A novel diterpene synthase from the sandfly *Lutzomyia longipalpis* produces the pheromone sobralene

Charles Ducker[a], Cameron Baines[a], Jennifer Guy[a], Antônio Euzébio Goulart Santana[b], John A. Pickett[c]*, Neil J. Oldham[a]*

[a]School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, UK

[b]Center of Engineering and Agrarian Science, Federal University of Alagoas, Maceio, Brazil

[c]School of Chemistry, Cardiff University, Main Building, Park Pl, Cardiff, CF10 3AT, UK

* Corresponding authors:

Professor Neil J. Oldham, School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, UK.

neil.oldham@nottingham.ac.uk

+44 (0)115 9513542

Professor John A. Pickett, School of Chemistry, Cardiff University, Main Building, Park Pl, Cardiff, CF10 3AT, UK

pickettj4@cardiff.ac.uk

+44 (0)29 2251 0838

Abstract

The phlebotomine sandfly, *Lutzomyia longipalpis*, a major vector of the *Leishmania* parasite, uses terpene pheromones to attract conspecifics for mating. Examination of the *L. longipalpis* genome revealed a putative terpene synthase (TPS), which – upon heterologous expression in, and purification from, *E. coli* – yielded a functional enzyme. The TPS, termed *Ll*TPS, converted geranyl diphosphate (GPP) into a mixture of monoterpenes with low efficiency, of which β-ocimene was the major product. (*E,E*)-farnesyl diphosphate (FPP) principally produced small amounts of (*E*)-β - farnesene, whilst (*Z,E*)- and (*Z,Z*)-FPP yielded a mixture of bisabolene isomers. None of these mono- and sesquiterpenes are known volatiles of *L. longipalpis*. Notably, however, when provided with (*E,E,E*)-geranylgeranyl diphosphate (GGPP), *Ll*TPS gave sobralene as its major product. This diterpene pheromone is released by certain chemotypes of *L. longipalpis*, in particular those found in the Ceará state of Brazil. Minor diterpene-components were also seen as products of the enzyme that matched those seen in a sandfly pheromone extract.

Significance

Terpenes are widely used in nature for chemical communication, but our understanding of how these structurally diverse natural products are produced by insects is only now beginning to emerge. Males of the sandfly, *Lutzomyia longipalpis*, use terpene pheromones to lure females and other males to mating sites. This insect has attracted considerable attention due to its role as a vector for the *Leishmania* parasite, which causes the neglected tropical disease leishmaniasis. In this study a diterpene synthase that produces the pheromone component sobralene is identified, heterologously expressed and functionally characterized. This represents the first discovery of a terpene synthase (TPS) from *Lutzomyia*, and the first insect diterpene synthase. It offers the potential for sustainable production of this compound through biocatalysis.

Introduction

Terpenes, a class of isoprenoids, comprise a highly diverse group of natural products often possessing pronounced biological activity (1). Within the eukaryotes, plants and fungi are recognized as the major producers, but amoebae and insects also utilize terpenes for a range of ecological functions(2, 3). Terpenes are biosynthesized by two main routes: the mevalonate (MVA) pathway and the methylerythritol phosphate (MEP) pathway(4, 5), with both resulting in the production of dimethylallyl diphosphate (DMAPP, **1**) and isopentenyl diphosphate (IPP, **2**), which are the C_5 building blocks of terpene hydrocarbons and the wider class of isoprenoid metabolites (Fig. 1). It is believed that animals are only capable of employing the MVA pathway whilst plants, for example, can utilize both the MVA and MEP routes (6). DMAPP and IPP are assembled into chains of varying length by a class of enzymes known as isoprenyl diphosphate synthases (IDSs). DMAPP plus one unit of IPP produces geranyl diphosphate (GPP, C_{10}, **3**), DMAPP plus two units of IPP yields farnesyl diphosphate (FPP, C_{15}, **4**) and DMAPP plus three units of IPP gives geranylgeranyl diphosphate (GGPP, C_{20}, **5**) (Fig. 1). IDSs responsible for producing a specific length of isoprenyl chain are often named for their product, e.g., farnesyl diphosphate synthase (FPPS), geranylgeranyl diphosphate synthase (GGPPS) etc. Terpenes may then be produced by the action of terpene synthases (TPSs) upon isoprenyl diphosphates via removal of inorganic diphosphate (PPi) leading to an allyl cation intermediate (Fig. 1) (7, 8). Controlled rearrangement and/or elimination of a proton from this species can result in one

