborneol biosynthesis. *Angew. Chem. Int. Ed. Engl.* **52**, 2100-2104, doi:10.1002/anie.201209173 (2013).

17 Dickschat, J. S. Bacterial terpene cyclases. *Nat. Prod. Rep.* **33**, 87-110, doi:10.1039/c5np00102a (2016).

18 Chou, W. K., Ikeda, H. & Cane, D. E. Cloning and characterization of Pfl_1841, a 2-methylenebornane synthase in Pseudomonas fluorescens PfO-1. *Tetrahedron* **67**, 6627-6632, doi:10.1016/j.tet.2011.05.084 (2011).

19 Giglio, S., Chou, W. K., Ikeda, H., Cane, D. E. & Monis, P. T. Biosynthesis of 2-methylisoborneol in cyanobacteria. *Environ. Sci. Technol.* **45**, 992-998, doi:10.1021/es102992p (2011).

20 Yamada, Y. *et al.* Terpene synthases are widely distributed in bacteria. *Proc. Natl Acad. Sci. USA* **112**, 857-862, doi:10.1073/pnas.1422108112 (2015).

21 Schumann, R. & Pendleton, P. Dehydration products of 2-methylisoborneol. *Water Res.* **31**, 1243-1246, doi:10.1016/S0043-1354(96)00330-2 (1997).

22 Zhou, Y. J. *et al.* Modular pathway engineering of diterpenoid synthases and the mevalonic acid pathway for miltiradiene production. *J. Am. Chem. Soc.* **134**, 3234-3241 (2012).

23 Ignea, C., Pontini, M., Maffei, M. E., Makris, A. M. & Kampranis, S. C. Engineering monoterpene production in yeast using a synthetic dominant negative geranyl diphosphate synthase. *ACS Synth. Biol.* **3**, 298-306, doi:10.1021/sb400115e (2014).

24 Stanley Fernandez, S. M., Kellogg, B. A. & Poulter, C. D. Farnesyl diphosphate synthase. Altering the catalytic site to select for geranyl diphosphate activity. *Biochemistry* **39**, 15316-15321 (2000).

25 Jiang, Z., Kempinski, C., Bush, C. J., Nybo, S. E. & Chappell, J. Engineering Triterpene and Methylated Triterpene Production in Plants Provides Biochemical and Physiological Insights into Terpene Metabolism. *Plant Physiol.* **170**, 702-716, doi:10.1104/pp.15.01548 (2016).

26 Springer, M., Weissman, J. S. & Kirschner, M. W. A general lack of compensation for gene dosage in yeast. *Mol.Syst. Biol.* **6**, 368 (2010).

27 Koksal, M., Chou, W. K., Cane, D. E. & Christianson, D. W. Structure of 2-methylisoborneol synthase from Streptomyces coelicolor and implications for the cyclization of a noncanonical C-methylated monoterpenoid substrate. *Biochemistry* **51**, 3011-3020, doi:10.1021/bi201827a (2012).

28 Chou, W. K., Gould, C. A. & Cane, D. E. Incubation of 2-methylisoborneol synthase with the intermediate analog 2-methylneryl diphosphate. *J. Antibiot. (Tokyo)* **70**, 625-631, doi:10.1038/ja.2017.24 (2017).

29 Citron, C. A., Barra, L., Wink, J. & Dickschat, J. S. Volatiles from nineteen recently genome sequenced actinomycetes. *Org. Biomol. Chem.* **13**, 2673-2683, doi:10.1039/c4ob02609h (2015).

30 Kampranis, S. C. *et al.* Rational conversion of substrate and product specificity in a salvia monoterpene synthase: structural insights into the evolution of terpene synthase function. *Plant Cell* **19**, 1994-2005 (2007).

31 Phillips, M. A., Wildung, M. R., Williams, D. C., Hyatt, D. C. & Croteau, R. cDNA isolation, functional expression, and characterization of (+)-alpha-pinene synthase and (-)-alpha-pinene synthase from loblolly pine (Pinus taeda): stereocontrol in pinene biosynthesis. *Arch. Biochem. Biophys.* **411**, 267-276 (2003).

