PNAS RESEARCH ARTICLE | BIOCHEMISTRY

# Canonical terpene synthases in arthropods: Intraphylum gene transfer

Xinlu Chen[a], John M. Urban[b,c], Jens Wurlitzer[d], Xiuting Wei[e], Jin Han[a], Sarah E. O'Connor[d], Jeffrey D. Rudolf[e], Tobias G. Köllner[d], and Feng Chen[a,1]

Affiliations are included on p. 10.

Edited by Nancy Moran, The University of Texas at Austin, Austin, TX; received July 3, 2024; accepted November 11, 2024

**Insects employ terpenoids for communication both within and between species. While terpene synthases derived from isoprenyl diphosphate synthase have been shown to catalyze terpenoid biosynthesis in some insects, canonical terpene synthases (TPS) commonly found in plants, fungi, and bacteria were previously unidentified in insects. This study reveals the presence of *TPS* genes in insects, likely originating via horizontal gene transfer from noninsect arthropods. By examining 361 insect genomes, we identified *TPS* genes in five species of the Sciaridae family (fungus gnats). Additionally, *TPS* genes were found in Collembola (springtails) and Acariformes (mites) among diverse noninsect arthropods. Selected TPS enzymes from Sciaridae, Collembola, and Acariformes display monoterpene, sesquiterpene, and/or diterpene synthase activities. Through comprehensive protein database search and phylogenetic analysis, the *TPS* genes in Sciaridae were found to be most closely related to those in Acariformes, suggesting transfer of *TPS* genes from Acariformes to Sciaridae. In the model Sciaridae *Bradysia coprophila*, all five *TPS* genes are most highly expressed in adult males, suggesting a sex- and developmental stage-specific role of their terpenoid products. The finding of *TPS* genes in insects and their possible evolutionary origin through intraphylum gene transfer within arthropods sheds light on metabolic innovation in insects.**

terpene synthase | evolution | horizontal gene transfer | arthropods

## Significance

Insects, the most speciose group of animals, rely on infochemicals for communication. Terpenoids are a diverse class of insect infochemicals. Yet our understanding of terpenoid biosynthesis in insects is very limited. Here, we report that insects in the Sciaridae family employ canonical terpene synthase (*TPS*) genes, like plants, fungi, and bacteria, for terpenoid biosynthesis. Sciaridae *TPSs* are most closely related to *TPSs* from mites, implying a possibility of horizontal gene transfer from mites to the common ancestor of Sciaridae. In a model insect *Bradysia coprophila*, *TPS* genes were found to encode functional enzymes and have the highest levels of expression in adult males. Our findings shed light on the evolution of terpenoid biosynthesis in insects.

Insects are the largest clade of animals and contain more described species than any other taxonomic class (1). Different from other animals, insects produce a diverse array of secondary metabolites (2), with terpenoids being one prominent chemical class (3, 4). Terpenoids play vital roles in intra- and interspecies communications in insects, such as serving as sex pheromones (5), aggregation pheromones (6), alarm pheromones (7), and trail pheromones (8). Therefore, understanding terpene biosynthesis may be an important avenue to understanding rapid radiations and adaptation in insects.

As a diverse chemical class, terpenoids are not just rich in insects. Plants are better-known producers of terpenoids (9). In addition, many species of fungi (10) and bacteria (11, 12) produce terpenoids. Despite the vast chemical diversity of terpenoids in different lineages of life, their universal building blocks isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) are produced by only two pathways: the mevalonate pathway and the methylerythritol phosphate pathway. From IPP and DMAPP, isoprenyl diphosphates of various chain length, including geranyl diphosphate (GPP, C10), farnesyl diphosphate (FPP, C15), and geranylgeranyl diphosphate (GGPP, C20) are formed by the action of isoprenyl diphosphate synthase (IDS). GPP, FPP, and GGPP are then converted to monoterpenes, sesquiterpenes, and diterpenes, respectively, by the action of terpene synthases. Canonical terpene synthases (TPSs) have been well studied in plants (13–15), fungi (16), and bacteria (11). Having a modular structure, TPSs consist of one or multiple domains known as α, β, and γ. TPSs that contain only α domain occur widely in bacteria (11, 12), fungi (16), and nonseed plants (14). In contrast, αβγ-tridomain TPSs occur widely in plants (13), sporadically in fungi (17), and rarely in bacteria (18). Plants are unique in that they also contain αβ-didomain TPSs (15).

Interestingly, insects appear to employ a different enzyme system for terpene biosynthesis. Whereas no canonical TPS has been identified in insects, IDS-like enzymes have been demonstrated to catalyze terpene biosynthesis in multiple insect species. The bark beetle *Ips pini* contains an IDS-like enzyme that produces the monoterpene myrcene from GPP (19, 20). Similarly, two IDS-like enzymes involved in the production of the monoterpene (*E*)-β-ocimene, an antiaphrodisiac pheromone, were recently discovered in *Heliconius postman* butterflies (21). Furthermore, flea beetle *Phyllotreta striolata* uses an IDS-like enzyme to synthesize the sesquiterpene aggregation pheromone (6*R*,7*S*)-himachala-9,11-diene (22), and the harlequin bug *Murgantia histrionica*

Author contributions: J.D.R., T.G.K., and F.C. designed research; X.C., J.M.U., J.W., X.W., and J.H. performed research; X.C., J.M.U., J.W., X.W., J.H., S.E.O., J.D.R., T.G.K., and F.C. analyzed data; and J.M.U., J.D.R., T.G.K., and F.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.



[1]To whom correspondence may be addressed. Email: fengc@utk.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2413007121/-/DCSupplemental.

Published December 13, 2024.

encodes an IDS-like terpene synthase that synthesizes sesquipiperitol as precursor for an aggregation pheromone (23). The discovery of multiple IDS-like terpene synthases in several insect orders suggests that this type of terpene synthases may occur widely in insects (3). It is noteworthy to mention that IDS-like terpene synthases have also been identified in fungi (24), suggesting that the evolution of terpene synthases from IDSs has occurred independently multiple times.

Considering the vast chemical diversity of terpenoids in insects, it is tempting to ask whether IDS-like terpene synthases are the only enzymes that are responsible for terpene biosynthesis in this class of animals. The recent, somewhat surprising discovery of canonical *TPS* genes in social amoebae (25), red algae (26), soft corals (27, 28), even viruses (29), indicates that they are more widely distributed than previously thought. Along this line, we asked whether canonical *TPS* genes occur in insects and contribute to terpenoid diversity. It is worth noting that a recent study detected canonical *TPS* genes in two species of mites: *Dinothrombium tinctorium* and *Leptotrombidium deliense*, which are noninsect arthropods (30), although in-depth investigation of these genes is still needed. The primary goal of this study was to systematically analyze the sequenced genomes of insects to determine whether canonical *TPS* genes occur in insects, and if so, to determine their relatedness to *TPSs* of other organisms in order to understand their evolution.

## Results

**Identification of *TPS* Genes in the Sequenced Genome of Insects of the Sciaridae Family.** To determine whether *TPS* genes occur in insects, an HMMER search was performed on the sequenced genomes of 361 species of insects (*SI Appendix*, Table S1) belonging to 14 orders. While *TPS* genes were absent in 13 of the 14 orders, they were detected in five species in the order of Diptera (Fig. 1*A*). The 112 species in the Diptera analyzed belong to two suborders: Brachycera (73 species) and Nematocera (39 species). All five *TPS*-containing species belong to the Nematocera suborder. The 39 species in the Nematocera suborder belong to three superfamilies: Culicomorpha (30 species), Psychodomorpha (two species), and Bibionomorpha (seven species). All five *TPS*-containing species belong to the Bibionomorpha superfamily. The seven species in the Bibionomorpha superfamily can be further divided into two families: Cecidomyiidae (two species) and Sciaridae (five species). All five *TPS*-containing species are from the Sciaridae family. These include *Bradysia coprophila* (five *TPS* genes, designated *BcTPS1-5*), *Bradysia odoriphaga* (three *TPS* gene, designated *BoTPS1-3*), *Pseudolycoriella hygida* (two *TPS* genes, designated *PhTPS1-2*), *Phytosciara flavipes* (26 *TPS* genes, designated *PfTPS1-26*), and *Trichosia splendens* (one *TPS* gene, designated *TsTPS*).

**Phylogenetic Analysis of the Sciaridae *TPS* Genes.** A phylogenetic analysis of the Sciaridae TPSs was performed using red algae EaMTPSL1 from *Erythrolobus australicus* (26) as an outgroup. The Sciaridae TPSs formed two clades S1 and S2: The S1 clade contains TPSs from all five species and the S2 clade contains TPSs from three species *B. coprophila*, *B. odoriphaga,* and *P. flavipe* (Fig. 1*B*). In clade S1, there are three groups: Group 1 contains TPS from one species; group 2 contains TPSs from *B. coprophila*, *B. odoriphaga,* and *P. flavipe*, and group 3 is specific to *P. flavipe*. These suggest two possible scenarios regarding the evolution of *TPS* genes in Sciaridae. In one scenario, ancestor Sciaridae contained two *TPS* genes: one leading to clade S1 and the other leading to clade S2, the latter of which was subsequently lost in