or several products from a potentially vast array of cyclic and acyclic structures, depending upon the environment within a particular TPS active site.

Fig. 1. The origin of terpene natural products. (A) Isoprenyl diphosphate synthesis from DMAPP (**1**) and IPP (**2**) catalyzed by IDS enzymes, including FPPS and GGPPS, and leading to FPP (**4**) and GGPP (**5**). (B) terpene synthesis via carbocation intermediates catalyzed by the action of TPS enzymes. (C) structures of the terpene pheromones of *L. longipalpis*, (*S*)-9-methylgermacrene-B (**6**), (1*S*,3*S*,7*R*)-3-methyl-α-himachalene (**7**) and sobralene (**8**).

Although many TPSs derived from plants and microorganisms have been identified, far fewer examples from insects are known. In a pioneering study, Beran *et al.* in this journal (2016) showed that insect TPSs appear to have evolved from IDS ancestors through gene duplication (9). Thus, in insects, TPSs are much closer to IDSs than is the case in plants, fungi or bacteria. Recently, Rebholtz *et al.* analyzed sequences of putative or confirmed IDS/TPS genes from 8 insect orders to produce a comprehensive phylogeny (10). From this work a tendency for a greater number of amino acid substitutions in IPP-binding motifs (IBMs) has been shown to be associated with the evolution from IDSs to TPSs, which appears to have occurred independently across the insects. Examination of the number of amino acid substitutions in IBMs is emerging as a useful tool in identifying candidate TPSs from insect genomic data.

The phlebotomine sandfly *Lutzomyia longipalpis* is a vector for the *Leishmania* parasite, the causative pathogen of the neglected tropical disease leishmaniasis (11). Although, it should be noted, that it is more correct to use the plural, leishmaniases, as this is a group of diseases with a range of symptoms. Cutaneous leishmaniasis is the most common form globally, and results in disfiguring skin ulcers, whereas visceral leishmaniasis, which targets the body's internal organs, is the most severe form, being invariably fatal if untreated (12, 13). Male sandflies produce sex/aggregation pheromones, which they use to attract both females and males to lek mating sites (14). There are several defined geographical chemotypes of *Lutzomyia longipalpis* in Brazil (15), which are associated with differing pheromone chemistry: those that produce the homosesquiterpenes (*S*)-9-methylgermacrene-B (**6**) or (1*S*,3*S*,7*R*)-3-methyl-α-himachalene (**7**), those that employ diterpenes, of which sobralene (**8**) is the only structurally characterized member, and those that utilize a mixture of these homosesquiterpene (characterized originally by Pickett) and diterpene components (16–21). Control of the sandfly is a recognized means of leishmaniasis prevention, and the hope is that pheromones may be used to assist mass trapping. Hamilton and colleagues have shown that (±)-9-methylgermacrene-B, produced by chemical means (22), is, indeed, effective for trapping *Lutzomyia longipalpis* in the field (23–26). Due to their structural complexity, there remain, however, significant barriers to the effective synthetic production of the three pheromone components for use in widespread insect control. This consideration, together with the interesting chemical structures of these terpenes, has motivated us to explore their biosynthesis in the insect and to identify TPS enzymes responsible for pheromone production.

