Terpenes synthesize genes in eukaryotes beyond plants and fungi: Occurrence in social amoeba

Xinlu Chen, Tobias G. Kölner, Qidong Jia, Ayla Norris, Balaji Santhanam, Patrick Rabe, Jeroen S. Dickschat, Gad Shaulsky, Jonathan Gershenson, and Feng Chen

*Department of Plant Sciences, University of Tennessee, Knoxville, TN 37996; †Department of Biochemistry, Max Planck Institute for Chemical Ecology, D-07745 Jena, Germany; ‡Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN 37996; §Graduate Program in Structural Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, TX 77030; ¶Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; and ¶Kekulé-Institute of Organic Chemistry and Biochemistry, University of Bonn, 53121 Bonn, Germany

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved August 30, 2016 (received for review June 27, 2016)

Terpenes are structurally diverse natural products involved in many ecological interactions. The pivotal enzymes for terpene biosynthesis, terpene synthases (TPSs), had been described only in plants and fungi in the eukaryotic domain. In this report, we systematically analyzed the genome sequences of a broad range of nonplant/nonfungus eukaryotes and identified putative TPS genes in six species of amoeba, five of which are multicellular social amoeba belonging to the order Dictyosteliida. A phylogenetic analysis revealed that amoebal TPSs are evolutionarily more closely related to fungal TPSs than to bacterial TPSs. The social amoeba Dictyostelium discoideum was selected for functional study of the identified TPSs. D. discoideum grows as a unicellular organism when food is abundant and switches to vegetative growth to multicellular development upon starvation. We found that expression of most D. discoideum TPS genes was induced during development. Upon heterologous expression, all nine TPSs from D. discoideum showed sesquiterpene synthase activities. Some also exhibited monoterpane and/or diterpene synthase activities. Direct measurement of volatile terpenes in cultures of D. discoideum revealed essentially no emission at an early stage of development. In contrast, a bouquet of terpenes, dominated by sesquiterpenes including β-barbatene and (E,E)-α-farnesene, was detected at the middle and late stages of development, suggesting a development-specific function of volatile terpenes in D. discoideum. The patchy distribution of TPS genes in the eukaryotic domain and the evidence for TPS function in D. discoideum indicate that the TPS genes mediate lineage-specific adaptations.

Significance

Many living organisms use terpenes for ecological interactions. Terpenes are biosynthesized by terpene synthases (TPSs), but classic TPS genes are known to exist only in plants and fungi among the eukaryotes. In this study, TPS genes were identified in six species of amoeba with five of them being multicellular social amoeba. Amoebal TPSs showed closer relatedness to fungal TPSs than bacterial TPSs. In the social amoeba Dictyostelium discoideum, all nine TPS genes encoded active enzymes and most of their terpene products were released as volatiles in a development-specific manner. This study highlights a wider distribution of TPS genes in eukaryotes than previously thought and opens a door to studying the function and evolution of TPS genes and their products.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences for the biochemically characterized terpene synthases reported in this paper have been deposited in the GenBank database (accession nos. KX364374–KX364382).

1To whom correspondence should be addressed. Email: fengc@utk.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1610379113/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1610379113
Identified Eukaryotic Terpene Synthases: Evolutionary Relatedness and Motifs. To understand the evolutionary relatedness of the identified eukaryotic TPSs with known TPSs, a phylogenetic tree was constructed that includes, besides the eukaryotic TPSs described here, representative bacterial and fungal TPSs, and microbial type TPSs from the lycophyte Selaginella moellendorffii (8). Notably, the TPSs from the five species of Dictyostelidida clustered together (clade I), whereas the seven TPSs from N. gruberi clustered in a separate, but closely related clade (clade II) (Fig. 2). Together, the amoebal TPSs showed closer relatedness to fungal TPSs than to bacterial TPSs (Fig. 2).

TPSs contain several highly conserved motifs that are important for catalytic activity including the aspartate-rich “DDxx(x)D/E” motif and the “NDxxSxxxD/E” motif, both of which are involved in complexing metal ions to coordinate the binding of the isoprenyl diphosphate substrate in the active site (20, 21). Both motifs are also highly conserved among all newly identified eukaryotic TPSs (Table S3). In addition, the diphosphate sensor that is involved in substrate recognition and critical for catalytic activity (Arginine) (6, 22) was also highly conserved (Table S3).

Expression Patterns of Individual Terpene Synthase Genes in Dictyostelium discoideum. D. discoideum was selected as a model system to explore the function of the newly identified eukaryotic TPSs. As a social amoeba, D. discoideum has a distinctive life cycle (Fig. S1). It propagates vegetatively as a unicellular organism when food (bacteria in the natural environment) is abundant. Upon starvation, D. discoideum transitions into multicellular development in a highly coordinated process that causes individual cells to aggregate and differentiate with formation of a multicellular slug that migrates and finally turns into a fruiting body. This process lasts approximately 24 h (23).