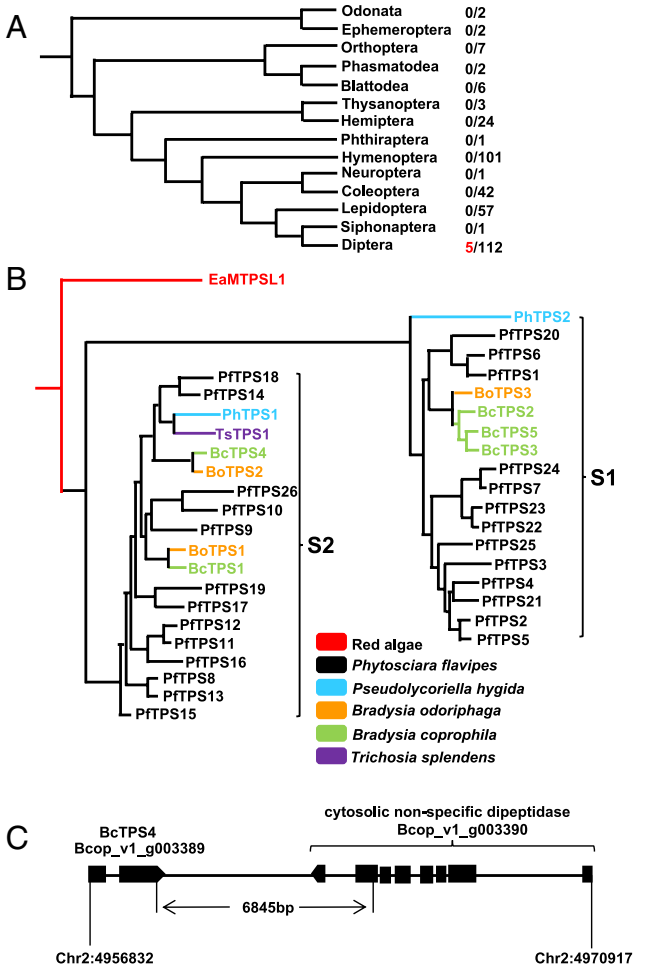


**Fig. 1.** Occurrence and analysis of canonical terpene synthases in insects. (*A*) distribution of canonical *TPS* genes among insects. A total of 361 species (*SI Appendix*, Table S1) were analyzed. These insects' genomes were arranged according to their known phylogeny (31). The first number (before the slash) indicates the number of species in certain lineages that were determined to contain *TPS* genes. The second number (after the slash) indicates the total number of species in that lineage that were analyzed. Within the Diptera order, the presence and absence of *TPS* genes among infraorder are shown. Within the Bibionomorpha infraorder, the presence and absence of *TPS* genes among families are shown. (*B*) Phylogenetic tree of TPSs from five species of Sciaridae including *B. coprophila*, *B. odoriphaga*, *P. flavipes*, *P. hygida,* and *T. splendens*. EaMTPSL1, a red algae TPS, was used as an outgroup. S1 and S2 are two clades formed by the TPSs from the five species of Sciaridae family. (*C*) evidence that selected *BcTPS* genes are nuclear genes of the *B. coprophila* genome. Schematic representation of bcTPS4 and its neighbor gene in the genome of *B. coprophila*. A genomic DNA fragment of 6,046 bp (delimited by arrows), which included BcTPS4, a part of its neighboring gene encoding a cytosolic nonspecific dipeptidase and the intergenic region, was amplified by PCR and confirmed by sequencing.

*T. splendens* and *P. hygida*. In the second scenario, the ancestor to Sciaridae contained only one *TPS* gene leading to clade S1, but the common ancestor of *B. coprophila*, *B. odoriphaga,* and *P. flavipe* would need to acquire a new *TPS* gene. In either case, the gene in clade S1 underwent one duplication in the common ancestor of *B. coprophila*, *B. odoriphaga,* and *P. flavipe* after its divergence from other lineages.

**TPSs Identified in Sciaridae Are Not from Microbial Contamination.** The mere existence of *TPS* genes in Sciaridae among diverse taxa of insects (Fig. 1) raised the question of whether these genes actually originate from Sciaridae or instead from microbial contamination. To exclude the latter possibility, three analyses were performed. In the first analysis, we examined the

exon–intron organization of all Sciaridae *TPS* genes based on gene annotation. Except one gene that does not contain any introns, all other genes contain at least one intron (*SI Appendix*, Fig. S1). The existence of introns in all five *TPS* genes from *B. coprophila* was further supported by RNA-seq data, which contained many reads that were anchored into both exons of a given *BcTPS* gene with split alignments spanning across the introns of the annotated gene structures (*SI Appendix*, Fig. S2), from both sexes and all life stages (32).

In the second analysis, we performed genomic PCR using primers based on the sequence of a representative gene, *BcTPS4* (Bcop_v1_g003389) from *B. coprophila* and the gene immediately upstream of *BcTPS4* and that of the gene immediately downstream Bcop_v1_g003390. This gene encodes a cytosolic nonspecific dipeptidase (CNSD) that is most similar to CNSD proteins in other insects (*SI Appendix*, Fig. S3). When the corresponding PCR product was amplified and sequenced, the assembly containing *BcTPS4* and its immediate neighboring genes was found to be accurate (Fig. 1*C*).

In the third analysis, we examined a variety of genome-mapping datasets, including short reads (Illumina), long reads (PacBio and Nanopore), optical maps (BioNano Genomics), and Hi-C links (32–34). The coverage from all datasets across the loci with *BcTPS* genes was found to be consistent with the coverage across the chromosomes they resided on (*SI Appendix*, Fig. S4). Moreover, there are no breaks in the Hi-C signal at the *BcTPS* gene loci, which show Hi-C interaction frequencies that follow the expected chromosome-wide pattern of more interactions with nearby genes than faraway genes on the same chromosome (*SI Appendix*, Fig. S4). Finally, when zooming in on the *BcTPS* loci, we identified individual long reads and ultralong optical maps that contain *BcTPS* genes and nearby highly conserved Dipteran genes, providing demonstrations of both on the same long DNA molecules from distinct technologies, DNA preparations, and sources (*SI Appendix*, Fig. S4). All of our inspections of genome-wide datasets demonstrated that the *BcTPS* genes are part of the nuclear chromosomes in *B. coprophila* (*SI Appendix*, Fig. S4). The results from all three approaches rule out that the *TPS* genes identified in Sciaridae are due to microbial contamination.

**Sciaridae TPSs Are Most Related to TPSs from Three Species of Mites in the Oribatida Order.** To understand how Sciaridae TPSs are related to TPSs from other organisms, individual Sciaridae TPSs were used as queries for BlastP searches against the nonredundant protein database at NCBI (https://blast.ncbi.nlm.nih.gov) excluding Sciaridae. Among the five BcTPSs from *B. coprophila*, BcTPS1 and BcTPS3 had top hits from *Oppia nitens*, while the other three BcTPSs had top hits from *Medioppia subpectinata* (Table 1). For the three BoTPSs from *B. odoriphaga*, BoTPS1 and BoTPS2 had top hits from *O. nitens* and BoTPS3 had a top hit from *M. subpectinata* (Table 1). For the two *PhTPS* genes from *P. hygida*, *PhTPS1* had its top hit from *O. nitens* whereas the top hit for *PhTPS2* was from *Oppiella nova*. For the 26 *PfTPS* genes from *P. flavipes*, three had top hits from *M. subpectinata* and ten from *O. nova*. The remaining half had tops hits from *O. nitens*. The single TPS from *T. splendens* (TsTPS) had its top hit from *O. nitens*. To summarize, top hits for the 37 Sciaridae TPSs were from three species: *O. nova, O. nitens,* and *M. subpectinata* (Table 1). All three species belong to the order of Oribatida (oribatid mites) of noninsect arthropods.

**Canonical *TPS* Genes in Noninsect Arthropods.** Existence of *TPS* genes in three species of the order of Oribatida (Table 1), together with the previous identification of *TPS* genes in *D. tinctorium* and *L. delicense* (30), made it interesting to ask how Sciaridae TPSs are evolutionarily related to TPSs of noninsect arthropods. To answer this question, we first set out to identify species of noninsect arthropods that contain *TPS* genes. Genome sequences for a total of 76 species of noninsect arthropods were searched for *TPS* genes using HMMER. *TPS* genes were detected from a total of 14 noninsect arthropod species (Fig. 2*A*). Two of the 14 species, *Allacma fusca* and *Folsomia candida*, belong to the subclass of Collembola, commonly known as springtails. It is worth noting that a total of three species in Collembola were analyzed. The third species, *Orchesella cincta*, does not contain any *TPS* gene. *F. candida* and *O. cincta* belong to the Entomobryomorpha order, while *A. fusca* belongs to the Symphypleona order (Fig. 2*B*).

**Table 1. Top hits of Sciaridae TPSs in GenBank**

| Gene* | Source of top hit (% identity) | Accession of top hit |
|---|---|---|
| BcTPS1 | *Oppia nitens* (37%) | XP_054158546 |
| BcTPS2 | *Medioppia subpectinata* (27%) | CAD7636257 |
| BcTPS3 | *Medioppia subpectinata* (27%) | CAD7636257 |
| BcTPS4 | *Oppia nitens* (37%) | XP_054158545 |
| BcTPS5 | *Medioppia subpectinata* (27%) | CAD7636257 |
| BoTPS1 | *Oppia nitens* (39%) | XP_054156256 |
| BoTPS2 | *Oppia nitens* (39%) | XP_054158545 |
| BoTPS3 | *Medioppia subpectinata* (27%) | CAD7636257 |
| PfTPS1 | *Oppiella nova* (30%) | CAD7644878 |
| PfTPS2 | *Medioppia subpectinata* (26%) | CAD7636257 |
| PfTPS3 | *Medioppia subpectinata* (36%) | CAD7623260 |
| PfTPS4 | *Oppiella nova* (35%) | CAD7655710 |
| PfTPS5 | *Medioppia subpectinata* (26%) | CAD7636257 |
| PfTPS6 | *Oppiella nova* (28%) | CAD7644878 |
| PfTPS7 | *Oppiella nova* (38%) | CAD7655710 |
| PfTPS8 | *Oppia nitens* (36%) | XP_054158545 |
| PfTPS9 | *Oppia nitens* (36%) | XP_054158545 |
| PfTPS10 | *Oppia nitens* (37%) | XP_054158546 |
| PfTPS11 | *Oppia nitens* (38%) | XP_054158545 |
| PfTPS12 | *Oppia nitens* (36%) | XP_054156256 |
| PfTPS13 | *Oppia nitens* (35%) | XP_054158545 |
| PfTPS14 | *Oppia nitens* (36%) | XP_054158545 |
| PfTPS15 | *Oppia nitens* (36%) | XP_054158546 |
| PfTPS16 | *Oppia nitens* (31%) | XP_054158546 |
| PfTPS17 | *Oppia nitens* (34%) | XP_054158546 |
| PfTPS18 | *Oppia nitens* (32%) | XP_054158546 |
| PfTPS19 | *Oppia nitens* (401%) | XP_054158546.1 |
| PfTPS20 | *Oppiella nova* (24%) | CAD7656351 |
| PfTPS21 | *Oppiella nova* (29%) | CAD7652697 |
| PfTPS22 | *Oppiella nova* (38%) | CAD7655710 |
| PfTPS23 | *Oppiella nova* (37%) | CAD7655710 |
| PfTPS24 | *Oppiella nova* (37%) | CAD7655710 |
| PfTPS25 | *Oppiella nova* (37%) | CAD7655710 |
| PfTPS26 | *Oppia nitens* (31%) | XP_054156267 |
| PhTPS1 | *Oppia nitens* (34%) | XP_054156267 |
| PhTPS2 | *Oppiella nova* (24%) | CAD7659323 |
| TsTPS | *Oppia nitens* (43%) | XP_054158545 |

*Bc: *Bradysia coprophila*; Bo: *Bradysia odoriphaga*; Pf: *Phytosciara flavipes*; Ph: *Pseudolycoriella hygida*; Ts: *Trichosia splendens*.