Here we report the first identification, expression and functional characterization of a terpene synthase from *L. longipalpis* that is capable of producing the pheromone component sobralene (**8**) from (*E,E,E*)-GGPP. To the best of our knowledge, this is the first identification of a true diterpene synthase from an insect. It is hoped that these findings may offer a potential route to sustainable production of the isomerically pure compound.

3

Results

Identification of a putative terpene synthase from the *L. longipalpis* genome. Examination of the published *L. longipalpis* genome revealed the presence of several candidate IDS/TPS genes (27). Of these XP_055691875.1, which is annotated in the genome as a farnesyl diphosphate synthase-like gene, attracted our particular interest as, when compared to FPPSs it showed significant amino acid substitutions in the IPP-binding motifs (IBMs) characterized by Rebholtz *et al*. The *L. longipalpis* FPPS, *Ll*FPPS, displayed full conservation of all 15 IPP-interacting residues (IBMs 1-5 and SARM, Fig. 2). This protein is annotated as an FPPS in the published *L. longipalpis* genome and functionally confirmed as such in this study (Fig. S1 and S2). In contrast, XP_055691875.1 retained only 3 of the 15, which made it a strong TPS candidate. This result was comparable to a characterized sesquiterpene TPS from the flea beetle *Phyllotreta striolata*, *Ps*TPS1, which maintained 4 of the 15 conserved residues (Beran et al.). Additionally, the sequence of XP_055691875.1 showed high identity with an RNA contig, identified as C94, from a transcriptomic study of the *L. longipalpis* pheromone gland (28).

Fig. 2. Sequence analysis of insect IDS and TPS enzymes. (A) Muscle alignment of XP_055677521.1 (*Ll*FPPS, aa108-341) and XP_055691875.1 (*Ll*TPS, aa34-263), compared against the FPPS of *Drosophila melanogaster* (*Dm*FPPS, aa116-349) and TPS1 from *Phyllotreta striolata* (*Ps*TPS1, aa75-307). Underlined residues denote location of DDxxD motifs, with yellow highlighting predicted IPP-binding motifs (IBMs) in IDS enzymes, and red colour identifying conserved residues reported to make direct contacts with IPP. (B) Phylogenetic tree of selected insect IDS and TPS enzymes inferred by maximum-likelihood analysis.

Expression and functional testing of putative TPS. The gene for XP_055691875.1 was synthesized commercially in pET100/D-TOPO plasmid, and soluble protein successfully expressed and purified from *E. coli* as described in the Materials and Methods (Fig. S2). Native electrospray-mass spectrometry of the enzyme showed that it was homodimeric in structure (predicted monomer mass 43140 Da, measured mass 86365 Da, Fig. S2C). Although this is not a surprising result given the similarity between insect TPSs and FPPSs and the fact that FPPSs are known to be dimeric, the quaternary structure of insect TPSs has not been previously determined experimentally. The difference between the predicted polypeptide dimer mass (86280 Da) and the measured value was 85 Da, which was likely due to copurification of 3-4 Mg^{2+} ions with the protein. Activity of the putative TPS was assayed by incubation with a panel of isoprenyl diphosphates (Table 1) followed by GC-MS analysis of the resulting volatiles. It was clear from these results that the protein product of gene XP_055691875.1 was, indeed, a TPS and that it was capable of producing mono-, sesqui- and diterpenes from the appropriate precursors. For this reason, from this point in the paper the enzyme is referred to as *Ll*TPS. IDS and TPS enzymes typically use one or more Mg^{2+} ions as a cofactor, which both bind and activate the diphosphate moiety. Occasionally, other divalent cations are employed (29, 30), but – of the divalent metal ion cofactors tested – Mg^{2+} gave the greatest product yield. Co^{2+} and Mn^{2+} resulted in an approximate factor-10 and factor-100 decrease in terpene products, respectively. If no metal ions were added, no activity was seen. Each metal ion produced a broadly similar profile of products. GPP was converted to small quantities of myrcene, limonene, and (*E*)- and (*Z*)-β-ocimene as the only true terpenes, with linalool and geraniol also present (Table 1, Table S1, Fig. S3 and S4). (*E,E*)-FPP also yielded weak signals, the largest of which was identified as (*E*)-β-