The D. discoideum genome contains 11 putative TPS genes, 9 of which show a full-length sequence and were designated DdTPS1 to DdTPS9. Analysis of the published gene expression...
dataset of *D. discoideum* (24) revealed that each TPS gene of *D. discoideum* was expressed at different times during multicellular development (Fig. 3). Specifically, as *D. discoideum* began to starve (0 h), mRNAs of DdTPS1, DdTPS2, DdTPS3, DdTPS6, and DdTPS7 were present at detectable but low levels, whereas mRNAs of DdTPS4, DdTPS5, DdTPS8, and DdTPS9 were almost undetectable, but in all cases, the mRNA abundance increased during development. During mound formation (approximately 8–12 h), DdTPS3 attained the highest level of expression among all nine genes, whereas DdTPS1 levels were still near the limit of detection. At the time of slug formation (approximately 16 h), DdTPS2 and DdTPS8 mRNAs reached their highest abundance, whereas the abundance of DdTPS3 started to decrease. Finally, during culmination (from 18 to 24 h), DdTPS2, DdTPS3, DdTPS5, and DdTPS8 levels decreased, whereas DdTPS1, DdTPS4, DdTPS6, and DdTPS9 mRNAs accumulated to higher levels, reaching their peaks at the ultimate stage of mature fruiting body (24 h).

**DdTPS Genes in Dictyostelium discoideum Encode Active Terpene-Producing Enzymes.** To further understand the function of *D. discoideum* terpene synthase genes, we characterized the biochemical activities of the enzymes they encode. Full-length cDNAs of DdTPS1–DdTPS9 were cloned and heterologously expressed in *Escherichia coli*. Individual recombinant DdTPS proteins were tested for terpene synthase activity. All nine enzymes were able to accept farnesyldiphosphate (FPP) as a substrate to produce either a single sesquiterpene or a mixture of compounds (Fig. 4). The major products of DdTPS1, DdTPS4, DdTPS5, and DdTPS7/9 were identified as (E,E)-α-farnesene, (E)-nerolidol, (E)-β-farnesene, and β-barbatene, respectively, whereas DdTPS2, DdTPS3, DdTPS6, and DdTPS8 produced unidentified sesquiterpenes. DdTPS1, DdTPS2, DdTPS3, and DdTPS9 were also able to convert geranyldiphosphate (GPP) into different mixtures of monoterpenes (Fig. S2). In addition, diterpene products from geranylgeranyldiphosphate (GGPP) could be observed for DdTPS1, DdTPS2, DdTPS3, DdTPS4, DdTPS5, and DdTPS9 (Fig. S3).

Dictyostelium discoideum Emits Terpene-Dominated Volatiles During Multicellular Development. The elaborate temporal regulation of mRNA abundance during different life stages suggests that the DdTPS genes may play a role in development. Monoterpenes and sesquiterpenes are generally volatile compounds that can be trapped by use of, e.g., a closed-loop stripping apparatus or solid-phase microextraction (SPME) and analyzed by gas chromatography/mass spectrometry (GC/MS) (25). Based on the expression patterns of individual DdTPS genes (Fig. 3) and the in vitro biochemical activities of the respective proteins (Fig. 4 and Figs. S2 and S3), we hypothesized that DdTPSs might be involved in producing volatile compounds during development. To test this hypothesis, we performed volatile profiling of *D. discoideum* cultures at 4-h intervals during the 24 h of development (Fig. 5).

Altogether, a total of 15 volatile compounds were detected from developing *D. discoideum* cultures (Fig. 5), including 11 terpenes and four nonterpene volatiles, one of which was identified as 2-phenylethanol. The terpene portion of the volatiles was dominated by nine detectable sesquiterpenes, of which four were identified as (E,E)-α-farnesene, calarenene, (E)-nerolidol, and β-barbatene (Fig. 5). Comparison of the mass spectra of the unidentified sesquiterpenes in the headspace extracts to those obtained enzymatically with the expressed DdTPSs allowed assignment of each of the compounds emitted by *D. discoideum* with confidence to a specific DdTPS. As such, each sesquiterpene in Fig. 5 was labeled with the same peak number as used in Fig. 4. In addition, the monoterpene linalool and one diterpene, which was identical to in vitro diterpene product of DdTPS5 (Fig. S3), were detected in the culture extracts.

The relative abundance of individual volatile terpenes during the 24 h of development was calculated based on three biological replicates (Fig. 5 and Fig. S4). At the beginning of development (0 h), essentially no volatile terpenes were detected, whereas after 4 h of development, the emission of traces of terpenes including (E)-nerolidol was detected. The production of terpenes by *D. discoideum* gradually increased during the next hours of development, but some compounds showed an early maximum production, e.g., calarenene peaked at 12 h and 16 h, whereas the production of other terpenes such as (E,E)-α-farnesene and β-barbatene exhibited a later maximum of production.