The other 12 *TPS*-containing species belong to the superorder Acariformes (mites). In the dataset we analyzed, the superorder Acariformes contained 15 species, which belong to three suborders: Astigmata (six species), Oribatida (three species), and Prostigmata (six species) (Fig. 2*C*). Except three species in the Prostigmata suborder (*Tetranychus urticae*, *Panonychus citri,* and *Fragariocoptes setiger*), all other species contain *TPS* genes.

**Phylogenetic Analysis of Arthropod TPSs: Alone or with TPSs from Other Organisms.** When arthropod TPSs (>250 amino acids in length) were subjected to phylogenetic analysis, three key observations could be made (Fig. 3). One was that Sciaridae TPSs were separated into two clades. Interestingly, both clades are more closely related to Acariformes (mites) *TPS* genes than Collembola (springtails) TPSs. These two clades correspond to the two clades S1 and S2 when the Sciaridae TPSs were analyzed alone (Fig. 1*B*). The Acariformes TPSs that are most closely related to the S1 clade are from *O. nitens*, *O. nova,* and other mites, and those that are most closely related to the S2 clade are from *M. subpectinata* and *O. nova*. The second observation was that Collembola *TPS* genes formed their own clade, which is further divided into two groups based on their species of origin. The third observation was that the relatedness of Acariformes TPSs is largely congruent with the species' phylogeny.

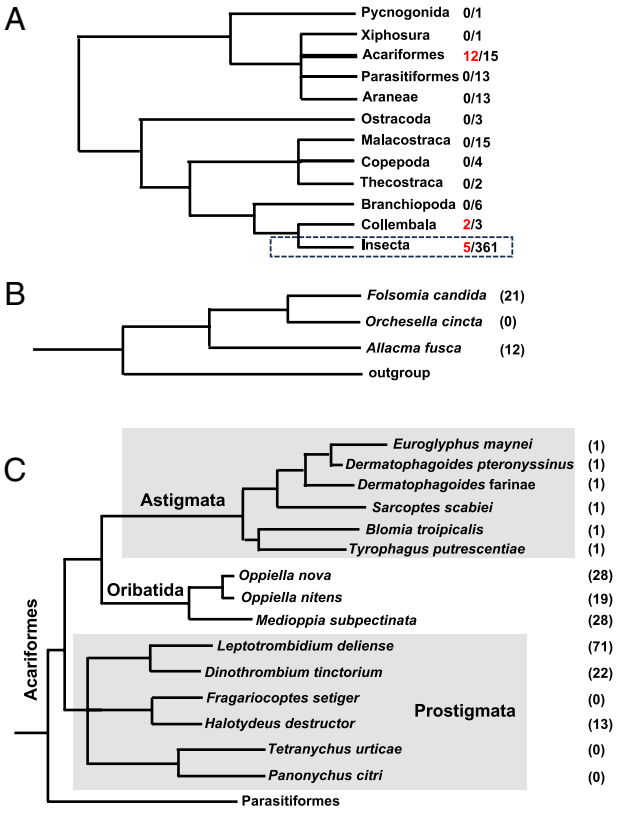


**Fig. 2.** Occurrence and phylogenetic analysis of canonical terpene synthases in noninsect arthropods. (*A*) Presence and absence of canonical terpene synthases in noninsect arthropods. A total of 76 species (*SI Appendix*, Table S2) were analyzed. The phylogeny of noninsect arthropods was adapted from phylogeny and evolution of Arthropods (35). The first number (before the slash) indicates the number of species in certain lineages that were determined to contain *TPS* genes. The second number (after the slash) indicates the total number of species in that lineage that were analyzed. (*B*) Presence and absence of canonical terpene synthases among the three species of Collembola analyzed. The number in parenthesis indicates the number of *TPS* genes identified in the species. (*C*) Presence and absence of canonical terpene synthases among the 15 species of Acariformes analyzed.

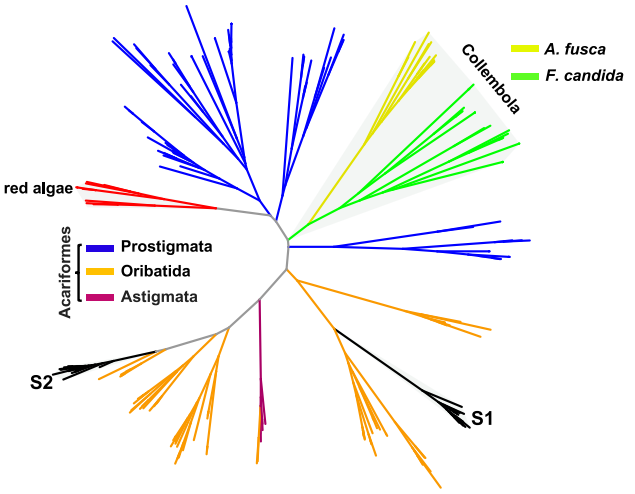


**Fig. 3.** Phylogenetic tree of arthropod TPSs. TPSs of red algae were used as outgroup. The phylogeny tree was built using RAxML with 1,000 bootstraps. Amino acid substitution model determined by ProtTest 3 is JTT+G + F. S1 and S2 stand for the two clades of TPSs from the Sciaridae family defined in Fig. 1*B*.

As noted earlier, *TPS* genes occur widely in bacteria (11) and fungi (16). To gain information about the relatedness of arthropod TPSs with TPSs from bacteria and fungi, a phylogenetic analysis that combined arthropod TPS with bacterial and fungal TPSs was performed. It is interesting to note that Sciaridae TPSs still showed higher affinity to TPSs from Acariformes than those from bacteria and fungi (Fig. 4).

In several species of insects, terpene biosynthesis is catalyzed by IDS-like terpene synthases (3). To understand the relatedness of IDS-like terpene synthases of insects and canonical arthropod TPSs, another phylogenetic analysis was performed. IDS-like terpene synthases are clustered with bona fide IDSs, whereas canonical TPSs from arthropods are clustered with TPSs from other organisms (*SI Appendix*, Fig. S5).

**Representative Insect *TPS* Genes Encode Active Terpene Synthases with Diverse Catalytic Activities.** The presence of canonical *TPS* genes in insects (Fig. 1) and noninsect arthropods (Fig. 2) poses an intriguing question about their functions. To gain an initial assessment of the biochemical functions of TPS enzymes, a subset of *TPS* genes from the three lineages (Insecta, Collembola, and Acariformes) (*SI Appendix*, Fig. S6) were synthesized, cloned into protein expression vector, and expressed in *Escherichia coli* to produce recombinant proteins. Purified recombinant proteins were assayed with the potential substrates GPP, (*E*,*E*)-FPP, and (*E*,*E*,*E*)-GGPP. Their respective products were analyzed using Gas Chromatography-Mass Spectrometry (GCMS) (Fig. 5). The relative abundance of terpene product(s) for each enzyme with each substrate is shown in *SI Appendix*, Table S3.

For the lineages of insects, the entire TPS family with five members (BcTPS1-5) from *B. coprophila* were analyzed (Fig. 5*A*). BcTPS1 was inactive with GPP. BcTPS2 catalyzed the conversion of GPP to four monoterpenes: β-myrcene (**1**, most abundant), (*Z*)-β-ocimene (**2**, least abundant), (*E*)-β-ocimene (**3**), and linalool (**4**). BcTPS3 produced the same four monoterpenes as BcTPS2, but with different relative abundances of the four products: **1** was least abundant and **4** was most abundant. In contrast, BcTPS5 produced only two monoterpenes with **4** as the major product and **1** being the minor product. BcTPS1, BcTPS2, BcTPS3, and BcTPS5 all accepted (*E*,*E*)-FPP as substrate to produce sesquiterpenes (Fig. 5*A*). BcTPS1 produced one major (**5**) and three minor products (**6**–**8**), which were unidentified sesquiterpenes. BcTPS2

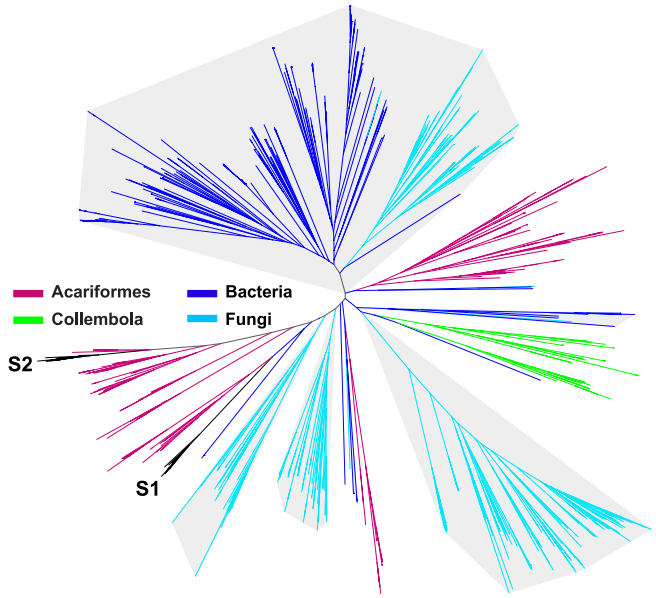


**Fig. 4.** Phylogenetic tree of arthropod TPSs together with bacterial TPS and fungal TPSs. The phylogeny tree was built using RAxML with 1,000 bootstraps. Amino acid substitution model determined by ProtTest 3 is LG+G. S1 and S2 stand for the two clades of TPSs from the Sciaridae family defined in Fig. 1*B*.

and BcTPS3 produced same four sesquiterpenes, β-elemene (**9**), germacrene A (**11**), and two unidentified sesquiterpenes (**10** and **12**). BcTPS5 converted (*E*,*E*)-FPP into (*E*)-β-caryophyllene (**13**). With the diterpene precursor (*E*,*E*,*E*)-GGPP, only BcTPS2 and BcTPS3 showed TPS activity. Whereas BcTPS3 produced several diterpenes, BcTPS2 produced one product (**14**, Fig. 5*A*). BcTPS4 was inactive with each of the three substrates.