farnesene (Table 1, Table S1, Fig. S5A, and Fig. S6). Minor peaks due to α-, β- and γ-bisabolenes were also seen. In contrast, (Z,E)- and (Z,Z)-FPP, the metabolically less familiar and generally less abundant isomers of FPP, gave significant quantities of bisabolenes. The major product from (Z,E)-FPP was (Z)-γ-bisabolene (ca. 50 % of the total), with β-bisabolene, (E)-α-bisabolene and (Z)-α-bisabolene constituting the remaining 50 % (Table 1, Table S1, Fig. S5B and Fig. S6). (Z,Z)-FPP yielded β-bisabolene as the major product (ca. 70 %) with minor amounts of (E)-α-bisabolene and (E)-γ-bisabolene (Table 1, Table S1, Fig. S5C and Fig. S6). Additionally, a GC peak with an EI mass spectrum very similar to that of helminthogermacrene was seen. No germacrene-B, γ-elemene or himachalenes (Fig. S9/10) were seen as products from any of the FPP isomers.

As described in the Introduction, the sesquiterpene pheromones **6** and **7** from *L. longipalpis* are, in fact, homosesquiterpenes, which are likely derived from (S)-8-methyl-FPP. This homologue of FPP possesses an additional methyl branch, which corresponds to the position and stereochemistry of the branches seen in **6** and **7**. To establish whether the presence of a methyl group on C8 of (E,E)- or (Z,E)-FPP caused *Ll*TPS to produce either the methylgermacrene-B (**6**) or the methylhimachalene (**7**), instead of the bisabolenes, the (E,E)- and (Z,E)- isomers of 8-methyl-FPP were generated as substrates for *Ll*TPS in coupled enzyme assays. (±)-(E,E)-8-methyl-FPP was produced *in situ* by incubation of (±)-4-methylGPP with IPP in the presence of *Ll*FPPS, the *L. longipalpis trans*-FPP synthase (Fig. S8), before addition of *Ll*TPS and incubation for a further 16 hrs. (±)-(Z,E)-8-methyl-FPP was similarly produced using the *cis*-FPPS *Ps*IDS3 in place of *Ll*FPPS before addition of *Ll*TPS. *Ps*IDS3 is a known (Z,E)-FPP synthase from the flea beetle *Phyllotreta striolata*. GC-MS analysis of the pentane extracts from both incubations did reveal the presence of homosesquiterpenes (Fig. S9/10), but the EI spectra of these products did not match those of either (S)-9-methylgermacrene-B (**6**) (Fig. S9) or (1S,3S,7R)-3-methyl-α-himachalene (**7**). It is likely that they are homobisabolenes, by analogy with the products from the standard isoprenyl pyrophosphates (E,E)- or (Z,E)-FPP.

***Ll*TPS produces the diterpene pheromone sobralene from (E,E,E)-GGPP**. Incubation of *Ll*TPS with (E,E,E)-GGPP yielded a range of diterpene products (Fig. 3A and Fig. S11). The major component exhibited an EI mass spectrum identical to that of sobralene (**8**), a known pheromone component of certain *L. longipalpis* chemotypes. Comparison of the GC retention time and EI mass spectrum with that seen for authentic sobralene from an extract of male *L. longipalpis* collected near Ico, Ceará state, Brazil confirmed the identification (Fig. 3B and Fig. S11). Sandflies from this region, which is approximately 400 km from Sobral, had previously been shown to produce sobralene (Fig. S12). Other product peaks present in the GC-MS chromatogram of the *Ll*TPS assay included one with an MS very close to that of cembrene A (**9**, also known as neocembrene) and two with spectra very similar to those of verticillene isomers (**10**, **11**) (31), as well as two unknown diterpenes (Fig. S11). These compounds were also seen as minor products in the male sandfly extract (Fig. 3B).