**Discussion**

This first report of the occurrence of canonical terpene synthase genes in the social amoebae raises questions about the functions of the terpene products in these organisms. The fact that TPS gene expression and terpene volatile emission in *D. discoideum* are restricted to specific periods during multicellular development suggests possible roles for these compounds if the unique biology of *D. discoideum* is considered in light of the known functions of volatile terpenes in other organisms.

One possible function of volatile terpenes emitted from *D. discoideum* is to attract other organisms to facilitate spore dispersal, resembling the function of volatile terpenes from the fruiting bodies of fungi (26). Forming fruiting bodies by social amoebae is considered to be an adaptation for spore dispersal (27). This hypothesis was directly supported by experimental studies in which fruiting bodies were shown to increase the rate at which spores are acquired by a model invertebrate *Drosophila*
Sesquiterpene synthase activity of D. discoideum terpene synthases. Genes were heterologously expressed in E. coli, and crude protein extracts were incubated with the substrate FPP. Enzyme products were collected by using solid-phase microextraction and analyzed by GC/MS. GC traces (Left) and mass spectra of major products (Right) are shown. 1, (E,E)-α-farnesene*; 2, unidentified sesquiterpene hydrocarbon; 3, β-maaliene; 4, aristolene; 5, calarene; 6, unidentified sesquiterpene hydrocarbon; 7, unidentified sesquiterpene hydrocarbon; 8, unidentified sesquiterpene hydrocarbon; 9, (E)-nerolidol*; 10, β-elemene*; 11, (E)-β-farnesene*; 12, unidentified sesquiterpene hydrocarbon; 13, β-barbatene*; 14, unidentified sesquiterpene; cont, contamination. Compounds marked with asterisks (*) were identified by using authentic standards. Each assay was repeated at least three times, and a representative GC chromatogram is shown.

Although the primary vectors for D. discoideum spore dispersal are unknown (27), it will be an interesting future subject to identify such vectors in nature and then determine whether volatile terpenes have a role in attracting such vectors to facilitate spore dispersal. Consistent with this hypothesis, β-barbatene (Fig. 5) emitted from the fruiting bodies of the bracket fungus Fomitopsis pinicola has been implicated in attracting insects for spore dispersal (26).

Another possible function of D. discoideum volatile terpenes is defense. Social amoebae are preyed on by nematodes (28). They have evolved multiple defense mechanisms, which include the synthesis of a protective extracellular matrix called the slime sheath and the formation of protective coats at the surface of spores (28). In addition, individual amoebae protect themselves by secreting compounds that repel nematodes (28). It will be interesting to see whether any of the volatile terpenes emitted from D. discoideum serve such a function as well. Consistent with this hypothesis, (E,E)-α-farnesene (Fig. 5) emitted from the leaves of the model plant Arabidopsis thaliana has been implicated in defense against insects (29).

The third possibility is that D. discoideum terpenes may function as signals to coordinate multicellular development. The roles of terpenes in signaling have been relatively well-studied in plants (30). As volatile compounds, terpenes can signal over a distance in either multicellular organisms or multicellular aggregates. Previous studies showed that ammonia, a volatile by-product of gluconeogenesis, is involved in regulating several stages of D. discoideum development, including aggregation (31), slug migration (32), and culmination (33). It is certainly intriguing to ask whether volatile terpenes could have similar functions, with the diversity of chemical structures helping to provide functional specificity.

It is interesting that all of the five species of social amoebae (i.e., all from Dictyostelida) with sequenced genomes contain TPS genes, whereas the three species from the genus Entamoeba (Table S1 and Fig. 1), which also belong to the supergroup Amoebozoa, do not contain any TPS gene. The genus Entamoeba...
are amoeobae but not social amoeobae, suggesting that TPS genes may provide adaptive functions for social amoeoba, which share a unique lifestyle. However, N. gruberi of the supergroup Excavata, which is also an amoeoba having no multicellular development (34), contains TPS genes (Fig. 1). Sharing similarity in lifestyle, N. gruberi and Entamoeba are evolutionarily distantly related. It will be interesting to investigate the biochemical and biological function of N. gruberi TPS genes and to understand how they may confer a fitness advantage.

To briefly summarize, we have found classic terpene synthase genes in six species of amoeoba among a broad range of non-plant/nonfungus eukaryotes. Amoebal TPSs are more closely related to fungal TPSs and the microbial type TPSs from plants than bacterial TPSs (Fig. 2). The social amoeoba D. discoideum is an organism other than plants, fungi, and bacteria from which classic terpene synthase genes have been functionally characterized (Fig. 4). This study provides insights into the occurrence, function and evolution of TPS genes, particularly in eukaryotes, and it is expected to stimulate important future research.