From the lineage of Collembola, three TPSs from two species, including *AfTPS2* from *A. fusca* and *FcTPS2* and *FcTPS5* from *F. candida*, were demonstrated to encode active enzymes (Fig. 5*B*). AfTPS2 and FcTPS2 produced **4** as a major product and **1** as a minor one, in addition, FcTPS2 produced an unidentified monoterpene (**15**). FcTPS5 produced a single monoterpene, geraniol (**16**). AfTPS2, FcTPS2, and FcTPS5 accepted (*E*,*E*)-FPP as substrate to produce sesquiterpenes (Fig. 5*B*). FcTPS5 produced one major product, farnesol (**19**), while AfTPS2 and FcTPS2 produced two sesquiterpenes, (*E*)-β-farnesene (**17**) and nerolidol (**18**), with **17** being the major product of FcTPS2. AfTPS2 and FcTPS2 accepted also (*E*,*E*,*E*)-GGPP as substrate to produce diterpenes (**20**–**22**, Fig. 5*B*).

In Acariformes, three *TPS* genes from three species, including *DtTPS1* from *D. tinctorium*, *OniTPS2* from *O. nitens*, and *OnoTPS5* from *O. nova*, were demonstrated to encode active enzymes (Fig. 5*C*). DtTPS1 accepted GPP to produce nine monoterpenes. Among them, limonene (**28**), tricyclene (**23**), and γ-terpinene (**29**) were the major products, and **1**, α-pinene (**24**), camphene (**25**), β-pinene (**26**), α-phellandrene (**27**), and 2-carene (**30**) appeared as minor ones. OnoTPS5 produced a single monoterpene, **2**. DtTPS1, OniTPS2, and OnoTPS5 accepted (*E*,*E*)-FPP as substrate to produce sesquiterpenes (Fig. 5*C*). OniTPS2 produced eight sesquiterpenes. Among them, α-copaene (**34**) and α-muurolene (**38**) were the most abundant products, while β-copaene (**36**), γ-muurolene (**37**), and four unidentified compounds (**33**, **35**, **39**, and **40**) were minor products. DtTPS1 produced three products, **9**, selina-4(15),7-diene (**31**), and germacrene D (**32**). OnoTPS5 produced (*Z*,*E*)-α-farnesene (**41**). The three genes from mites did not use (*E*,*E*,*E*)-GGPP as substrate to produce diterpenoids.

Among the 10 arthropod TPSs that were active with one or more substrates tested (Fig. 5), the percentage of the enzymes active with GGPP was the lowest: Only BcTPS2, BcTPS3, and AfTPS2 were active, and their diterpene products remained unidentified after GCMS analysis. To determine the chemical identities of their respective diterpene products, these three genes were expressed in an engineered *E. coli* system optimized for diterpene production (36). For both BcTPS2 and AfTPS2, we were successful to produce enough quantities for NMR analysis. It is interesting to note that for BcTPS2 the *E. coli* system did not yield compound **14,** the major product from the in vitro system, as the main product. Its main product from the *E. coli* system was determined to have $R_f$ of 0.78 (hexanes) and $[\alpha]_D^{21}$ of −41.31 (c 0.01, $CH_2Cl_2$). Based on its $^1H$ and $^{13}C$ NMR data (*SI Appendix*, Figs. S7–S11 and Table S4), this diterpene (**42**) was determined to be cattleyene. Its $^1H$ and $^{13}C$ spectra matched literature values for cattleyene (37). The product of AfTPS2 was determined to have $R_f$ of 0.87 (hexanes) and $[\alpha]_D^{21}$ of −31.89 (c 0.01, $CH_2Cl_2$). Based on its $^1H$ and $^{13}C$ NMR data (*SI Appendix*, Figs. S12–S16 and Table S5), this diterpene (**21**) was determined to be phomopsene. Its $^1H$ and $^{13}C$ spectra matched literature values for phomopsene (38).

**Expression Patterns of Individual *TPS* Genes and Relevant *IDS* Genes in *B. coprophila*.** The life cycle of *B. coprophila* encompasses four distinct stages: egg, larva, pupa, and adult. The life cycle begins with the female laying eggs in suitable habitats, often in moist organic matter such as decomposing vegetation, compost, or dung. After a brief incubation period, the eggs hatch into larvae, which are the most active and crucial stage for nutrient cycling in their environment. The larvae of *B. coprophila* are typically found in the top layers of soil, where they feed on decaying organic material. As they mature, the larvae enter the pupal stage, undergoing metamorphosis within a protective cocoon. Finally, the adult *B. coprophila* emerges from the pupal casing, ready to repeat the cycle by laying eggs and perpetuating the species. This life cycle is intricately tied to the insect's ecological role in breaking down organic matter and highlights its adaptability to various environments rich in decaying substrates. Whereas the entire lifecycle of *B. coprophila* is 4 to 5 wk long, the adult stage only lasts on the scale of days. Since they do not appear to eat as adults, the primary function of this stage seems to be mating and reproduction.

To gain more insights into the biological role *BcTPS* genes, we analyzed the expression levels of the *BcTPS* genes using available RNA-seq data for both males and females of *B. coprophila* spanning the four lifecycle stages: embryos, larvae, pupae, and adults (32). Interestingly, expression of each *BcTPS* gene is highest in adult males by 1 to 2 orders of magnitude compared to all other life stages in males and all life stages in females. *BcTPS1* is the highest expressed in adult males, followed by *BcTPS2*–*BcTPS5* in decreasing order (Fig. 6). Not only do adult males have the highest expression levels of all *BcTPS* genes, the two highest expressed *BcTPS* genes (*BcTPS1* and *BcTPS2*) are expressed at higher transcript levels than over half of all expressed genes in adult males (*SI Appendix*, Fig. S17), and two others (*BcTPS3* and *BcTPS4*) exhibit moderate expression levels. The high-level expression of the *BcTPS* genes in adult males, not only with respect to other life stages and females, but with respect to other genes in adult males, suggests they are important to the biological function of males at this stage. Since the function of adult *B. coprophila* appears primarily limited to mating, and since terpenoids have known roles in mating, we suspect the *BcTPS* gene products in adult males play a role in mating.

Among the four active BcTPSs (Fig. 5), three exhibited activities with two or more substrates in in vitro assays. Indeed many terpene synthases can use multiple substrates (15), although their

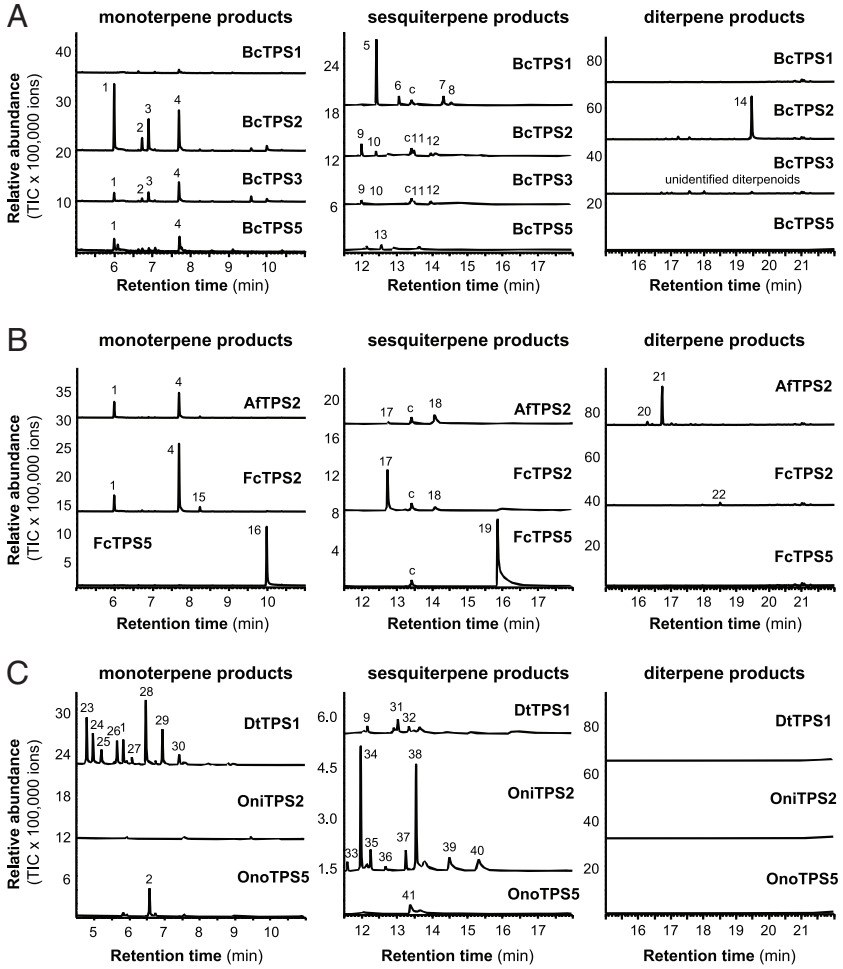


**Fig. 5.** Biochemical characterization of selected TPS enzymes from Insecta (*A*), Collembola (*B*) and Acariformes (*C*). Proteins were heterologously expressed in *E. coli* as N-terminal His-tag fusions, purified, and incubated with the potential substrates GPP, (*E*,*E*)-FPP, and (*E*,*E*,*E*)-GGPP. Reaction products were extracted with hexanes and incubated using gas chromatography-mass spectrometry. **1**, β-myrcene; **2**, (*Z*)-β-ocimene; **3**, (*E*)-β-ocimene; **4**, linalool; **5**-**8**, unidentified sesquiterpenoids; **9**, β-elemene; **10**, unidentified sesquiterpenoid; **11**, germacrene A; **12**, unidentified sesquiterpenoid; **13**, (*E*)-β-caryophyllene; **14**, unidentified diterpenoids; **15**, unidentified monoterpenoid; **16**, geraniol; **17**, (*E*)-β-farnesene; **18**, nerolidol; **19**, farnesol; **20**, unidentified diterpenoid; **21**, phomopsene; **22**, unidentified diterpenoid; **23**, tricyclene; **24**, α-pinene; **25**, camphene; 26, β-pinene; **27**, α-phellandrene; **28**, limonene; **29**, γ-terpinene; **30**, 2-carene; **31**, selina-4(15),7-diene; **32**, germacrene D; **33**, unidentified sesquiterpenoid 7; **34**, α-copaene; **35**, unidentified sesquiterpenoid 8; **36**, β-copaene; **37**, γ-muurolene; **38**, α-muurolene; **39**, unidentified sesquiterpenoid 9; **40**, unidentified sesquiterpenoid 10; **41**, (*Z*,*E*)-α-farnesene; c, contamination.

in vivo function is usually linked to a specific substrate. To gain information about the in vivo substrate of BcTPSs, we analyzed the expression of short-chain *IDS* genes in *B. coprophila*. Searching the genome sequence of *B. coprophila*, we identified one farnesyl diphosphate synthase gene (XP_037043359.1), which was designated as *BcFPPS*, and one geranylgeranyl diphosphate synthase gene (XP_037044777.1), which was designated as *BcGGPPS*. The biochemical function of BcFPPS is to synthesize FPP, which serves as the substrate for sesquiterpene production. Similarly, the biochemical function of BcGGPPS is to produce GGPP, the substrate for diterpene synthesis. The expression level of the *BcFPPS* is the highest in adult male (Fig. 6*F*), whereas the expression levels of *BcGGPPS* are relatively consistent across sexes and developmental stages (Fig. 6*G*).

