SCIENCE CHINA
Life Sciences

•RESEARCH PAPER•

July 2019 Vol.62 No.7: 947–958
https://doi.org/10.1007/s11427-019-9521-2

Recently duplicated sesterterpene (C25) gene clusters in *Arabidopsis thaliana* modulate root microbiota

Qingwen Chen[1,8†], Ting Jiang[1,4,8†], Yong-Xin Liu[1,4], Haili Liu[2], Tao Zhao[3], Zhixi Liu[1,8], Xiangchao Gan[5], Asis Hallab[6], Xuemei Wang[1,8], Juan He[1,8], Yihua Ma[1,8], Fengxia Zhang[1], Tao Jin[7], M. Eric Schranz[3], Yong Wang[2], Yang Bai[1,4,8*] & Guodong Wang[1,8*]

[1]*State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China;*
[2]*Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China;*
[3]*Biosystematics Group, Wageningen University, 6708 PB Wageningen, The Netherlands;*
[4]*CAS-JIC Centre of Excellence for Plant and Microbial Sciences (CEPAMS), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China;*
[5]*Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Köln, Germany;*
[6]*Forschungszentrum Jülich, Plant Sciences (IBG-2) Wilhelm-Johnen-Straße, 52428 Jülich, Germany;*
[7]*China National Genebank-Shenzhen, the Beijing Genomics Institute (BGI-Shenzhen), Shenzhen 518083, China;*
[8]*College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing 100039, China*

Received February 21, 2019; accepted March 25, 2019; published online May 10, 2019

Land plants co-speciate with a diversity of continually expanding plant specialized metabolites (PSMs) and root microbial communities (microbiota). Homeostatic interactions between plants and root microbiota are essential for plant survival in natural environments. A growing appreciation of microbiota for plant health is fuelling rapid advances in genetic mechanisms of controlling microbiota by host plants. PSMs have long been proposed to mediate plant and single microbe interactions. However, the effects of PSMs, especially those evolutionarily new PSMs, on root microbiota at community level remain to be elucidated. Here, we discovered sesterterpenes in *Arabidopsis thaliana*, produced by recently duplicated prenyltransferase-terpene synthase (PT-TPS) gene clusters, with neo-functionalization. A single-residue substitution played a critical role in the acquisition of sesterterpene synthase (sesterTPS) activity in Brassicaceae plants. Moreover, we found that the absence of two root-specific sesterterpenoids, with similar chemical structure, significantly affected root microbiota assembly in similar patterns. Our results not only demonstrate the sensitivity of plant microbiota to PSMs but also establish a complete framework of host plants to control root microbiota composition through evolutionarily dynamic PSMs.

plant specialized metabolites, microbiota, sesterterpene, terpene synthase

Citation: Chen, Q., Jiang, T., Liu, Y.X., Liu, H., Zhao, T., Liu, Z., Gan, X., Hallab, A., Wang, X., He, J., et al. (2019). Recently duplicated sesterterpene (C25) gene clusters in *Arabidopsis thaliana* modulate root microbiota. Sci China Life Sci 62, 947–958. https://doi.org/10.1007/s11427-019-9521-2

†Contributed equally to this work
*Corresponding authors (Yang Bai, email: ybai@genetics.ac.cn; Guodong Wang, email: gdwang@genetics.ac.cn)

INTRODUCTION

In nature, land plants featured with continually expanding plant specialized metabolites (PSMs) have co-evolved with a diverse community of soil microorganisms, and recruit specific root microbiota, which play pivotal roles in plant health and productivity (Berendsen et al., 2012; Müller et al., 2016; Verbon and Liberman, 2016; Leach et al., 2017). Meanwhile, plants face strong selective pressure to control their root microbiota to adapt or survive in natural environments, given the huge diversity and evolutionary dynamism of microbes. Thus, plants have evolved multiple mechanisms to control root microbiota, as humans have done to control their gut microbiota (Foster et al., 2017). Several recent studies have identified processes that control root microbiota assembly in *Arabidopsis thaliana*, including salicylic acid dependent immune system and phosphate starvation pathways (Lebeis et al., 2015; Castrillo et al., 2017). However, the key processes characterized to date can only explain limited variation observed in root microbiota, suggesting that host plants almost certainly use additional molecular mechanisms to manipulate root microbiota.

PSMs have long been proposed to function in key roles that mediate interactions between host plants and microbes in their environment (van Dam and Bouwmeester, 2016). The shared lineage and tissue specificity and the evolutionary dynamism of PSM pool and plant microbiota also support this hypothesis (Bulgarelli et al., 2012; Lundberg et al., 2012; Schlaeppi et al., 2014). However, the mutual relationships between PSMs and root microbiota, i.e., how and to what extent they are related, are largely unknown. Among various PSMs, terpenoids are the largest and the most diverse group; more than 70,000 terpenoids have been structurally characterized (Vickers et al., 2014). Despite their enormous structural and functional diversity, all terpenoids originate via *de novo* synthesis from two common C5 precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). These precursors are subsequently modified by prenyltransferase (PT) and terpene synthase (TPS) enzymes, which together generate the diversity of terpenoid backbones (Chen et al., 2011). There are 32 *TPS* genes in the *Arabidopsis thaliana* genome, 18 of which are expressed at various levels in root tissues (Tholl and Lee, 2011), implying that terpenoid diversity is critically important for root physiological processes in *Arabidopsis thaliana*. Interestingly, three PT-TPS gene clusters had been discovered by genome sequence analysis in *Arabidopsis thaliana* fifteen years ago (Aubourg et al., 2002). Till recently, we and other groups elucidated the biochemical function of these PT-TPS gene clusters, which showed unprecedented sesterterpene synthase (sesterTPS) activity utilizing geranylfarnesyl pyrophosphate (GFPP, C25) as a substrate (Wang et al., 2016; Huang et al., 2017; Shao et al., 2017). However, the *in planta* functions of these PT-TPS gene clusters remain unknown.

Here, we provide genetic and biochemical evidences to characterize two root-specific sesterterpenoids produced by recently duplicated gene clusters (*TPS25-GFPPS3* and *TPS30-GFPPS4*) in *Arabidopsis thaliana*. We further found that a single-residue substitution played a critical role in the acquisition of sesterTPS activity in Brassicaceae plants. The absence of two pentacyclic ring sesterterpenoids in loss-of-*sesterTPS*-function mutants significantly affected the *Arabidopsis thaliana* root microbiota composition. The sesterterpenoid-mediated plant mechanism for controlling root microbiota establishes a conceptual framework of host plants to modulate root microbiota composition through evolutionarily new PSMs.