32 Iijima, Y. *et al.* The biochemical and molecular basis for the divergent patterns in the biosynthesis of terpenes and phenylpropenes in the peltate glands of three cultivars of basil. *Plant Physiol.* **136**, 3724-3736 (2004).

33 Iijima, Y., Gang, D. R., Fridman, E., Lewinsohn, E. & Pichersky, E. Characterization of geraniol synthase from the peltate glands of sweet basil. *Plant Physiol.* **134**, 370-379 (2004).

34 Lucker, J. *et al.* Monoterpene biosynthesis in lemon (Citrus limon). cDNA isolation and functional analysis of four monoterpene synthases. *Eur. J. Biochem.* **269**, 3160-3171 (2002).

35 Tsaballa, A. *et al.* Use of the de novo transcriptome analysis of silver-leaf nightshade (Solanum elaeagnifolium) to identify gene expression changes associated with wounding and terpene biosynthesis. *BMC Genomics* **16**, 504, doi:10.1186/s12864-015-1738-3 (2015).

36 Hyatt, D. C. *et al.* Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. *Proc. Natl Acad. Sci. USA* **104**, 5360-5365 (2007).

37 Ignea, C. *et al.* Improving yeast strains using recyclable integration cassettes, for the production of plant terpenoids. *Microb. Cell Fact.* **10**, 4 (2011).

38 Genee, H. J. *et al.* Software-supported USER cloning strategies for site-directed mutagenesis and DNA assembly. *ACS Synth. Biol.* **4**, 342-349, doi:10.1021/sb500194z (2015).

39 Ignea, C. *et al.* Positive genetic interactors of HMG2 identify a new set of genetic perturbations for improving sesquiterpene production in Saccharomyces cerevisiae. *Microb. Cell Fact.* **11**, 162 (2012).

40 Trikka, F. A. *et al.* Combined metabolome and transcriptome profiling provides new insights into diterpene biosynthesis in S. pomifera glandular trichomes. *BMC Genomics* **16**, 935, doi:10.1186/s12864-015-2147-3 (2015).

41 Finkelstein, J., Antony, E., Hingorani, M. M. & O'Donnell, M. Overproduction and analysis of eukaryotic multiprotein complexes in Escherichia coli using a dual-vector strategy. *Anal. Biochem.* **319**, 78-87 (2003).

42 Ignea, C. *et al.* Carnosic acid biosynthesis elucidated by a synthetic biology platform. *Proc. Natl Acad. Sci. USA* **113**, 3681-3686, doi:10.1073/pnas.1523787113 (2016).

43 Zaed, A. M. & Sutherland, A. Stereoselective synthesis of the bicyclic guanidine alkaloid (+)-monanchorin. *Org. Biomol. Chem.* **8**, 4394-4399, doi:10.1039/c0ob00219d (2010).

44 Winterfeldt, E. Applications of Diisobutylaluminium Hydride (DIBAL) and Triisobulyaluminium (TIBA) as Reducing Agents in Organic Synthesis. *Synthesis*, 617-630 (1975).

45 Dixit, V. M., Laskovics, F. M., Noall, W. I. & Poulter, C. D. Tris(tetrabutylammonium) hydrogen pyrophosphate. A new reagent for the preparation of allylic pyrophosphate esters. *J. Org. Chem.* **46**, 1967-1969 (1981).
46 Miyata, O., Ozawa, Y., Ninomiya, I. & Naito, T. Radical Cyclization in Heterocycle Synthesis. Part 10:1 A Concise Synthesis of (-)-Kainic Acid via Sulfanyl Radical Addition-Cyclization-Elimination Reaction. *Tetrahedron* **56**, 6199-6207 (2000).
47 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
48 Peitsch, M. C. ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochem. Soc. Trans.* **24**, 274-279 (1996).