## Discussion

This study revealed the existence of canonical terpene synthases (*TPS*) genes in insects, demonstrated via representatives that they encode active enzymes, and established an evolutionary model on the origin of insect *TPS* genes. Widely occurring in plants (15), fungi (16, 39), and bacteria (11), *TPS* genes were thought to be absent in insects. Instead, insects were believed to solely employ distinct IDS-like terpene synthases for terpenoid biosynthesis (3).

Through systematic data mining, this study confirmed the widespread absence of *TPS* genes in most insects. However, these genes were identified in five species, all belonging to the Sciaridae family within the Diptera order. The identification of *TPS* genes in springtails and mites (Fig. 4), both of which are noninsect arthropods, suggests that the presence of *TPS* genes in Sciaridae is not an isolated event among arthropods. The similarity of Sciaridae *TPS*s to *TPS*s from mites, rather than to those from springtails, bacteria, or fungi, implies horizontal gene transfer (HGT) of *TPS* from mites to Sciaridae. In *B. coprophila*, a model dark-winged fungus gnat species from the Sciaridae family, *TPS*s were demonstrated to encode functional enzymes and exhibit male adult-specific expression, suggesting that the acquired *TPS* genes have evolved adaptive functions in the recipient species.

Presence of *TPS* genes in all five species of the Sciaridae family analyzed suggests that *TPS* genes may be ubiquitous in the Sciaridae family. A number of inferences can be made from the phylogenetic analysis of the Sciaridae *TPS*s. First, it suggests the existence of two ancestral genes leading to the two extant groups. One of the ancestral copies existed in the common ancestor of all

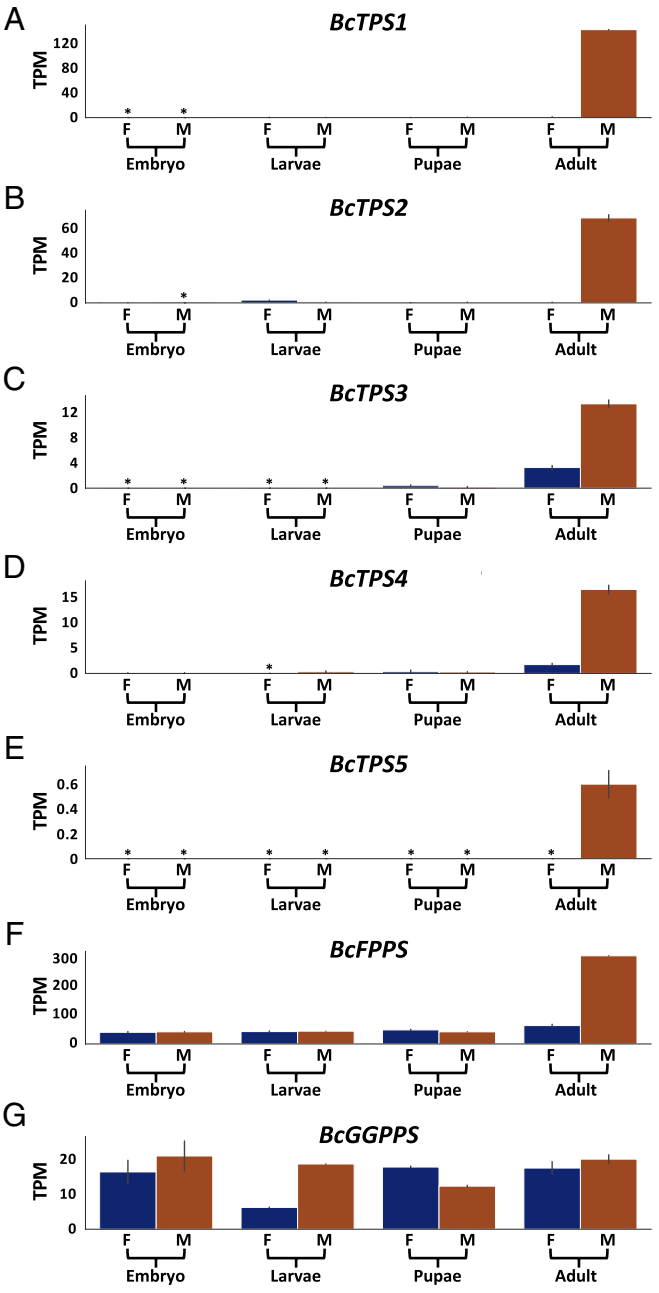


**Fig. 6.** Expression of five *BcTPS* genes (*A*–*F*) and two related isoprenyl diphosphate synthase genes (*F* and *G*) in *B. coprophila*. The two isoprenyl diphosphate synthase genes analyzed are farnesyl diphosphate synthase gene (*BcFPPS*) and geranylgeranyl diphosphate synthase gene (*BcGGPPS*). This analysis was based on published RNAseq data (32), which were obtained for both females (F, blue) and males (M, red) at four developmental stages during a complete life cycle of the insect including embryo, larvae, pupae, and adult. The expression levels of the seven genes are represented as TPM. Asterisk indicates that no transcript was detected. Differential expression analysis using edgeR demonstrated that the higher *BcTPS* expression levels in adult males are statistically significant compared to all other samples. See *SI Appendix*, Fig. S17 to compare *BcTPS* expression levels to the distribution of all genes.

the five species (leading to clade S1) and the other copy existed in the common ancestor of the four species after the divergence of their common ancestor from *T. splendens*. Another possibility is that both ancestral genes were present in the common ancestor of all five species, with one of the ancestral copies subsequently lost in *T. splendens*. The second observation is that the TPS family underwent lineage-specific expansion. While *TPS* gene duplication is apparent in the common ancestor of *B. coprophila* and *B. odoriphaga* (in clade S1, Fig. 1), species-specific expansion is evident in *B. coprophila* and *P. flavipe*. In *B. coprophila*, species-specific expansion is partly due to tandem duplications with *BcTPS2*, *BcTPS3,* and *BcTPS5* existing in tandem repeats (*SI Appendix*, Fig. S18).

The existence of *TPS* genes only in Sciaridae, but not in a wide range of insect taxa analyzed, suggests that these genes are derived from noninsect sources, most likely through horizontal gene transfer. Many studies have demonstrated important roles of horizontal gene transfer during the evolution of insects (40), through which insects have gained genes from a variety of sources including bacteria (41), viruses (42), fungi (43), and plants (44). Horizontal gene transfer as a mechanism responsible for the evolution of *TPS* genes has been inferred in a number of studies (45). In all cases, horizontal gene transfer of *TPS* genes involves microbial organisms, particularly with bacteria or fungi as donor (45). The similarity of Sciaridae TPSs to Acariformes TPSs was a surprising observation: The closer relatedness of Sciaridae TPSs to Acariformes TPSs than bacterial and fungal TPSs implies that Acariformes were the donor of *TPS* genes to Sciaridae. The most sensible explanation is that Sciaridae ancestor acquired *TPS* genes from the ancestor of Oribatid mites. Oribatid mites are diverse and ubiquitous microarthropods in soil; *O. nitens* is a polyphagous fungivore. Such horizontal gene transfer within arthropods is rarely proposed. In a recent evolutionary study of cytochrome *P450* genes of the insect *B. coprophila*, some of its genes were hypothesized to have acquired from springtails or mites (46). *P450* genes are ubiquitous in arthropods. As such, the inference of their evolution could be distorted due to different rates of evolution of *P450* genes in different species/lineages. In our study, the presence or absence of *TPS* genes is unambiguous. As such, our study presents a stronger case of horizontal gene transfer from noninsect arthropod to insects. More interestingly, the two clades of Sciaridae TPSs have respective homologs in the same species of mites (Fig. 3). This suggests the transfer of both copies from Acariformes to Sciaridae, rather than duplication-derived two ancestral copies in Sciaridae. Another scenario for the origin of *TPS* genes in arthropods is independent horizontal gene transfer of *TPS* genes into Acariformes, Collembola, and Sciaridae from same microorganisms, especially endosymbiont, such as *Rickettsia* or *Wolbachia*. Such endosymbionts are known to occur in arthropods (32). As they are found in eggs, endosymbiont genes can easily be horizontally transferred to the host. In our analysis, no *TPS* genes were found in *Rickettsia* and only one *TPS* gene was found in *Wolbachia*, which is, however, more similar to bacterial TPSs than to TPSs from arthropods. Thus, intraphylum horizontal gene transfer from Acariformes to Sciaridae is a better supported model to explain the origin of *TPS* genes in insects.

While biochemical activities of canonical TPSs from many organisms such as bacteria (11, 12), fungi (16) and plants (15) have been determined, the selected TPSs we analyzed in this study (Fig. 5) present a set of canonical TPSs from arthropods to be characterized. A number of observations can be made. First, as their counterparts in other lineages of organisms, arthropod TPSs exhibit diverse catalytic activities. Many of them are multiproduct enzymes. Second, their biochemical activities can reflect their evolution. Taking BcTPSs as example, phylogenetic analysis indicates that BcTPS1/4 and BcTPS2/3/5 have different evolutionary origins. Reflected in their sesquiterpene product profiles, there is no overlap between the products of BcTPS1 and BcTPS2/3/5 (Fig. 5). Third, duplication followed by functional divergence is apparent in *BcTPS2/3/5*. With the three of them clustered in

tandem (*SI Appendix*, Fig. S18), BcTPS5 has a distinct sesquiterpene product compared to those of BcTPS2 and BcTPS3 (Fig. 5).