RESULTS AND DISCUSSION

Evolutionarily new sesterterpene gene clusters in *Arabidopsis thaliana*

The *Arabidopsis thaliana* genome contains four geranylfarnesyl pyrophosphate synthases (GFPPS, C25) genes. Conspicuously, each *GFPPS* gene is adjacent to a terpene synthase (*TPS*) gene, followed by one or more *P450* genes, thereby forming TPS-GFPPS-P450 gene clusters. One large cluster includes three *TPS* genes, two GFPPS genes, and eight *P450* genes (TPS17-GGPPS5-TPS18-GFPPS1-TPS19-GFPPS2-P450$_{1-8}$, Figure 1A), while two small clusters include TPS25-GFPPS3 and TPS30-GFPPS4-P450 (Figure 1A) (Wang et al., 2016). The products of TPS18 (Compound 1, (+)-thalianatriene), TPS19 (Compound 2, (−)-retigeranin B), TPS25 (Compound 3, (−)-*ent*-quiannulatene) and TPS30 (Compound 4, (+)-astellatene) were identified using NMR techniques (Huang et al., 2017; Shao et al., 2017). These compounds are unprecedented *Arabidopsis thaliana* sesterterpenes largely similar to those isolated from fungi (Figure 1A). We next attempted to understand the evolutionary trajectory of the PT-TPS gene clusters. Phylogenomic synteny analysis revealed that an ancestral TPS-GFPPS cluster formed after an α whole-genome duplication event; orthologous clusters were present in Brassicaceae but not in any Cleomaceae (Figure 1B). Notably, TPS18-GFPPS1, TPS19-GFPPS2, TPS25-GFPPS3, and TPS30-GFPPS4 clusters were specific to the *Arabidopsis thaliana* genome via relatively recent duplications of the TPS17-GGPPS5 cluster; no counterparts for these gene clusters were found in *Arabidopsis lyrata* and other Brassicaceae genomes. Moreover, the TPS25-GFPPS3 and TPS30-GFPPS4-P450 clusters were located in dynamic chromosomal regions that are enriched in transposable elements (TEs). These results suggest that gene duplication and TE-mediated recombination contribute to the recent formation of TPS25-GFPPS3 and TPS30-GFPPS4-P450 clusters in *Arabidopsis*

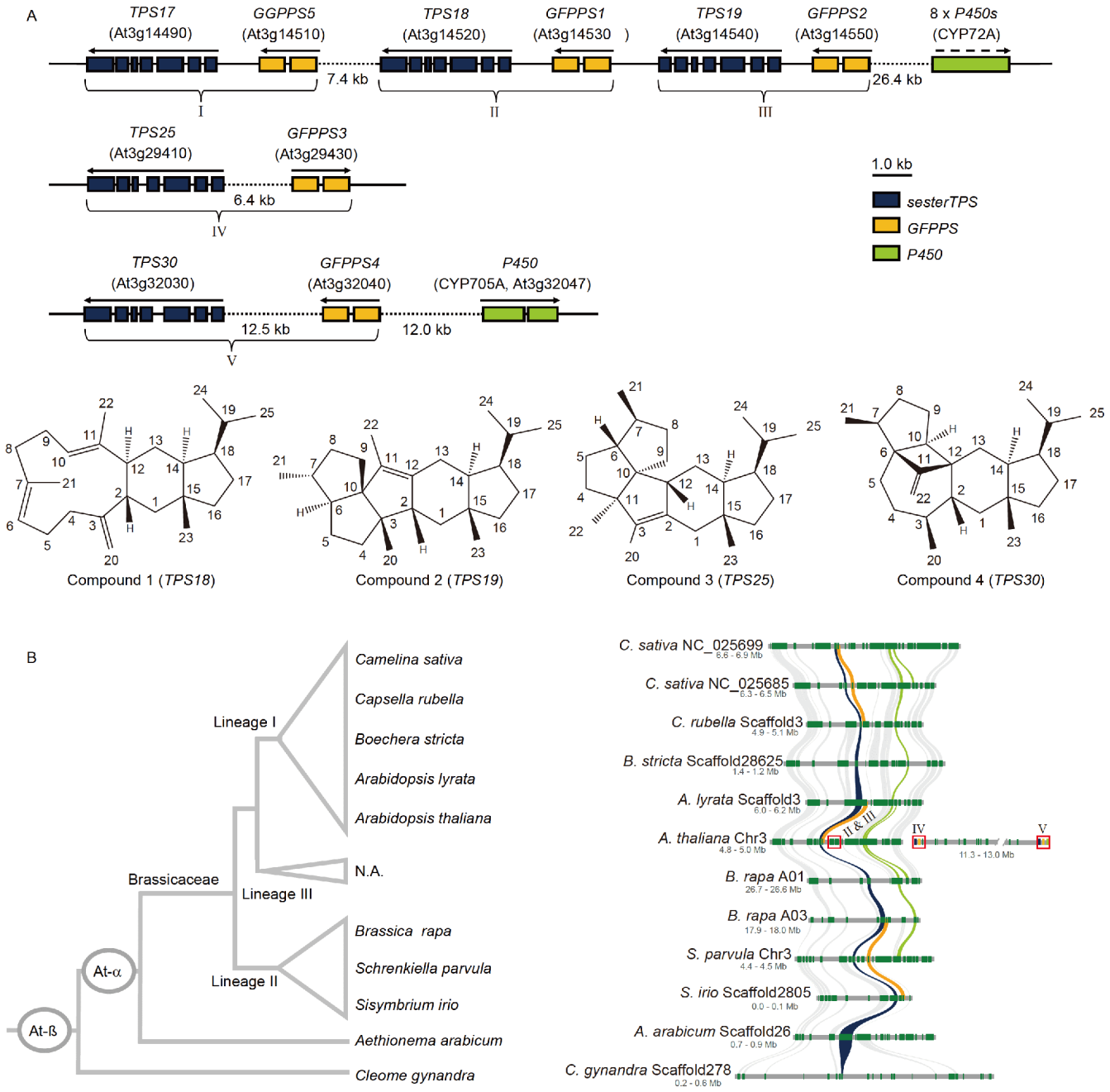
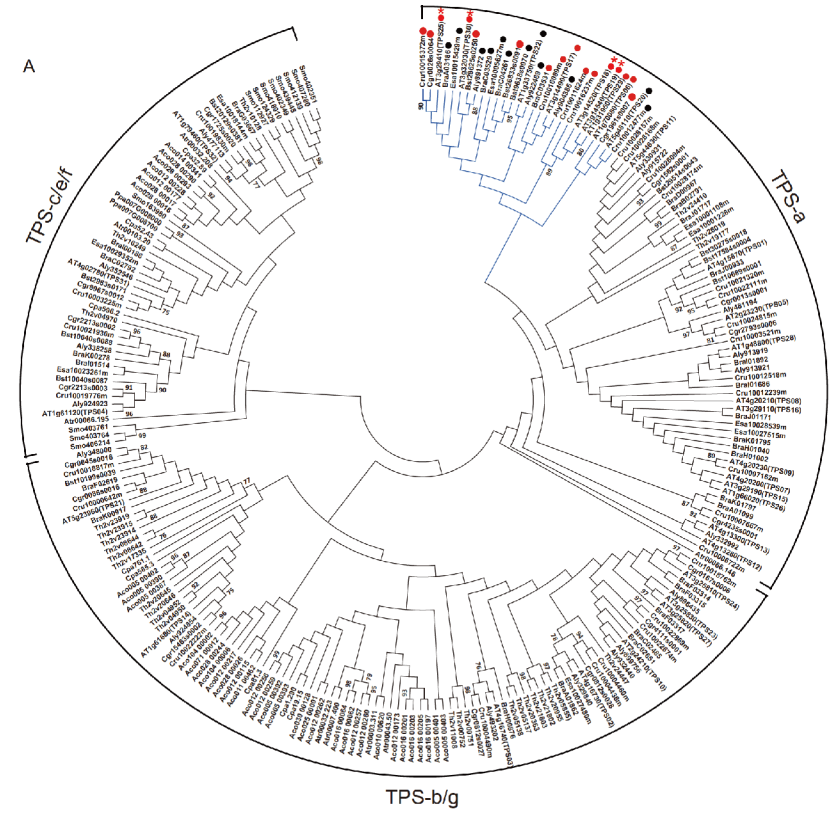


Figure 1 Brassicaceae-wide synteny of TPS-GFPPS-P450 gene clusters. A, Illustration of prenyltransferase-terpene synthase (PT-TPS) gene clusters in *Arabidopsis thaliana* and chemical structures of their corresponding products. Five putative sesterterpene (C25) gene clusters in the *Arabidopsis* genome. All of these gene clusters are located on chromosome 3. *GGPPS5* (*At3g14510*) is a pseudogene. Dotted lines represent multiple genes in this region. The chemical structures of four sesterterpenes generated by the *Arabidopsis* clustered sesterTPS are shown below. B, Simplified molecular phylogenetic tree (left) of Brassicaceae species that were used for the synteny analysis. The positions of the well-known ancient α and β whole-genome duplication events are indicated. Phylogenomic synteny analysis (right) of TPS-GFPPS-P450 gene clusters (cyan: *TPS*; orange: *GFPPS*; light green: *P450*) from 10 representative Brassicaceae and Cleomaceae genomes. Chromosome segments are represented by gray bars, and genes are represented by dark green bricks. The positions of the *TPS18-GFPPS1*, *TPS19-GFPPS2*, *TPS25-GFPPS3*, and *TPS30-GFPPS4* clusters are highlighted with red boxes. N.A. indicates that genome information of Brassicaceae lineage III species is not publicly available.

thaliana.