Fig. 3. Diterpene synthase activity of *Ll*TPS. (A) GC-MS chromatogram of the diterpene products resulting from incubation of (E,E,E)-GGPP with *Ll*TPS. (B) GC-MS chromatogram of an extract of *L. longipalpis* males collected near Ico, Ceará state, Brazil. (C) EI mass spectrum of sobralene peak from chromatogram (A). (D) EI mass spectrum of authentic sobralene peak from chromatogram (B).

Discussion

***L. longipalpis* possesses an IDS-derived TPS.** It has long been known that insects employ a wide variety of terpenes and terpenoids as semiochemicals, whether that be as pheromones, for

communication with conspecifics, or as allomones, kairomones and synomones to mediate interspecific interactions (3). What has emerged only relatively recently is the understanding that insects are themselves capable of biosynthesizing terpenes without the direct input of plants or microorganisms. Identification of a dual function IDS/TPS in bark beetles of the genus *Ips* was the first major insight into terpene biosynthesis in insects (32), but it was the study by Beran *et al.* on IDSs and TPSs from the flea beetle *Phyllotreta striolata* that provided a detailed understanding of the evolutionary relationship between these two classes of enzymes (9). Males of this species of flea beetle, which feeds on the Brassicales, release *γ*-cadinene and himachalene-type sesquiterpenoids that serve as aggregation pheromones. This behavior leads to mass infestation in the field, with serious damage to commercially important crops, such as cabbage and mustard. Two active IDSs and four active TPSs were identified in *P. striolata* and, although the two enzyme types separated into two different clades, the exon-intron structures of the IDS and TPS genes were conserved suggesting that the TPSs had evolved from the IDSs. Further evidence for this proposal was provided by Rebholz *et al.* in their large-scale bioinformatic analysis of insect IDS/TPS enzymes (10). Despite the small number of functionally characterized insect TPS sequences available, the authors were able to identify that disruptive amino acid substitutions in isopentenyl diphosphate-binding motifs (IBMs) were characteristic of TPSs, whilst IDS maintained essential consensus residues in these regions. They concluded that TPSs had emerged multiple times and independently across the insects. The identification here of *Ll*TPS from *L. longipalpis* as a TPS is consistent with these ideas and here we show multiple, 12 out of 15, amino acid substitutions in the IBM, when compared to the FPPS from this species, which conserved all 15 IPP-interacting residues.

***Ll*TPS is a functional terpene synthase.** *Ll*TPS produced monoterpenes, sesquiterpenes and diterpenes when provided with the appropriate isoprenyl diphosphate substrates *in vitro*. GPP is not a particularly abundant isoprenylPP metabolite in the context of most cellular environments, except when monoterpenes are actively being synthesized, and so it is likely that its role as a substrate is not biologically relevant in sandflies, which are not reported to release monoterpenes. The production of monoterpenoid alcohols by *Ll*TPS is most likely due to the small size of GPP, which can allow water molecules to enter readily the enzyme active site leading to quenching of the carbocation intermediate. Turning to sesquiterpenes, the production of (*E*)-β-farnesene from (*E,E*)-FPP, albeit with relatively low efficiency, is an interesting observation. There are no reports of this sesquiterpene as a *Lutzomyia* volatile, but it is a known feeding stimulant for the insect (33). Given this fact and the ubiquitous nature of (*E,E*)-FPP, it is quite possible that (*E*)-β-farnesene is released from the male sandfly pheromone gland, perhaps at levels below the current limit of detection.