Prior to our work, several diterpene synthases catalyzing the formation of cattleyene and phomopsene have been reported. These include cattleyene synthase (CyS) from bacterium *Streptomyces cattleya* (37), phomopsene synthase (PmS) from bacterium *Allokutzneria albata* (38) and phomopsene synthase gene (PaPS) from the fungus *Phomopsis amygdali* (47). PaPS is unusual containing two α-domains. Neither BcTPS2 nor AfTPS2 shares significant sequence homology with the bacterial or fungal TPSs that synthesize cattleyene or phomopsene. Sequence identities of BcTPS2–CyS, AfTPS2–PmS, and AfTPS2–PaPS are all about 20%, suggesting convergent evolution of such diterpenes synthases in bacteria, fungi, and arthropods.

For the two mites (*L. delicense* and *D. tinctorium*) in which *TPS* genes were previously identified, multiple possible biological roles of the terpenoid product of TPSs have been suggested (30). Given the limited natural predators of these mites and their tendency to be quickly rejected by predators when offered in laboratory settings, terpenoids may serve as repellents. Additionally, considering the instances of cannibalism among adults and the ectoparasitism of free-living stages by larvae of these species, the terpenoids they produce might play a role in chemical communication within their species or even between closely related species, potentially aiding in kin recognition.

While much work is needed to elucidate the biological role of terpenoids produced by TPSs in Sciaridae, in the model insect *B. coprophila*, *BcTPS* genes appear to have evolved a sex- and developmental stage-specific function (Fig. 6). It will be interesting to determine whether the products of TPSs serve as a male sex pheromone in *B. coprophila*. The function of the adult stage of *B. coprophila* appears to be almost exclusively for reproduction, since adults do not seem to eat (unlike *Drosophila*, for example), the adult life stage is very short compared to other flies (e.g., on the order of days compared to weeks for *Drosophila*), and successful mating between adult males and females appears to result in dying soon after (much sooner than would otherwise occur). The high levels of expression of *BcTPS* genes in adult males (Fig. 6) suggest a possible role of their terpene products in attraction and/or reproduction. The similar pattern of expression of *BcFPPS* (Fig. 6*F*), not *BcGGPPS* (Fig. 6*G*), suggests that BcTPSs most likely function as sesquiterpene synthase in vivo. Considering the diversity of arthropods, particularly insects (48), and the relatively small number of species analyzed in this study, it would not be surprising if canonical *TPS* genes were discovered in other insects and noninsect arthropods, which will help clarify their evolutionary origin and provide new insights into the biological significance of such canonical *TPS* genes.

## Methods

**Sequence Retrieval and Analysis.** Insect genome sequences deposited at GenBank (https://www.ncbi.nlm.nih.gov/genbank/) were downloaded on 08/13/2023. These insects' genomes were arranged according to their known phylogeny (31). For insect species with sequenced genomes available for multiple strains, only the genome sequence of one selected strain was retained. Our final database includes the proteome sequences of 359 insect species (*SI Appendix*, Table S1). For noninsect arthropods, we compiled data from a total of 76 species (*SI Appendix*, Table S2). Both the insect and noninsect arthropod databases were searched against the Pfam database (version 35.0; pfam-legacy.xfam.org) using HMMER 3.0 (http://hmmer.org). Given the high sequence divergence of terpene synthases, we initially searched for α-domain-containing terpene synthases (TPSs) using three HMM profiles: Terpene_synth_C (PF03936), Terpene_syn_C_2 (PF19086), and TRI5 (PF06330), with an e-value cutoff of 1e-3. Subsequently, each putative α-domain-containing TPS was used as a query to search the conserved domain database (https://www.ncbi.nlm.nih.gov/cdd/). Only sequences with domain hits to the "Isoprenoid_Biosyn_C1 superfamily", "Terpene_cyclase_nonplant_C1", or "Terpene_syn_C_2 superfamily" with an e-value less than 1e-2 were classified as putative α-domain-containing TPSs. If any of these three conserved domains was the only domain detected, the protein was defined as an α-domain-only TPS. For the two species from the Sciaridae family, *P. flavipes* and *T. splendens*, which lack annotated proteomes, we conducted a tblastn analysis to identify *TPS* genes. Two criteria were used to define putative full-length intact *TPS* genes: 1) the presence of two motifs conserved in known TPSs, namely the "DDxxD" motif and the "NSE/DTE" motif, and 2) a minimum length cutoff of 250 amino acids.

**Genomic PCR for Validation of Representative *TPS* Gene from *B. coprophila* as a Nuclear Gene.** Genomic DNA from *B. coprophila* was extracted separately from pooled adult males and from pooled adult females using DNAzol (Thermo Fisher). The line of *B. coprophila* used was the same described for the reference genome (34). A pair of primers (forward sequence 5′-CTCGAGTGTACGGCGACAGAATG-3′ and reverse sequence 5′-GACAGATCGAAAGCGCCAAGAATG-3′) were used for PCR using the mixed male and female genomic DNA as template. The amplicon was amplified using PfuUltra II Hotstart (https://www.agilent.com/) and cloned into pGEM-T Easy vector (Promega) and fully sequenced.

**Genomic Validation of *TPS* Genes in the *B. coprophila* Reference Genome Using Computational Biology.** We analyzed the chromosomal context of the selected *BcTPS* gene models from automated annotations (see expression section below for more details) (32) that were lifted over to the updated chromosome-scale reference, Bcop_v2 (GenBank GCA_014529535.2; WGS VSDI02; release date 2023-01-04) (34). The presence of the *BcTPS* genes in the chromosome-scale scaffolds argues against contamination since these scaffolds were constructed by breaking contigs at sites where the expected Hi-C signal was disrupted or malformed, which is expected in the case of contamination, followed by Hi-C-guided scaffolding as described previously (34). Nevertheless, to further investigate whether there were disruptions in expected Hi-C signal emanating from the *BcTPS* genes, previously published Hi-C reads from male pupae available in the NCBI BioProject database under the accession number of PRJNA291918 (34) were mapped as in ref. 34. Hi-C interactions between a given *BcTPS* gene and all other sites in the genome were obtained when at least one of the mates from Hi-C paired-end reads mapped within the boundaries of the gene. Hi-C interaction frequencies between a given *BcTPS* gene with other areas of the same chromosome were counted in 1 Mb bins (*SI Appendix*, Fig. S3 *A–C*) or 10 kb bins (*SI Appendix*, Fig. S3 *D–F*) that emanated away from the gene start and end sites. A pseudocount of 1 was added to all bins in order to convert to log10 for better visualization of faraway contacts, as Hi-C interactions decay exponentially with distance. Thus, bins with a log10 of 0 had no interactions (only a pseudocount of 1). The *BcTPS* genes have typical Hi-C interaction frequencies that decay with distance, and include interactions with highly conserved Dipteran genes (see below), and are overall supported by the Hi-C data as being part of the nuclear chromosomes, not a separate contamination source.

To interrogate alignment signals and rule out disruptions in coverage over the BcTPS regions, previously published genomic DNA (gDNA) datasets from a variety of technologies and life stages were mapped to the chromosome-scale scaffolds (Bcop_v2). Oxford Nanopore MinION and PacBio long reads from embryonic gDNA (32) available in the NCBI BioProject PRJNA291918 were mapped with Minimap2 (49). BioNano Genomics long-range optical maps from male pupae gDNA (32) also available in the NCBI BioProject PRJNA291918 were mapped with Maligner (50) similar to that in ref. 32. Illumina paired-end reads (33) from male and female adult gDNA were mapped as in ref. 33. Genomic DNA sequencing coverage bigWigs from each dataset for visualization in IGV (51) were obtained using BEDtools (52) and UCSC Kent Utilities (53). PacBio and Nanopore long read coverage was limited to reads with a minimum length of 5 kb and minimum MAPQ of 40. BioNano optical map coverage was computed from molecules with a minimum length of 150 kb that had alignments with a maximum Maligner M-score of −10. Disruptions in coverage from all datasets were seen over gaps in the assembly as expected, but not over BcTPS regions, indicating again that the *BcTPS* genes within the *B. coprophila* scaffolds are likely not from contamination.

As a final analysis, we queried whether there were individual long reads and long-range optical maps that connected *BcTPS* genes with "conserved fly genes" on long and ultralong DNA molecules. "Conserved fly genes" were *B. coprophila* genes in the NCBI annotation that had one or more orthologs in all five fly species queried, which spanned the Dipteran tree (*B. coprophila*, *P. hygida*, *Aedes aegypti*, *Anopheles gambiae*, and *Drosophila melanogaster*), as determined by OrthoFinder (54). The individual long reads and optical maps selected around the BcTPS loci were visualized in IGV as BED intervals or BAM alignments. Selected PacBio, Nanopore, and optical map alignments at BcTPS loci had lengths up to 24.4 kb, 48.4 kb, and 402 kb, respectively. Corresponding to DNA molecules that contained both *BcTPS* and conserved fly genes, or connected them by at least partial overlaps, were many optical maps and at least 1–2 nanopore reads per gene and 1–2 PacBio reads per gene (except *BcTPS-4*). There were many examples of two overlapping long reads from either technology that connected *BcTPS* genes to conserved fly genes. Overall, the evidence supports *BcTPS* genes to be on the same DNA molecules as conserved fly genes, and not from a contaminating source.

**Phylogenetic Analysis.** Proper datasets of terpene synthases were subject to phylogenetic reconstruction. MAFFT v.7.520 (55) was used to perform the multiple sequence alignment with maxiterate 1000. The maximum-likelihood phylogenetic trees were built with RAxML through the ISAAC-NG platform (https://oit.utk.edu/hpsc/isaac-open-enclave-new-kpb) employing the appropriate amino acid substitution model predicted using ProtTest 3 (56), along with 1,000 bootstrap replicates. The resulting trees were visualized using iTOL (version 6.7.5) (https://itol.embl.de/).