A single amino acid determines the substrate specificity of sesterTPSs

As TPSs are determinant enzymes for the diversity of the terpene backbone, we explored the phylogeny of sesterTPSs in 246 TPSs from 13 representative species on plant evolutionary history. Four aforementioned *Arabidopsis* sesterTPSs and their close homologues were only present in Brassicaceae genomes and formed an evolutionarily conserved subclade (28 members) within the TPS-a family (Figure 2A),



Enzyme	GFPP (C25)	GGPP (C20)	FPP (C15)	GPP (C10)
TPS18	✓	×	×	×
TPS19	✓	×	×	×
TPS25	✓	×	×	×
TPS30	✓	×	×	×
TPS18^{G328W}	×	✓	×	×
TPS19^{G325W}	×	✓	×	×
TPS25^{G328W}	×	×	✓	✓
TPS30^{P328W}	×	×	×	×

Figure 2 Substrate specificity of sesterTPSs from brassicaceae plants. A, Phylogenetic analysis of 246 TPS proteins obtained from 13 representative species on plant evolutionary history. To simplify the classification of TPS proteins, three TPS clades (TPS-a, TPS-b/g and TPS-c/e/f) are shown here. Bootstrap values (based on 500 replicates) >75% are shown for corresponding nodes. All 28 putative sesterTPS proteins in blue form a Brassicaceae-specific clade. The 16 TPS proteins showing sesterTPS activity and 12 TPS proteins showing no sesterTPS activity in this clade are marked in red dots and black dots, respectively. Four clustered sesterTPSs from *Arabidopsis* are marked with additional red asterisks. Species abbreviations: Aco, *Aquilegia coerulea* Goldsmith; Aly, *Arabidopsis lyrata*; At, *Arabidopsis thaliana*; Atr, *Amborella trichopoda*; Cgr, *Capsella grandiflora*; Cpa, *Carica papaya*; Cru, *Capsella rubella*; Bst, *Boechera stricta*; Bra, *Brassica rapa*; Esa, *Eutrema salsugineum*; Ppa, *Physcomitrella patens*; Smo, *Selaginella moellendorffii*; Tha, *Tarenaya hassleriana*. It should be noteworthy that *AtTPS20* is a pseudogene in *Arabidopsis thaliana* (Col-0 ecotype). B, Amino acid sequence alignment of four *Arabidopsis thaliana* clustered sesterTPSs and three plant TPSs with solved crystal structures. Tobacco 5-epi-aristolochene synthase (TEAS, Protein Data Bank entry 5EAS), (4*S*)-limonene synthase from spearmint ((4*S*)-LS, Protein Data Bank entry 20NH), and 1,8-cineole synthase from *Salvia fruticosa* (SfCinS1, Protein Data Bank entry 2J5C). Sixteen residues lining the active site, mainly based on the 5EAS structure information, are denoted by asterisks. C, Substrate specificities of native and mutated *Arabidopsis* sesterTPSs. All four sesterTPSs showed specific sesterTPS (GFPP, C25) activity; however, mutagenized variants lost sesterTPS activity. Notably, TPS18^{G328W}, TPS19^{G325W} and TPS25^{G328W} showed monoTPS (GPP, C10), sesquiTPS (FPP, C15), or diTPS (GGPP, C20) activities.

which use geranyl diphosphate (GPP, C10), farnesyl diphosphate (FPP, C15) and/or geranylgeranyl diphosphate (GGPP, C20) as substrates (Chen et al., 2011). Importantly, we found that 16 out of 28 members of the sesterTPS clade used geranylfarnesyl diphosphate (GFPP, C25) as a substrate (Figure 2A and Figure S1 in Supporting Information), suggesting that functional sesterTPSs in Brassicaceae utilizing GFPP may evolve from TPSs utilizing GPP/FPP/GGPP.

To explore the evolution of sesterTPS in Brassicaceae plants, we scanned for conserved sites under recent positive selection in the sesterTPS clade (47 TPS genes were identified from eight Brassicaceaen genomes, see MATERIALS AND METHODS for detailed information) (Gan et al., 2016). Interestingly, only a single site (G328 (Glycine) in TPS18/TPS25, G325 in TPS19, and P328 (Proline) in TPS30) showed significant signals of selection (adjusted *P*-value<0.004), suggesting that positive selection was a driving force for the functional diversification of plant sesterTPS. Moreover, we aligned four clustered sesterTPSs in *Arabidopsis thaliana* (TPS18/TPS19/TPS25/TPS30) to monoTPSs and sesquiTPS with crystal structure (Starks et al., 1997; Kampranis et al., 2007; Srividya et al., 2015) to explore the enzymatic mechanism for sesterTPS substrate specificity, given that all known plant TPS-a members share a similar active-site scaffold (Christianson, 2017). Among 16 conserved residues in substrate binding and catalysis, the tryptophan (W) residue (position 273 in 5EAS; Figure 2B) in monoTPS and sesquiTPS proteins was replaced with G in TPS18/TPS19/TPS25 or P in TPS30 (Figure 2B). Together with positive selection result, this residue position (G328 in TPS18 and TPS25, G325 in TPS19 and P328 in TPS30) played a critical role during the evolutionary acquisition of sesterTPS activity among Brassicaceae plants. We hypothesized that the small side chains of G and P likely provide a large active-site cavity for the larger GFPP substrate. To test this hypothesis, we mutated G328 in TPS18 and TPS25, G325 in TPS19 and P328 in TPS30 to W. All four sesterTPSs showed specific activity towards GFPP (C25); however, mutagenized variants lost sesterTPS activity. More importantly, TPS18^{G328W}, TPS19^{G325W}and TPS25^{G328W} showed monoTPS (GPP, C10), sesquiTPS (FPP, C15) or diTPS (GGPP, C20) activities (Figure 2C, Figures S2–S5 in Supporting Information).

Tissue specificity of sesterterpenes in *Arabidopsis thaliana*

qPCR analysis revealed that *TPS25* and *TPS30* were predominantly expressed in roots, while *TPS18* and *TPS19* were mainly expressed in flower/silique tissue. The corresponding *GFPPS*s showed similar degrees of tissue specificity as the clustered *sesterTPS* genes (Figure 3A). To test the subcellular localization of four clustered sesterTPSs, we expressed sesterTPS:GFP in protoplasts. The GFP signals for all four clustered sesterTPSs were detected exclusively in chloroplasts (Figure 3B), suggesting that plant sesterterpenes (C25) are derived from the plastid-localized 2-*C*-methylerythritol 4-phosphate (MEP) pathway rather than the cytosol-localized mevalonic acid (MVA) pathway.