L. longipalpis is a species complex, with at least eight lineages (34). Populations may be characterized into chemotypes by their pheromone chemistry. The published *L. longipalpis* genome is based on the Jacobina chemotype (27), named for that city in the Bahia state of Brazil. This population characteristically produces (1*S*,3*S*,7*R*)-3-methyl-α-himachalene (**7**) as the major component of its pheromone, together with a small amount of the regular sesquiterpene α-himachalene(18). Thus, in mining genome data of the Jacobina chemotype, it was reasonable to expect to find a TPS that produces (homo)sesquiterpenes of this structural family. Experiments performed on the *Ll*TPS reported here showed no evidence for the production of α-himachalene from the three FPP isomers used. As described above, (*E*)-β-farnesene was detected from (*E,E*)-FPP. (*Z,E*)- and (*Z,Z*)-FPP, in contrast, gave bisabolene isomers in significant quantity. These monocyclic sesquiterpenes have not been reported in any *L. longipalpis* chemotype, and we suspect that this result may be artefactual, and brought about by exposure of the enzyme to uncommon (*Z,E*)- and (*Z,Z*)-isomers of FPP. That said, the flea beetle *P. striolata* possesses a *cis*-IDS and uses it to produce (*Z,E*)-FPP and from this (6*R*,7*S*)-himachala-9,11-diene (9). *Ll*TPS did not exhibit analogous activity,

however. It is noteworthy that a (+)-(*S,Z*)-α-bisabolene synthase has been found in insects, namely in the green stink bug *Nezara viridula* (*Nv*TPS1 in Figure 1B), but this uses (*E,E*)-FPP as a substrate to produce the bisabolene and not (*Z,E*)- or (*Z,Z*)-FPP (35, 36). The distinctly different product profiles seen between (*E,E*)- and (*Z,E*)-FPP indicates that rotation of the 2,3 bond in the carbocation intermediate is slow relative to the deprotonation step (Fig. S13) resulting in only minor amounts of bisabolenes seen with (*E,E*)-FPP and no (*E*)-β-farnesene seen with (*Z,E*)-FPP.

Most IDSs and TPSs characterized in the literature use Mg^{2+} as a cofactor, but it has been shown that some IDSs are sensitive to changes in divalent metal ions and can switch from FPP to GPP synthase activity upon addition of Co^{2+} or Mn^{2+} (29, 30). We examined if inclusion of either of these metals in an *Ll*TPS assay would affect the product profile, but no changes in the pattern of terpenes were seen, and Mg^{2+} was the most efficient cofactor by one to two orders of magnitude over Co^{2+} and Mn^{2+}, respectively.

Given that the major pheromone component in the Jacobina chemotype is methylhimachalene **7**, a homosesquiterpene (19), we speculated that *Ll*TPS may be capable of producing this from (*E,E*)- or (*Z,E*)-8-methylFPP (Fig. S8), even though such activity was not seen with the standard FPPs. Incubation of *Ll*TPS with either (*E,E*)- or (*Z,E*)-8-methylFPP gave homosesquiterpene products with quite different mass spectra (Fig. S9/10), from those reported for the methylhimachalene **7**, or indeed the methylgermacrene **6**, demonstrating that inclusion of the additional carbon did not shift the product distribution to the known sandfly homosesquiterpenes. The male sandflies collected near Ico, Ceará state, Brazil produced 9-methylgermacrene-B (Fig. S10) in addition to sobralene and so we postulate that either there was a missing essential component from our assay with 8-methylFPP and *Ll*TPS, which prevented production of the methylgermacrene, or – as seems more likely – this *L. longipalpis* chemotype has an additional TPS that produces the homosesquiterpene.