Materials and Methods

Sequence Retrieval and Analysis. A total of 168 species of nonplant/nonfungus eukaryotes with well-annotated genome sequences (Table S1) archived at the KEGG genome database (www.genome.jp/kegg/catalog/org_list.html) were downloaded as the genome dataset. Another dataset was the non-redundant (nr) protein database from NCBI, which was downloaded on April 19, 2016. Both databases were searched against the Pfam-A database locally by using HMMER 3.0 with an e-value of 1e-7. Sequences with best hits from the following three HMM profiles were identified as putative terpene synthases: Terpene.synth.C (PF03936) and Terpene Synthase N-terminal domain (PF01397), and TRIS (PF06330). For the search of the nr database, the terpene synthase hits identified from plants, fungi, archaea, and bacteria were removed. For phylogenetic reconstruction, known bacterial and fungal terpene synthases were retrieved from Pfam database (version 27). MAFFT (L-INS-i) was used to build the multiple sequence alignment with 1,000 iterations of improvement. The maximum-likelihood phylogenetic tree was built with RAxML through the CIPRES Science Gateway (https://www.phylo.org) by using the LG+G+F amino acid substitution model with 1,000 bootstrap replicates and then rendered by using FigTree (version 1.4.2).

Cloning of Full-Length cDNA of DdTPS Genes of Dictyostelium discoideum via RT-PCR. D. discoideum (strain AX4) was obtained from the Dictybase Stock Center (www.dictybase.org). D. discoideum was cocultured with live Klebsiella pneumoniae bacteria on SM agar plates (35) by following the protocol described in the Dictybase Stock Center. When slugs formed, D. discoideum cells were collected and used for total RNA isolation following the protocol described in ref. 36. Full-length cDNA of each DdTPS gene was amplified by RT-PCR using the gene specific primers listed in Table S4. PCR products were cloned into the pEXP-S-CT/TOPO vector (ThermoFisher Scientific) and confirmed by sequencing.

Terpene Synthase Enzyme Assays. Heterologous expression of DdTPS genes in E. coli, recombinant protein preparation, terpene synthase enzyme assays, and terpene product identification using a Hewlett-Packard 6890 gas chromatograph coupled to a Hewlett-Packard 5973 mass spectrometer were performed as described (8). Each expressed protein was assayed at least three times.

DdTPS Gene Expression Analysis. DdTPS gene expression was analyzed by using published RNA-seq data from developing D. discoideum (24). Transcript abundance was quantified as described in ref. 37. The data are available on
D. discoideum was grown on a thick lawn of K. aerogenes before the food source was removed to initiate the developmental program. The samples were obtained in 4-h intervals throughout the 24-h developmental program.

**Headspace Collection and GC/MS Analysis.** A mixture of D. discoideum spores and freshly grown K. pneumoniae was spread onto SM agar plates to initiate D. discoideum culture. Under our experimental conditions, D. discoideum progressed from spore germination to vegetative growth to the completion of multicellular development in ≈48 h. At 24 h, the appearance of the culture plate changed from opaque (from the bacterial lawn) to translucent, indicating the clearing of bacteria. This time point was defined as the start of multicellular development, after which D. discoideum progressed through the various described developmental stages in the next 24 h with the eventual formation of fruiting bodies (Fig. 5). SPME combined with GC/MS was used for volatile profiling of the D. discoideum cultures. During the 24 h of development, volatiles were collected once every 4 h (Fig. S1). Before each collection, the lid of the culture plate was removed and the plate was left in the hood for 1 min to dispose of accumulated volatiles. Then the lid was put back on and a SPME fiber coated with 100-μm polydimethylsiloxane was inserted into the headspace of the plate to start volatile collection. After 1 h, the SPME fiber was retracted and inserted into the injector port (a splitless injection and injector temperature of 250 °C) of a Shimadzu QP5050A quadrupole mass selective detector for chemical identification. Separation was performed on a Restek Rxi-5Sil MS column (20 m × 0.25 mm i.d. × 0.25 μm thickness; Restek) with helium as the carrier gas and a temperature program from 60 °C to 300 °C at 5 °C min⁻¹. The experiment was performed with three biological replicates.

**Acknowledgments.** We thank Wangdan Xiong for her assistance with the preparation of Fig. 3 and Fig. 5 and Natascha Rauch (Max Planck Institute for Chemical Ecology) for her technical assistance. This project was supported by a University of Tennessee AgResearch Innovation Grant (to F.C.) and the Max Planck Society (T.G.K. and J.G.). G.S. and B.S. acknowledge support from National Institutes of Health Grant PO1 HD039691.