To test whether *Arabidopsis* plants produce sesterterpenoids, we performed sesterterpenoid-profiling of eight *Arabidopsis* organs using gas chromatography-triple quadrupole-mass spectrometry (GC-QQQ-MS). We found that Compound 3 (TPS25) and Compound 4 (TPS30) specifically accumulated in root tissue (Figure 3C). Interestingly, we found that old roots had two-fold more Compound 3 (44.31±1.17 vs. 17.83±0.66 μg g^{-1} fresh weight (F.W.), *P*<0.001, student-test) and 9-fold more Compound 4 (20.55 ±0.69 vs. 2.18±0.093 μg g^{-1} F.W., *P*<0.001, student-test) comparing with young roots (Figure 3C). We obtained *Arabidopsis* mutants defective in *TPS25* and *TPS30* (*tps25-1*, *tps25-2*, *tps30-1*, *tps30-2*, two double mutants: *tps25/tps30-1*, and *tps25/tps30-2*; Figures S6–S9 in Supporting Information). Compounds 3 and 4 were not detected in the corresponding *TPS25* and *TPS30* mutants (Figure 3D, Figure S10 in Supporting Information). Based on these results, we concluded that *TPS25* and *TPS30* encode functional *bona fide* sesterTPS in *Arabidopsis thaliana*.

Sesterterpenoids modulate the composition of root microbiota

Due to the root specificities of Compounds 3 and 4 and the increase in their accumulation with time throughout root development, we tested the effects of these sesterterpenoids on root microbiota. We performed 16S rRNA gene profiling to compare the root bacterial microbiota compositions from wild-type (Col-0) plants and the *TPS25* and *TPS30 sesterTPS* mutants (*tps25-1*, *tps25-2*, *tps30-1*, *tps30-2*, *tps25*/*tps30-1*, *tps25*/*tps30-2*) grown in natural soil (Chang-ping Farm in Beijing, China) under climate-controlled conditions. For each plant genotype, we performed three independent biological replicates. Each biological replication was tested in 4–5 individual pots, each containing four plants. The Col-0 and *sesterTPS* mutant plants appeared healthy when harvested (Figure S11 in Supporting Information).

We found that loss-of-function mutants of the *TPS25* and *TPS30* genes affected root microbiota assembly. Constrained ordination analysis revealed significant differences in the root microbiota between Col-0 and the various *sesterTPS* mutants (Figure 4A, 9.28% constrained variance, canonical analysis of principal coordinates; Figure S12 in Supporting Information). A hierarchical clustering analysis (post hoc Bray-Curtis distance) also showed that the microbiota of Col-0 and *sesterTPS* single and double mutants each formed distinctive clusters (Figure 4A and B). Importantly, the ef-

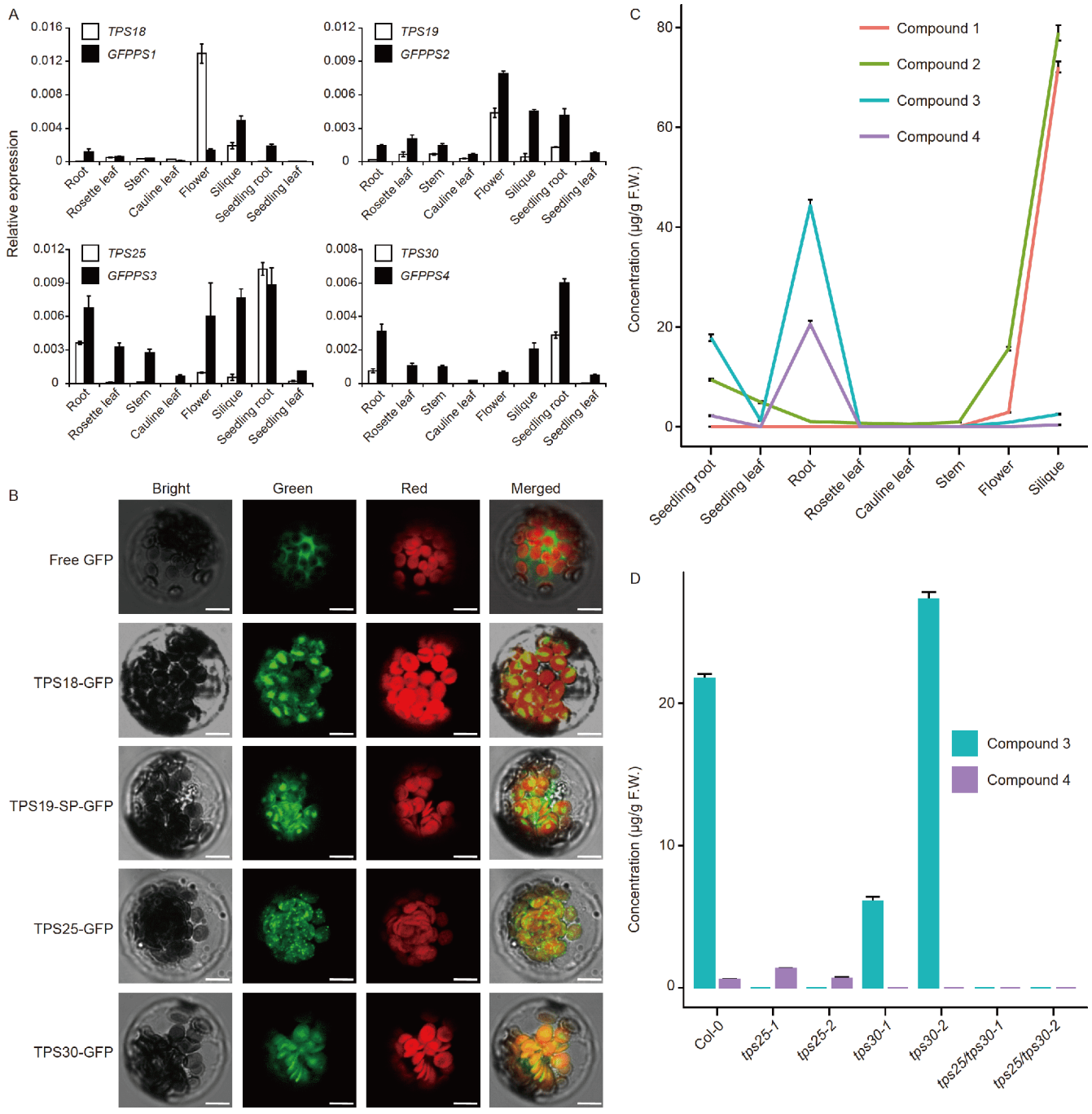


Figure 3 Characterization of four sesterTPSs in *Arabidopsis*. A, RT-qPCR analysis of two *GFPPS-sesterTPS* clustered genes in different organs of *Arabidopsis* plants. Samples were harvested from 8-week-old plants and 2-week-old seedlings. Transcript levels are expressed relative to the *Actin2* gene (n=3). B, Subcellular localization of TPS18-GFP, TPS19-SP-GFP, TPS25-GFP and TPS30-GFP in *Arabidopsis* leaf mesophyll protoplasts revealed by laser confocal microscopy. Chloroplasts are revealed by red chlorophyll autofluorescence. Scale bars=10 μm. C, Sesterterpene profiling (Compounds 1 to 4) in eight different organs of *Arabidopsis thaliana*. D, Sesterterpene profiles in roots of single and double mutants of *TPS25* and *TPS30*. Error bars represent the SE (n=3). F.W., fresh weight.