***Ll*TPS produces the diterpene pheromone sobralene from (*E,E,E*)-GGPP.** Taking the above results into account and recognizing that *Ll*TPS produces sobralene (**8**) as its major product from (*E,E,E*)-GGPP leads us to the proposal that the Jacobina chemotype harbors a gene for synthesizing this diterpene pheromone, even though it is not known to produce the compound. The exact status of the *L. longipalpis* species complex is a matter of some debate, and – as pointed out above – recent work has shown that at least eight linages exist, which may represent different species(37). Our finding of a sobralene TPS in the Jacobina chemotype may help to inform this discussion, and in the future we intend to identify other TPSs, including homosesquiterpene synthases, in the insect. Taking together the functional FPPS and TPS reported here with the GGPPS we previously described (38), *L. longipalpis* enzymes responsible for the biosynthesis of sobralene (**8**) from DMAPP (**1**) and IPP (**2**) are now elucidated. Gene expression analysis of TPSs in a range of chemotypes could provide more insights into how the different terpene profiles are tuned. In males of *L. longipalpis* the pheromone is released from a specialized exocrine gland located in the fourth abdominal segment of the insect (39). It is not clear whether *de novo* production of the terpenes takes place in these tissues, although genes associated with the MVA pathway have been found to be transcribed there (28). In the same study, an RNA contig (numbered C94), possessing high identity to *Ll*TPS, was seen in extracts of the gland. Taking these pieces of evidence together, it is likely that the pheromone gland is, indeed, the site of terpene biosynthesis.

In comparison with mono- and sesquiterpenes, true diterpenes appear relatively rare in insects. The simple acyclic diterpene springene is found in some ants, bees and wasps (3). Cembrene A, also known as neocembrene, (**9**) is a trail pheromone in termites of the genus *Nasutitermes* (40), is the queen pheromone of the pharaoh's ant *Monomorium pharonis* (41), and has been identified in

secretions of the ant *Monomorium chinense* (42). In addition to sobralene (**8**), other diterpenes have been found in chemotypes of the *L. longipalpis* complex, either as minor components, or as major products. With the exception of taxadiene (**12**), which is present in minor amounts in the extract analyzed by Palframan *et al.*(20), these are still to be structurally characterized. The products of *Ll*TPS included diterpenes with EI mass spectra identical to cembrene A (**9**) and verticillene (**10**, **11**) isomers (31), as well as two unknown diterpenes.

Intriguingly, these compounds were also present as minor components in extracts of *L. longipalpis* males found near Ico, Ceará state, Brazil; indeed, the qualitative match between the two GC chromatograms is striking (Fig. 3 A and B). Quantitatively, the pheromone extract contained proportionally more sobralene and lower amounts of the minor components than the enzyme assays. It has been shown that sobralene can be converted into the two observed verticillene isomers by treatment with acid, or long-term storage of a sample of sobralene in deuterochloroform (mildly acidic conditions) (43). Thus, although the enzyme assay and sandfly pheromone extract studied here were not directly exposed to acid, it is possible that the presence of these minor compounds is an artifact of the assay conditions, sample preparation and/or analysis. That said, enzyme assays were buffered at pH 7.2 and analyzed by GC-MS immediately following incubation. Moreover, passing the pentane fraction through a small silica gel column prior to analysis, or reducing the GC injection port temperature from 200 °C to 150 °C had no effect on the quantity of minor diterpene products seen. Examination of the probable mechanism of (*E,E,E*)-GGPP cyclization shows the relationship between the diterpene components seen (Fig. 4), which is broadly in-line with the proposals of Palframan and Pattenden (44). Following loss of the diphosphate group, the initial allyl cation can undergo ring closure to form a macrocycle. Loss of a proton may then lead to formation of an isopentenyl group and the product cembrene A (**9**). Alternatively, a second ring closure can occur to give the veticillenyl cation. Direct proton loss from this intermediate leads to the verticillenes (**10**, **11**), whilst proton migration from C11 to C7 (either directly or indirectly, via C3) followed by proton loss from C9 yields sobralene (**8**). Interestingly, taxadiene (**12**) was not detected as a product from *Ll*TPS, despite its close relationship with sobralene (**8**). This suggests that, in the *Ll*TPS active site, adoption of the sobralene-like conformation of the intermediate carbocation, and subsequent proton elimination to install the *Z*-8,9 double bond of **8**, is preferred over the taxadiene-like conformation and its accompanying ring closure, following the arguments presented by Hayes *et al.* (45).