**Heterologous Expression and Characterization of TPS Enzymes.** Selected Arthropod *TPS* genes were synthesized and cloned into protein expression vector pET28a. Liquid cultures of *E. coli* strain BL21 (DE3) (Invitrogen) harboring the different expression constructs were grown in lysogeny broth at 37 °C and 220 rpm until an OD600 of 0.7. Expression was induced by adding IPTG to a final concentration of 0.5 mM and the cultures were subsequently incubated at 18 °C and 220 rpm for 16 h. The cells were harvested by centrifugation at 3,200 g for 10 min, resuspended in extraction buffer [50 mM Tris-HCl pH 8, 500 mM NaCl, 20 mM imidazole, 5% (v/v) glycerol, 50 mM glycine, EDTA-free protease inhibitor (1 tablet/50 mL buffer, freshly added), and lysozyme (10 mg/50 mL buffer, freshly added)], and disrupted by sonication for 2 min (2 s on, 3 s off) on ice (Bandelin UW 2070). Cell debris was removed by centrifugation (35,000 g at 4 °C for 20 min) and the N-terminal His-tagged proteins were purified from the supernatant using NiNTA agarose (Qiagen) according to the manufacturer's protocol. The buffer of the eluted protein samples was exchanged for assay buffer [100 mM MOPS pH 7.5, 10% (v/v) glycerol] by using Amicon 10 K Concentrator columns (Merck Millipore).

TPS assays were carried out in 1 mL GC glass vials containing protein, 50 μM substrate [GPP, (*E, E*)-FPP, or (*E, E, E*)-GGPP] and 10 mM $MgCl_2$ in assay buffer. Assays were overlaid with 100 μL hexane and incubated for 1 h at room temperature. Analysis of enzyme products was conducted using an Agilent 6890 Series gas chromatograph coupled to an Agilent 5973 quadrupole mass selective detector (Agilent Technologies, Santa Clara, CA; injector temperature, 220 °C; interface temp, 250 °C; quadrupole temp, 150 °C; source temp, 230 °C; electron energy, 70 eV). Terpene products were separated using a ZB5 column (Phenomenex, Aschaffenburg, Germany; 30 m × 0.25 mm × 0.25 μm) and He as carrier gas (1.5 mL/min). 1 μL of the hexane phase was injected without split at an initial oven temperature of 60 °C. The temperature was held for 2 min and then increased to 250 °C with a gradient of 10 °C/min, and then further increased to 300 °C with a gradient of 100 °C/ min and a hold of 2 min.

**Isolation of Diterpenes from *E. coli* and NMR Analysis.** To isolate the diterpene products, *E. coli* BL21 Star (DE3) was cotransformed individually with pJR1064b [pCDF-Duet harboring genes to produce GGPP in vivo via an alternative isoprenoid precursor pathway (57)] and cultivated in LB medium with kanamycin (50 mg $L^{-1}$) and streptomycin (50 mg $L^{-1}$). The overnight cultured cells were then transferred into 12 L of fresh Terrific Broth (TB) media at 37 °C, and incubated with shaking at 200 rpm until reaching an $OD_{600}$ of 1.0. Isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.5 mM final concentration) and isoprenol (4 mM, final concentration) were added to the culture and the culture was incubated at 28 °C for an additional 24 h with shaking. Cells were harvested by centrifugation at 5,000 g for 10 min at 20 °C and subsequently transferred to a glass beaker. The cell pellets were extracted with acetone and the organic phase was extracted with the same volume of hexanes three times. The hexanes extractions were combined and concentrated in vacuo at room temperature. Subsequently, the resulting extract was redissolved in hexanes and purified by silica chromatography, employing an isocratic hexanes mobile phase. Fractions containing the target products were combined, individually, and subjected to further purification using preparative HPLC, which was conducted using an Agilent 1260 Infinity II LC equipped with an Agilent Eclipse XDB-C18 column (250 mm × 21.2 mm, 7 μm). All $^{1}H$, $^{13}C$, and 2D NMR ($^{1}H$-$^{1}H$ COSY, $^{1}H$-$^{13}C$ HSQC, and $^{1}H$-$^{13}C$ HMBC) experiments were conducted in $C_6D_6$ at 600 MHz for $^{1}H$ and 151 MHz for $^{13}C$ nuclei on a Bruker Avance III Ultrashield 600.

**Expression of *TPS* and *IDS* Genes in Transcriptomes of *B. coprophila*.** BcTPS1–5 protein sequences defined using methods above on the Bcop_v1 genome sequence (NCBI GenBank GCA_014529535.1, WGS VSDI01) (32) were subsequently used to fish out the corresponding BcTPS proteins in two automated gene annotation sets for *B. coprophila* using BLAST. Gene models corresponding to all five *BcTPS* genes were found in the protein set previously generated using Maker2 annotation pipeline (32), and gene models for four (*BcTPS 1-4*, not *5*) were identified in the NCBI *B. coprophila* Annotation Release 100 (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Bradysia_coprophila/100/): BcTPS1 (Bcop_v1_g001633; XP_037024969.1), BcTPS2 (Bcop_v1_g003386; XP_037027991.1), BcTPS3 (Bcop_v1_g003391; XP_037028003.1), BcTPS4 (Bcop_v1_g014573; XP_037049366.1), and BcTPS5 (Bcop_v1_g003389; N/A). Also identified were gene models for BcFPPS (XP_037043359.1) and BcGGPPS (XP_037044777.1). For further evaluation, we used the gene model sets that were lifted over to the chromosome-scale Bcop_v2 reference genome as described previously (34). To check for expression of the *BcTPS* genes and evaluate the gene models, we used published RNA-seq datasets available in the NCBI BioProject database under the accession number of PRJNA748150 (32) spanning four life stages (embryos, larvae, pupae, and adults) from both sexes as well as published RNA-seq datasets from i) irradiated and control female larvae (58) and ii) adult male testis (59). To evaluate the concordance between RNA-seq alignments and *BcTPS* gene models, including visually inspecting exon–intron junctions and UTRs, the RNA-seq datasets were mapped to the Bcop_v2 genome using STAR (version 2.7.10a_alpha_220818) (60), agnostic of gene annotation sets. The RNA-seq alignments in the resulting BAM files were directly examined for consistency with gene structures in both annotation sets using IGV (51). Following this evaluation, the gene models for *BcTPS 1-4*, *BcFPPS,* and *BcGGPPS* from "NCBI *B. coprophila* Annotation Release 100" were deemed better than the corresponding gene models from Maker2 (32) and selected for further analysis. Since the NCBI annotation set was missing a model for BcTPS5, the Maker2 model for BcTPS5 was simply added to the NCBI gene set before further transcriptome-wide RNA-seq analysis. *SI Appendix*, Fig. S2 shows the RNA-seq alignment evidence with respect to the chosen *BcTPS* gene models using a comprehensive merge of all RNA-seq datasets from four life stages for both sexes (32). To quantify transcript levels in each RNA-seq dataset for all genes in the NCBI annotation set (with BcTPS5 added in), RSEM (v1.3.1) (61) was used with STAR (version 2.7.10a_alpha_220818) (60) using the "rsem-calculate-expression" pipeline. Transcripts Per Million (TPMs) were extracted from the RSEM gene-level results, and the corresponding RSEM-calculated expected counts were used with EdgeR (62) for between-sample normalization and differential expression analysis. Plots of TPMs and EdgeR results were made in R and Python.

**Data, Materials, and Software Availability.** Raw fasta files, aligned files, and newick trees for phylogenetic analyses have been deposited in Figshare (63). Functionally characterized TPSs have been deposited in GenBank as PQ349591 (64), PQ349592 (65), PQ349593 (66), PQ349594 (67), PQ349595 (68), PQ349596 (69), PQ349597 (70), PQ349599 (71), PQ349600 (72), and PQ349601 (73). Raw NMR data have been deposited to NP-MRD (https://np-mrd.org) with deposition ID NP0019639 (74) for cattleyene and NP0341814 (75) for phomopsene. All other data are included in the manuscript and/or *SI Appendix*.

**ACKNOWLEDGMENTS.** The project was supported by a SPRINT award from the University of Tennessee, Institute of Agriculture (to F.C.), the Max Planck Society (funds to T.G.K.), and start-up from the University of Florida (to J.D.R.).

Author affiliations: [a]Department of Plant Sciences, University of Tennessee, Knoxville, TN 37996; [b]HHMI Research Laboratories, Carnegie Institution for Science, Baltimore, MD 21218; [c]Department of Embryology, Carnegie Institution for Science, Baltimore, MD 21218; [d]Department of Natural Product Biosynthesis, Max-Planck-Institute for Chemical Ecology, Jena 07745, Germany; and [e]Department of Chemistry, University of Florida, Gainesville, FL 32611