fects of *tps25*, *tps30*, and double mutants on the root microbiota composition were all validated in two independent mutant lines. The differential bacterial operational taxonomic units (OTUs) between mutants and Col-0 showed significant similarity, with overlaps of 44%–79% in two independent lines of each gene (Figure 4C). The differentially abundant OTUs mainly belonged to Proteobacteria, Actinobacteria, Firmicutes, Acidobacteria, Planctomycetes, and Verrucomicrobita (Figure 4D). These differences were found at a lower taxonomic OTU level, but not at a higher taxonomic phylum level (Figures S13 and S14 in Supporting Information). Notably, *TPS25* and *TPS30* deficiencies af-

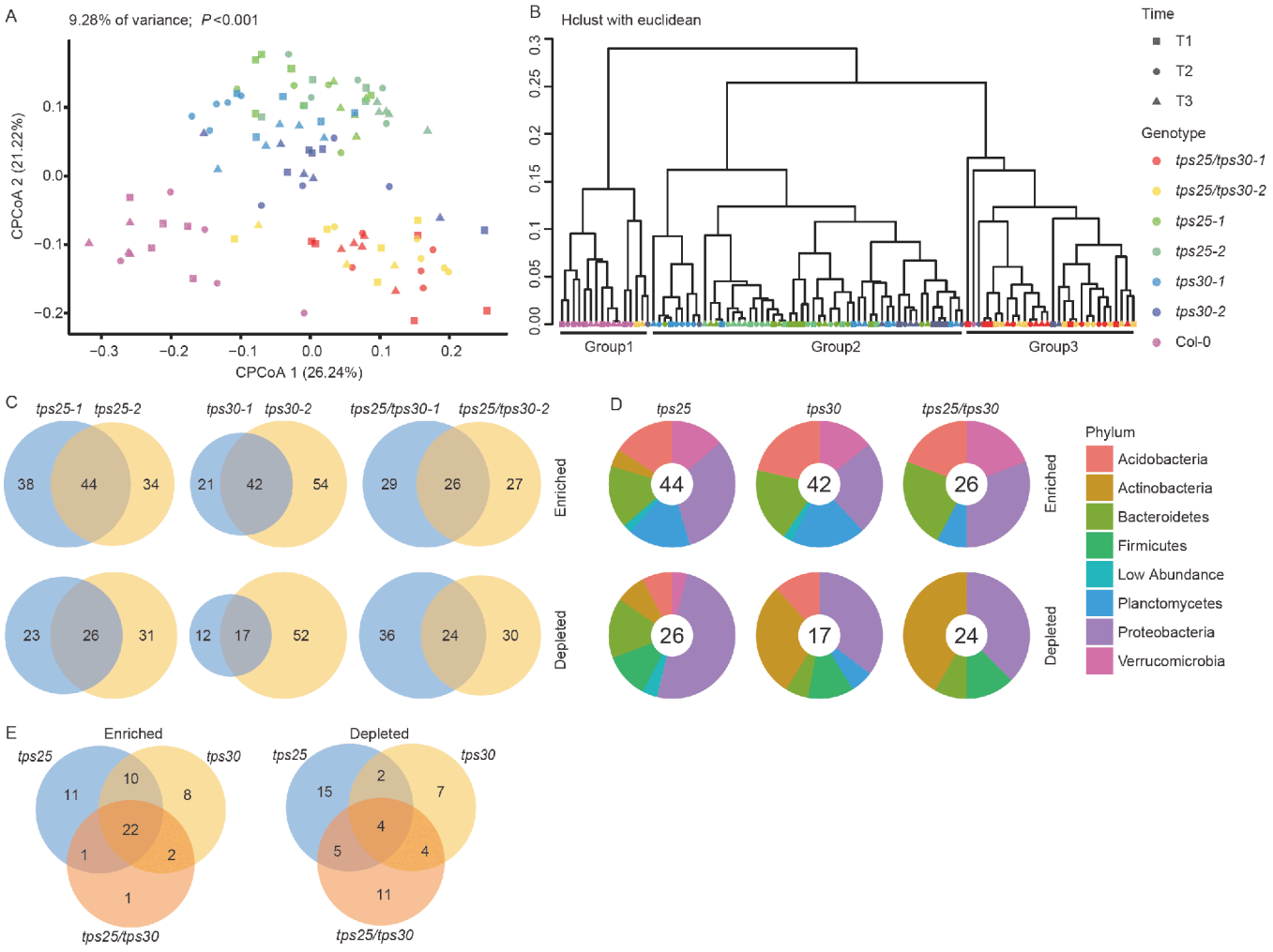


Figure 4 Loss-of-function mutations in the sesterterpenoid biosynthesis genes *TPS25* and *TPS30* result in similar effects on root microbiota assembly. A, Constrained principle coordination analysis (CPCoA) based on Bray-Curtis dissimilarity for 16S rRNA gene profiling data showing that *TPS*25 and *TPS*30 single mutants (*tps25-1*, *tps25-2*, *tps30-1*, *tps30-2*) and double mutants (*tps25/tps30*-1 and *tps25/tps30*-2) assembled distinct root microbiota comparing with wild-type plants (Col-0). Importantly, similar patterns were demonstrated with two independent mutant lines for *TPS25*, *TPS30* and double mutants. Independent biological experiments (T1, T2, and T3) are indicated by shapes. One biological replicate for each genotype consisted of root organ materials from 4-5 independent pots. The total replicate numbers are as follows: Col-0 (n=14); *tps25-1* (n=15); *tps25-2* (n=15); *tps30-1* (n=15); *tps30-2* (n=14); *tps25/tps30*-1 (n=15); *tps25/tps30*-2 (n=13). B, Hierarchical clustering of post hoc Bray-Curtis distances of root microbiota from Col-0, *tps25*, *tps30* and *tps25/tps30* mutants clusters in 3 groups. Group 1 mainly contains Col-0 samples, group 2 mainly contains *TPS25* and *TPS30* single mutants (*tps25*-1, *tps25-2*, *tps30-1*, *tps30-2*), and group 3 mainly contains double mutants (*tps25/tps30*-1 and *tps25/tps30*-2). C, Two independent lines of *TPS25* mutants (*tps25-1* and *tps25-2*), *TPS30* (*tps30-1* and *tps30-2*) mutants and double mutants (*tps25/tps30*-1 and *tps25/tps30*-2) showed significant overlaps of enriched and depleted OTUs compared with Col-0, validating the *sesterTPS* function to modulate root microbiota composition. D, Phylum-level distributions of bacterial OTUs identified as either enriched or depleted in each mutant comparing with Col-0. The number of total differential OTUs is indicated in the centre. E, Overlap of enriched or depleted OTUs in *tps25* and *tps30* single mutants and *tps25/tps30* double mutants. Remarkably, the overlapping proportion of enriched OTUs in various mutant plants was significantly higher than that of depleted OTUs.

fected root microbiota in similar patterns based on two lines of evidence: first, the root microbiota of the *tps25* and *tps30* mutants formed closely related clusters (Figure 4A and B); second, the differentially abundant OTUs between Col-0 and *tps25* roots showed a significant overlap with those between Col-0 and *tps30* mutants, especially among the enriched OTUs (Figure 4E, Figure S15 in Supporting Information). This trend was further confirmed by the enriched or depleted OTUs in *tps25/tps30* mutants (Figure 4E). Together, our results demonstrate that *TPS25* and *TPS30* affect root microbiota assembly; moreover, the extensive overlap of the bacterial taxa in both mutants strongly suggests that these recently duplicated sesterterpene gene clusters influence root microbiota assembly through related mechanisms. One possible explanation is that TPS25 product (Compound 3) and TPS30 product (Compound 4) share a similar chemical structure, both belong to pentacyclic structure: Compound 3 is a 5/5/5/6/5 pentacyclic sesterterpene and Compound 4 is a 5/4/7/6/5 pentacyclic sesterterpene (Figure 1A).

Land plants continually make new chemical molecules for fitness due to the never-ending selective pressure from plant-environment interactions (Kliebenstein and Osbourn, 2012).

In this study, we demonstrate that the recently duplicated GFPPS-sesterTPS clusters are responsible for sesterterpene production in *Arabidopsis thaliana*. We further provide genetic evidence to demonstrate that two root-specific sesterterpenes have significant impacts on root microbiota assembly (Figure 4A, 9.28% constrained variance). Our results suggest that the dynamic processes of duplication and functional divergence of metabolic genes (or gene clusters in this study) endow host plants with a powerful and flexible mechanism to control root microbiota. In addition to sesterterpenoids (C25), plants produce a myriad of other terpenoids (C10, C15, C20, C30, etc.) and other classes of PSMs (e.g., alkaloids, polyphenols, and other compounds) (Hartmann, 2007). Thus, it will be important to further test the effects of different classes of PSMs on controlling root microbiota by plant hosts. Our work opens a new avenue for the potential application of PSMs in modulating the root microbiota with the long-term goal of enhancing plant health and productivity.

MATERIALS AND METHODS

Plant materials and chemicals

The *Arabidopsis thaliana* lines (Col-0 and various transgenic plants under Col-0 background) were grown on soil in a growth room or on 1/2 solidified MS medium plates in a growth chamber at 22 °C with a 16-h light/8-h dark photoperiod. The T-DNA insertion mutants, *tps25-1* (SALK_123505), *tps25-2* (SALK_020266), and *tps30-1* (CS805958) were ordered from ABRC (http://abrc.osu.edu/). The *tps30-2* and *tps30-3* mutants were generated using the CRISPR/Cas9 technique. The *tps30/tps25-1* and *tps30/tps25-2* double mutants were generated by mutating *TPS30* using the CRISPR/Cas9 technique in the *tps25-1* mutant background. All authentic reference standards used in this study were purchased from Sigma. All DNA constructs generated in this study were verified by sequencing.

Heterologous expression in *Escherichia coli*, TPS protein purification, Terpene synthase assays, and enzyme characterization

All truncated sester*TPS* genes (removal of signal peptide sequence) were cloned into the pMAL-c2x vector (New England Biolabs). An MBP tag was added at the C termini of the truncated *TPS* genes. Construct transformation into *E. coli* Rosetta (DE3), protein induction, and MBP-tagged TPS protein purification were carried out by following standard protocols. Quantification and evaluation of the relative purity of the recombinant proteins was performed using SDS-PAGE with BSA (Sigma) as a standard. Assays testing whether or not the four *Arabidopsis* sesterTPSs could recognize GPP, FPP, or GGPP as substrates *in vitro* employed a previously described SPME-GC-MS method for analyzing volatile terpenes (Chen et al., 2015). It should be noteworthy that all other cDNA sequences of putative Brassicaceae sesterTPSs were synthesized by TaKaRa Bio.

RT-qPCR analysis and subcellular localization

RNA extraction, reverse transcription reaction, and RT-qPCR were performed as described previously (Wang et al., 2008; Li et al., 2015). The *Actin 2* (*At3g18780*) gene was chosen as internal standard gene due to its relatively stable expression across all tested samples. GFP fusion protein construct generation (pJIT163-hGFP vector), *Arabidopsis* leaf protoplast preparation and transformation, and laser scanning confocal microscope based image collection were all performed as described previously (Xu et al., 2013).

GC-MS analysis

For quantification of sesterterpenes in *Arabidopsis* using GC-QQQ-MS (Agilent 7890B/7000C), 100 or 200 mg of fresh plant tissue was collected and ground into a fine powder in liquid nitrogen. The resulting samples were extracted with 8 mL of EtOAc for 1 h, the supernatant was dried by nitrogen gas and re-dissolved into 200 μL of EtOAc. 1 μL of analytical sample was loaded on GC-QQQ-MS for analysis and quantification. The initial oven temperature was held at 60 °C for 1 min, then ramped at 10 °C min^{-1} to 325 °C, and kept at 325 °C for 10 min. The inlet temperature was 270 °C. Positive mode EI ionization was used. The temperatures of the ion source and quadrupole were set at 230 and 150 °C, respectively. Mass spectra were acquired within a scanning range of 50−600 *m*/*z*. The MRM settings for Compounds 1–4 are described in Table S1 in Supporting Information. The endogenous amounts of these four sesterterpenes were quantified based on a standard curve generated using the Compound 4 that we purified.

Phylogenetic and phylogenomic synteny analysis

The sequences of *Arabidopsis* TPS proteins and the TPS proteins from 11 other plant species were extracted from the TAIR (http://www.arabidopsis.org) and Phytozome 11 databases (http://www.phytozome.net). Twenty-seven Tarenaya TPS sequences were extracted from http://genomevolution.org/CoGe/CoGeBlast.pl. A maximum likelihood tree was constructed using MEGA6 software (Tamura et al., 2013). Shared synteny analysis of GFPPS-TPS-P450 gene clusters using 10 representative Brassicaceae and Cleomaceae genomes was carried out as described previously (Zhao et al., 2017).

For molecular evolutionary analysis of sesterTPS, total 47

sesterTPS genes were identified in eight Brassicaceaen genomes, including *Cardamine hirsuta*, *Capsella rubella*, *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Aethionema arabicum*, *Brassica rapa*, *Schrenkiella parvula* and *Eutrema salsugineum*, using a reciprocal BLAT and Markov clustering (MCL) based method (Gan et al., 2016). Coding sequences were aligned based on the respective multiple amino acid sequence alignment generated with MAFFT (v6.851b) (Katoh et al., 2002). Subsequently, the resulting multiple coding sequence alignment (MSA) was manually curated and two un-aligned blocks from the sequence of gene aet_18203 of species *A. arabicum* were removed. It was then submitted to phylogenetic maximum likelihood reconstruction (Murrell et al., 2012), with the resulting tree rooted using the midpoint-rooting method. Based on this phylogeny and the MSA, a mixed effects model of evolution (Murrell et al., 2012) was fitted and queried for amino acids showing signs of Darwinian selection.

Bacterial 16S rRNA gene profiling of *Arabidopsis* roots and unplanted control soil