1. J. A. Thomas, Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**, 339–357 (2005).
2. E. D. Morgan, *Biosynthesis in Insects* (Royal Society of Chemistry, 2010).
3. Z. Rebholz *et al.*, Emergence of terpene chemical communication in insects: Evolutionary recruitment of isoprenoid metabolism. *Protein Sci.* **32**, e4634 (2023).
4. D. Tholl, Z. Rebholz, A. V. Morozov, P. E. O'Maille, Terpene synthases and pathways in animals: Enzymology and structural evolution in the biosynthesis of volatile infochemicals. *Nat. Prod. Rep.* **40**, 766–793 (2023).
5. J. H. Tumlinson *et al.*, Sex pheromones produced by male boll weevil: Isolation, identification, and synthesis. *Science* **166**, 1010–1012 (1969).
6. D. K. Zahn, J. A. Moreira, J. G. Millar, Identification, synthesis, and bioassay of a male-specific aggregation pheromone from the harlequin bug, *Murgantia histrionica*. *J. Chem. Ecol.* **34**, 238–251 (2008).
7. X. Song, Y. G. Qin, Y. Yin, Z. X. Li, Identification and behavioral assays of alarm pheromone in the vetch aphid *Megoura viciae*. *J. Chem. Ecol.* **47**, 740–746 (2021).
8. R. K. Vander Meer, F. Alvarez, C. S. Lofgren, Isolation of the trail recruitment pheromone of *Solenopsis invicta*. *J. Chem. Ecol.* **14**, 825–838 (1988).
9. E. Pichersky, R. A. Raguso, Why do plants produce so many terpenoid compounds? *New Phytol.* **220**, 692–702 (2018).
10. M. B. Quin, C. M. Flynn, C. Schmidt-Dannert, Traversing the fungal terpenome. *Nat. Prod. Rep.* **31**, 1449–1473 (2014).
11. Y. Yamada *et al.*, Terpene synthases are widely distributed in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 857–862 (2015).
12. J. D. Rudolf, T. A. Alsup, B. Xu, Z. Li, Bacterial terpenome. *Nat. Prod. Rep.* **38**, 905–980 (2021).
13. Q. Jia *et al.*, Origin and early evolution of the plant terpene synthase family. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2100361119 (2022).
14. Q. Jia *et al.*, Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12328–12333 (2016).
15. F. Chen, D. Tholl, J. Bohlmann, E. Pichersky, The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* **66**, 212–229 (2011).
16. C. Schmidt-Dannert, Biosynthesis of terpenoid natural products in fungi. *Adv. Biochem. Eng. Biotechnol.* **148**, 19–61 (2015).
17. M. J. Fischer, C. Rustenhloz, V. Leh-Louis, G. Perrière, Molecular and functional evolution of the fungal diterpene synthase genes. *BMC Microbiol.* **15**, 221 (2015).
18. X. Chen *et al.*, Discovery of bifunctional diterpene cyclases/synthases in bacteria supports a bacterial origin for the plant terpene synthase gene family. *Hortic. Res.* **11**, uhae221 (2024).
19. A. B. Gilg, J. C. Bearfield, C. Tittiger, W. H. Welch, G. J. Blomquist, Isolation and functional expression of an animal geranyl diphosphate synthase and its role in bark beetle pheromone biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9760–9765 (2005).
20. A. B. Gilg, C. Tittiger, G. J. Blomquist, Unique animal prenyltransferase with monoterpene synthase activity. *Naturwissenschaften* **96**, 731–735 (2009).
21. K. Darragh *et al.*, A novel terpene synthase controls differences in anti-aphrodisiac pheromone production between closely related *Heliconius* butterflies. *PLoS Biol.* **19**, e3001022 (2021).
22. F. Beran *et al.*, The aggregation pheromone of *Phyllotreta striolata* (Coleoptera: Chrysomelidae) revisited. *J. Chem. Ecol.* **42**, 748–755 (2016).
23. J. Lancaster *et al.*, De novo formation of an aggregation pheromone precursor by an isoprenyl diphosphate synthase-related terpene synthase in the harlequin bug. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E8634–E8641 (2018).
24. G. Wei *et al.*, Evolution of isoprenyl diphosphate synthase-like terpene synthases in fungi. *Sci. Rep.* **10**, 14944 (2020).
25. X. Chen *et al.*, Terpene synthase genes in eukaryotes beyond plants and fungi: Occurrence in social amoebae. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12132–12137 (2016).
26. G. Wei *et al.*, Terpene biosynthesis in red algae Is catalyzed by microbial type but not typical plant terpene synthases. *Plant Physiol.* **179**, 382–390 (2019).
27. P. D. Scesa, Z. Lin, E. W. Schmidt, Ancient defensive terpene biosynthetic gene clusters in the soft corals. *Nat. Chem. Biol.* **18**, 659–663 (2022).
28. I. Burkhardt, T. de Rond, P. Y. Chen, B. S. Moore, Ancient plant-like terpene biosynthesis in corals. *Nat. Chem. Biol.* **18**, 664–669 (2022).
29. Y. Jung *et al.*, Function and structure of a terpene synthase encoded in a giant virus genome. *J. Am. Chem. Soc.* **145**, 25966–25970 (2023).
30. X. Dong *et al.*, Genomes of trombidid mites reveal novel predicted allergens and laterally transferred genes associated with secondary metabolism. *Gigascience* **7**, giy127 (2018).
31. B. Misof *et al.*, Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
32. J. M. Urban *et al.*, High contiguity de novo genome assembly and DNA modification analyses for the fungus fly, *Sciara coprophila*, using single-molecule sequencing. *BMC Genomics* **22**, 643 (2021).
33. R. B. Baird *et al.*, Recent evolution of a maternally acting sex-determining supergene in a fly with single-sex broods. *Mol. Biol. Evol.* **40**, msad148 (2023).
34. J. M. Urban, S. A. Gerbi, A. C. Spradling, Chromosome-scale scaffolding of the fungus gnat genome (Diptera: *Bradysia coprophila*). bioRxiv [Preprint] (2022). https://doi.org/10.1101/2022.11.03.515061 (Accessed on 10 January 2024).
35. G. Giribet, G. D. Edgecombe, The phylogeny and evolutionary history of arthropods. *Curr. Biol.* **29**, R592–R602 (2019).
36. B. Xu, W. Ning, X. Wei, J. D. Rudolf, Mutation of the eunicellane synthase Bnd4 alters its product profile and expands its prenylation ability. *Org. Biomol. Chem.* **20**, 8833–8837 (2022).
37. J. Rinkel, S. T. Steiner, J. S. Dickschat, Diterpene biosynthesis in Actinomycetes: Studies on cattleyene synthase and phomopsene synthase. *Angew. Chem. Int. Ed. Engl.* **58**, 9230–9233 (2019).
38. L. Lauterbach, J. Rinkel, J. S. Dickschat, Two bacterial diterpene synthases from *Allokutzneria albata* produce bonnadiene, phomopsene, and allokutznerene. *Angew. Chem. Int. Ed. Engl.* **57**, 8280–8283 (2018).
39. R. Chen *et al.*, Systematic mining of fungal chimeric terpene synthases using an efficient precursor-providing yeast chassis. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023247118 (2021).
40. Y. Li *et al.*, HGT is widespread in insects and contributes to male courtship in lepidopterans. *Cell* **185**, 2975–2987.e10 (2022).
41. F. Husnik *et al.*, Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* **153**, 1567–1578 (2013).
42. L. Gasmi *et al.*, Horizontally transmitted parasitoid killing factor shapes insect defense to parasitoids. *Science* **373**, 535–541 (2021).
43. N. A. Moran, T. Jarvik, Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* **328**, 624–627 (2010).
44. C. Gilbert, F. Maumus, Sidestepping darwin: Horizontal gene transfer from plants to insects. *Curr. Opin. Insect Sci.* **57**, 101035 (2023).
45. Q. Jia *et al.*, Terpene synthase genes originated from bacteria through horizontal gene transfer contribute to terpenoid diversity in fungi. *Sci. Rep.* **9**, 9223 (2019).
46. R. Feyereisen, J. M. Urban, D. R. Nelson, Aliens in the CYPome of the black fungus gnat, *Bradysia coprophila*. *Insect Biochem. Mol. Biol.* **159**, 103965 (2023).
47. T. Toyomasu *et al.*, Biosynthetic gene-based secondary metabolite screening: A new diterpene, methyl phomopsenonate, from the fungus *Phomopsis amygdali*. *J. Org. Chem.* **74**, 1541–1548 (2009).
48. D. Grimaldi, M. S. Engel, *Evolution of the Insects* (Cambridge University Press, 2005).
49. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
50. L. M. Mendelowitz, D. C. Schwartz, M. Pop, Maligner: A fast ordered restriction map aligner. *Bioinformatics* **32**, 1016–1022 (2016).
51. H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform.* **14**, 178–192 (2013).
52. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
53. W. J. Kent, A. S. Zweig, G. Barber, A. S. Hinrichs, D. Karolchik, BigWig and BigBed: Enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).
54. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
55. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
56. D. Darriba, G. L. Taboada, R. Doallo, D. Posada, ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).
57. Z. Li *et al.*, First trans-eunicellane terpene synthase in bacteria. *Chem* **9**, 698–708 (2023).
58. J. M. Urban *et al.*, *Bradysia* (*Sciara*) *coprophila* larvae up-regulate DNA repair pathways and down-regulate developmental regulators in response to ionizing radiation. *Genetics* **226**, iyad208 (2024).
59. C. N. Hodson, K. S. Jaron, S. Gerbi, L. Ross, Gene-rich germline-restricted chromosomes in black-winged fungus gnats evolved through hybridization. *PLoS Biol.* **20**, e3001559 (2022).
60. A. Dobin *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
61. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323 (2011).
62. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
63. X. Chen *et al.*, Canonical terpene synthases in arthropods: intra-phylum gene transfer (2024). Figshare. https://doi.org/10.6084/m9.figshare.27248763.v33. Deposited 22 November 2024.
64. X. Chen *et al.*, Bradysia coprophila sesquiterpene synthase (TPS1) mRNA, complete cds. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/PQ349591. Deposited 17 September 2024.
65. X. Chen *et al.*, Bradysia coprophila mono/sesqui/diterpene synthase (TPS2) mRNA, complete cds. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/PQ349592. Deposited 17 September 2024.
66. X. Chen *et al.*, Bradysia coprophila mono/sesquiterpene synthase (TPS3) mRNA, complete cds. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/PQ349593. Deposited 17 September 2024.
67. X. Chen *et al.*, Bradysia coprophila mono/sesquiterpene synthase (TPS5) mRNA, complete cds. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/PQ349594. Deposited 17 September 2024.
68. X. Chen *et al.*, Dinothrombium tinctorium mono/sesquiterpene synthase (TPS1) mRNA, complete cds. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/PQ349595. Deposited 17 September 2024.
69. X. Chen *et al.*, Oppia nitens sesquiterpene synthase (TPS2) mRNA, complete cds. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/PQ349596. Deposited 17 September 2024.
70. X. Chen *et al.*, Oppiella nova mono/sesquiterpene synthase (TPS5) mRNA, complete cds. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/PQ349597. Deposited 17 September 2024.
71. X. Chen *et al.*, Folsomia candida mono/sesquiterpene synthase (TPS2) mRNA, complete cds. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/PQ349599. Deposited 17 September 2024.
72. X. Chen *et al.*, Folsomia candida mono/sesquiterpene synthase (TPS5) mRNA, complete cds. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/PQ349600. Deposited 17 September 2024.
73. X. Chen *et al.*, Allacma fusca mono/sesqui/diterpene synthase mRNA (TPS2), complete cds. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/PQ349601. Deposited 17 September 2024.
74. X. Chen *et al.*, NP-Card for Cattleyene (NP0019639). Natural Products Magnetic Resonance Database. https://np-mrd.org/natural_products/NP0019639. Deposited 19 September 2024.
75. X. Chen *et al.*, NP-Card for Phomopsene (NP0341814). Natural Products Magnetic Resonance Database. https://np-mrd.org/natural_products/NP0341814. Deposited 19 September 2024.