Arabidopsis thaliana plants were grown in natural soil (40°5′49″, 116°24′44″E, Beijing, China) under greenhouse conditions (22 °C, 50% humidity, 16-h light/8-h dark photoperiod). Roots of 5 week-old plants (just after bolting) were harvested, washed twice in water, washed twice in PBS, shaken twice at 180 r min^{-1} for 20 min, and frozen at −80 °C (Bulgarelli et al., 2012). Frozen root and corresponding bulk soil samples were homogenized using a Bertin Precellys®24 tissue lyser (four times, at 6,200 r min^{-1} for 30 s each time). DNA was extracted using a FastDNA SPIN Kit for soil (MP Biomedicals) according to product instructions. The DNA concentration was measured using a PicoGreen dsDNA Assay Kit (Life Technologies, USA), and diluted to 3.5 ng μL^{-1}. 16S rRNA genes were amplified using primers targeting the variable regions V5-V7 (799F and 1193R) (Bulgarelli et al., 2012). Each sample was amplified in triplicate using 30 μL reactions containing 3 μL template, 0.75 U PrimeSTAR HS DNA Polymerase, 1× PrimSTAR Buffer (Takara, Japan), 0.2 mmol L^{-1} dNTPs, 10 pmol L^{-1} of barcoded forward and reverse primers with linker regions. The PCR conditions were as below. Denaturation at 98 °C for 30 s, followed by 25 cycles of 98 °C for 10 s, 55 °C for 15 s and 72 °C for 1 min, and a final elongation at 72 °C for 5 min. If no visible amplication from negative control (no template added), triplicate PCR products were pooled and purified using an AMPure XP Kit (Beckman Coulter, USA), and measured by Nanodrop (NanoDrop™ 2000C, Thermo Scientific, USA), and adjusted 10 ng μL^{-1}. During the second step PCR, 3 μL template was amplified in triplicates by 0.75 U PrimeSTAR HS DNA Polymerase, 1× PrimSTAR Buffer (Takara), 0.2 mmol L^{-1} dNTPs, 10 pmol L^{-1} of barcoded forward and reverse primers with linker regions. The PCR conditions were as below. Denaturation at 98 °C for 30 s, followed by 8 cycles of 98 °C for 10 s, 55 °C for 15 s and 72 °C for 1 min, and a final elongation at 72 °C for 5 min. PCR products of triplicates were subsequently combined, purified, and sequenced with the Illumina HiSeq 2500 Ultra-High-Throughput platform.

Bioinformatics analysis

The 16S rRNA gene sequences were processed using QIIME 1.9.1 (Caporaso et al., 2010b), USEARCH 8.0 (Edgar, 2010), and in-house scripts as described previously (Zhang et al., 2018). QIIIME mapping files are provided in External Database S1). Paired-end Illumina reads were first evaluated by FastQC. Forward and reverse reads were joined by join_paired_ends.py script. Barcodes were extracted by extract_barcodes.py script. Merged reads were checked to remove low quality sequences and demultiplexed by split_libraries_fastq.py script (using–max_bad_run_length 3–min_per_read_length_fraction 0.75–max_barcode_errors 0-q 19). Primers were removed by cutadapt 1.9.1. After dereplication and removal of singletons, high-quality reads were clustered into OTUs by the cluster_otus function of USEARCH at 97% identity (Edgar, 2013). Chimeric sequences were removed by the UCHIME algorithm, and the "gold" database (http://drive5.com/uchime/gold.fa) was used as a reference (Edgar et al., 2011). Another filter step was performed to remove non-bacterial 16S sequences by aligning all OTU representative sequences to the Greengenes database using PyNAST (align_seqs.py script). Unaligned 16S sequences were discarded at a threshold of 75% identity using the filter_fasta.py script (DeSantis et al., 2006; Caporaso et al., 2010a). Based on the high confidence 16S representative sequences, an OTU table was generated by USEARCH (-usearch_global and uc2otutab.py scripts). The taxonomy of the representative sequences was classified with the RDP classifier (Wang et al., 2007).

Phylogenetic tree construction: OTU representative sequences were aligned by Clustal Omega, and conserved regions were filtered with the filter_alignment.py script. A phylogenetic tree was constructed with the make_phylogeny.py script with the FastTree algorithm. Alpha diversity (Shannon diversity index, number of observed OTUs, and chao1 index) were calculated after resampling each sample to 50,000 reads (QIIME script single_rarefaction.py and alpha_diversity.py). Beta diversity (Bray-Curtis and weighted and unweighted UniFrac) were calculated for a cumulative sum scaling (CSS) transformed OTU table (QIIME script normalize_table.py and beta_diversity.py).

The principal coordinate analysis (PCoA) was performed using classical multidimensional scaling of beta diversity distance matrices with the cmdscale function in R. Con-

strained PCoA was implemented using the capscale function in the vegan package, and plant genotypes were used as a constrained factor. The genotype effect on root microbiota composition was determined by a permutation-based ANOVA test implemented with the anova.cca function, using 5,000 permutations.

Analysis of the differential OTU abundance and taxa were performed using a negative binomial generalized linear model in the edgeR package (Robinson et al., 2010). We first obtained normalization factors with the calcNormFactors function and then used the estimateGLMCommonDisp and estimateGLMTagwiseDisp functions to estimate, common and tag-wise dispersions for a Negative Binomial Generalized Linear Model (GLM). We fitted a negative binomial generalized log-linear model with OTU read counts by the glmFit function to test differential OTU abundance; corresponding *P*-values were corrected for multiple tests using a false discovery rate (FDR) set at 0.05 (Benjamini and Hochberg, 1995).

Scatter plots, boxplots, and pie charts were generated using the ggplot2 package. Venn diagrams were visualized with the Venndiagram package (Chen and Boutros, 2011). Hierarchical clustering of samples was based on Euclidean distance, and was visualized with the ggdendro package.

Accession Numbers

Bacterial 16S rRNA sequencing data were uploaded to the NCBI SRA database, with accession PRJNA400677. Scripts were deposited in https://github.com/microbiota/Chen2019SCLS. All primers used in this study are listed in Table S2 in Supporting Information.

Compliance and ethics *The author(s) declare that they have no conflict of interest.*

Acknowledgements *This work was supported by the Priority Research Program of the Chinese Academy of Sciences (ZDRW-ZS-2019-2, QYZDB-SSW-SMC021), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA08000000, XDB11020700), the National Program on Key Basic Research Projects (2013CB127000), and the State Key Laboratory of Plant Genomics of China (2016A0219-11, SKLPG2013A0125-5). We thank Dr. Jay D Keasling (University of California, Berkeley) for providing the pMBIS plasmid.*

References

Aubourg, S., Lecharny, A., and Bohlmann, J. (2002). Genomic analysis of the terpenoid synthase (*AtTPS*) gene family of *Arabidopsis thaliana*. Mol Genets Genom 267, 730–745.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Statist Soc-Ser B 57, 289–300.

Berendsen, R.L., Pieterse, C.M.J., and Bakker, P.A.H.M. (2012). The rhizosphere microbiome and plant health. Trends Plant Sci 17, 478–486.

Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F., Rauf, P., Huettel, B., Reinhardt, R., Schmelzer, E., et al. (2012). Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. Nature 488, 91–95.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G. L., and Knight, R. (2010a). PyNAST: A flexible tool for aligning sequences to a template alignment. Bioinformatics 26, 266–267.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010b). QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7, 335–336.

Castrillo, G., Teixeira, P.J.P.L., Paredes, S.H., Law, T.F., de Lorenzo, L., Feltcher, M.E., Finkel, O.M., Breakfield, N.W., Mieczkowski, P., Jones, C.D., et al. (2017). Root microbiota drive direct integration of phosphate stress and immunity. Nature 543, 513–518.

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

Chen, H., and Boutros, P.C. (2011). VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinf 12, 35.

Chen, Q., Fan, D., and Wang, G. (2015). Heteromeric geranyl (geranyl) diphosphate synthase is involved in monoterpene biosynthesis in *Arabidopsis* flowers. Mol Plant 8, 1434–1437.

Christianson, D.W. (2017). Structural and chemical biology of terpenoid cyclases. Chem Rev 117, 11570–11648.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72, 5069–5072.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461.

Edgar, R.C. (2013). UPARSE: Highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10, 996–998.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27, 2194–2200.

Foster, K.R., Schluter, J., Coyte, K.Z., and Rakoff-Nahoum, S. (2017). The evolution of the host microbiome as an ecosystem on a leash. Nature 548, 43–51.

Gan, X., Hay, A., Kwantes, M., Haberer, G., Hallab, A., Ioio, R.D., Hofhuis, H., Pieper, B., Cartolano, M., Neumann, U., et al. (2016). The Cardamine hirsuta genome offers insight into the evolution of morphological diversity. Nat Plants 2, 16167.

Hartmann, T. (2007). From waste products to ecochemicals: Fifty years research of plant secondary metabolism. Phytochemistry 68, 2831–2846.

Huang, A.C., Kautsar, S.A., Hong, Y.J., Medema, M.H., Bond, A.D., Tantillo, D.J., and Osbourn, A. (2017). Unearthing a sesterterpene biosynthetic repertoire in the Brassicaceae through genome mining reveals convergent evolution. Proc Natl Acad Sci USA 114, E6005–E6014.

Kampranis, S.C., Ioannidis, D., Purvis, A., Mahrez, W., Ninga, E., Katerelos, N.A., Anssour, S., Dunwell, J.M., Degenhardt, J., Makris, A. M., et al. (2007). 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.

Katoh, K., Misawa, K., and Kuma, K.I. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucl Acids Res 30, 3059–3066.

Kliebenstein, D.J., and Osbourn, A. (2012). Making new molecules—Evolution of pathways for novel metabolites in plants. Curr Opin Plant Biol 15, 415–423.

Leach, J.E., Triplett, L.R., Argueso, C.T., and Trivedi, P. (2017). Communication in the phytobiome. Cell 169, 587–596.

Lebeis, S.L., Paredes, S.H., Lundberg, D.S., Breakfield, N., Gehring, J.,

McDonald, M., Malfatti, S., Glavina del Rio, T., Jones, C.D., Tringe, S. G., et al. (2015). Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. Science 349, 860–864.

Li, W., Zhang, F., Chang, Y., Zhao, T., Schranz, M.E., and Wang, G. (2015). Nicotinate *O*-glucosylation is an evolutionarily metabolic trait important for seed germination under stress conditions in *Arabidopsis thaliana*. Plant Cell 27, 1907–1924.

Lundberg, D.S., Lebeis, S.L., Paredes, S.H., Yourstone, S., Gehring, J., Malfatti, S., Tremblay, J., Engelbrektson, A., Kunin, V., Del Rio, T.G., et al. (2012). Defining the core *Arabidopsis thaliana* root microbiome. Nature 488, 86–90.

Müller, D.B., Vogel, C., Bai, Y., and Vorholt, J.A. (2016). The plant microbiota: Systems-level insights and perspectives. Annu Rev Genet 50, 211–234.

Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S.L. (2012). Detecting individual sites subject to episodic diversifying selection. PLoS Genet 8, e1002764.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: A bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Schlaeppi, K., Dombrowski, N., Oter, R.G., Ver Loren van Themaat, E., and Schulze-Lefert, P. (2014). Quantitative divergence of the bacterial root microbiota in *Arabidopsis thaliana* relatives. Proc Natl Acad Sci USA 111, 585–592.

Shao, J., Chen, Q.W., Lv, H.J., He, J., Liu, Z.F., Lu, Y.N., Liu, H.L., Wang, G.D., and Wang, Y. (2017). (+)-Thalianatriene and (−)-retigeranin B catalyzed by sesterterpene synthases from *Arabidopsis thaliana*. Org Lett 19, 1816–1819.

Srividya, N., Davis, E.M., Croteau, R.B., and Markus Lange, B. (2015). Functional analysis of (4*S*)-limonene synthase mutants reveals determinants of catalytic outcome in a model monoterpene synthase. Proc Natl Acad Sci USA 112, 3332–3337.

Starks, C.M., Back, K., and Chappell, J. (1997). Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. Science 277, 1815–1820.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis Version 6.0. Mol Biol Evol 30, 2725–2729.

Tholl, D., and Lee, S. (2011). Terpene specialized metabolism in *Arabidopsis thaliana*. Arabidopsis Book 9, e0143.

van Dam, N.M., and Bouwmeester, H.J. (2016). Metabolomics in the rhizosphere: Tapping into belowground chemical communication. Trends Plant Sci 21, 256–265.

Verbon, E.H., and Liberman, L.M. (2016). Beneficial microbes affect endogenous mechanisms controlling root development. Trends Plant Sci 21, 218–229.

Vickers, C.E., Bongers, M., Liu, Q., Delatte, T., and Bouwmeester, H. (2014). Metabolic engineering of volatile isoprenoids in plants and microbes. Plant Cell Environ 37, 1753–1775.

Wang, C., Chen, Q., Fan, D., Li, J., Wang, G., and Zhang, P. (2016). Structural analyses of short-chain prenyltransferases identify an evolutionarily conserved GFPPS clade in Brassicaceae plants. Mol Plant 9, 195–204.

Wang, G., Tian, L., Aziz, N., Broun, P., Dai, X., He, J., King, A., Zhao, P. X., and Dixon, R.A. (2008). Terpene biosynthesis in glandular trichomes of hop. Plant Physiol 148, 1254–1266.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ MicroBiol 73, 5261–5267.

Xu, H., Zhang, F., Liu, B., Huhman, D.V., Sumner, L.W., Dixon, R.A., and Wang, G. (2013). Characterization of the formation of branched short-chain fatty acid: CoAs for bitter acid biosynthesis in hop glandular trichomes. Mol Plant 6, 1301–1317.

Zhang, J., Zhang, N., Liu, Y.X., Zhang, X., Hu, B., Qin, Y., Xu, H., Wang, H., Guo, X., Qian, J., et al. (2018). Root microbiota shift in rice correlates with resident time in the field and developmental stage. Sci China Life Sci 61, 613–621.

Zhao, T., Holmer, R., de Bruijn, S., Angenent, G.C., van den Burg, H.A., and Schranz, M.E. (2017). Phylogenomic synteny network analysis of MADS-box transcription factor genes reveals lineage-specific transpositions, ancient tandem duplications, and deep positional conservation. Plant Cell 29, 1278–1292.

SUPPORTING INFORMATION

Figure S1 Functional characterization of 16 plant sesterTPSs in *E.coli* system.

Figure S2 Terpene synthase assays of TPS18 and TPS18^{G328W} analyzed by GC-MS.

Figure S3 Terpene synthase assays of TPS19 and TPS19^{G325W} analyzed by GC-MS.

Figure S4 Terpene synthase assays of TPS25 and TPS25^{G328W} analyzed by GC-MS.

Figure S5 Terpene synthase assays of TPS30 and TPS30^{P328W} analyzed by GC-MS.

Figure S6 Characterization of *TPS25* T-DNA insertion mutants.

Figure S7 Characterization of *TPS30* T-DNA insertion mutant (*tps30-1*|CS805958).

Figure S8 *tps30-2* and *tps30/tps25-1* double mutants generated via CRISPR/Cas9.

Figure S9 *tps30-3* and *tps30/tps25-2* double mutants generated via CRISPR/Cas9.

Figure S10 Sesterterpene profiling in various tissues of *Arabidopsis* and different transgenic plants by GC-QQQ-MS.

Figure S11 Phenotypes of sesterTPS mutants and Col-0 when grown in natural soil (Changping farm in Beijing) under climate control conditions.

Figure S12 Sample diversity (α/β-diversity) measurements among each genotype.

Figure S13 Stack plot showing relative abundance distribution of the OTUs in phylum.

Figure S14 Taxonomic difference and overlap of specific differential OTUs between *TPS* mutants and Col-0.

Table S1 The MRM settings for four sesterterpene (C25) backbones analyzed by GC-QQQ-MS in this study

Table S2 The primers used in this study

The supporting information is available online at http://life.scichina.com and https://link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.