Fig. 4. Scheme showing the proposed *Ll*TPS-catalyzed route to sobralene (**8**) from (*E,E,E*)-GGPP and its relationship with cembrene A (**9**), verticillenes (**10**, **11**) and taxadiene (**12**). Alternative routes from common intermediates are shown by red or green arrows. Note: H$^+$ transfer from C11 to C7 may occur directly or via C3, and that transannular ring closure and H$^+$ loss to form taxadiene is a two-step process.

Prior to this study, there were no reports of insect diterpene synthases in the literature, so it is pleasing to confirm that insects are also capable of making this class of terpenes without reliance upon the enzymatic machinery of plants or microorganisms. Currently, there is no synthetic chemical route to sobralene (**8**), and it is unlikely that sustainable and commercially viable production will be achieved by this means. Having identified a sobralene producing enzyme, we are now exploring its potential as a viable source of the pheromone for vector control.

Materials and Methods

Phylogenetic analyses. FPPS-like homologues were identified from the *L. longipalpis* genome (Jacobina, NCBI accession PRJNA20279) using blastp searches with TPS enzymes from *P. striolata* as the query. Hits were assessed for the presence of IPP binding motifs (IBMs) indicative of IDS activity, leading to the characterisation of XP_055677521.1 as a likely FPPS (*Ll*FPPS) and XP_055691875.1 as a likely TPS (*Ll*TPS). These sequences were aligned with various functionally characterized insect IDS and TPS enzymes using MUSCLE. MEGA11 was used to construct the maximum likelihood tree (LG+G+I) and phylogeny was tested by bootstrap analysis (1000 replicates). Sequences were as follows: *Dm*FPPS (NP_477380.1), *Dm*GGPPS (AAC05273.1), *Ll*GGPPS (XP_055688734.1), *Aa*FPPS (XP_001663796.1), *Ps*IDS1 (ALL35400.1), *Ps*IDS3 (ALL35406.1), *Ps*TPS1 (ALL35411.1), *Ps*TPS2 (ALL35414.1), *Ps*TPS3 (ALL35417.1), *Ps*TPS4 (ALL35420.1), *Ip*GPPS/TPS (Q58GE8.1), *Nz*TPS1 (CAH1405470.1), *Nz*TPS2 (WBV93233.1), *Mh*TPS (A0A343W969.1), *Hh*TPS1 (QBK50746.1), *Hh*TPS2 (QBA82488.1), *Mo*TPS (A0A7D0AGU9.1), *Hm*TPS1 (HmelOS) and *Hm*TPS2 (HMEL037108g1).

Protein expression and purification. Assessment of protein sequences for N-terminal localization sequences (TargetP-2.0 (46) and iPSORT (47)) revealed that *Ll*FPPS was predicted to contain mitochondrial targeting motifs. Constructs encoding full length *Ll*TPS (aa 1-340) and truncated *Ll*FPPS (aa 72-411) for recombinant protein expression in bacteria (pET100/D-TOPO, N-terminally His-tagged) were synthesized by GeneArt (Thermo Fisher Scientific, Loughborough, UK). These were transformed into BL21(DE3)pLysS *E. coli* and 20ml LB cultures were grown from a single colony (100 µg/mL ampicillin, 34 µg/mL chloramphenicol). Following 16 h at 37 $^{\circ}$C with shaking (200 rpm), these were added to 1 L media and incubated at 37 $^{\circ}$C until reaching OD_{600} 0.5, at which point IPTG was added to 0.75 mM and cultures were incubated at either 25 $^{\circ}$C for 5 hrs (*Ll*TPS) or 37°C for 3 hrs (*Ll*FPPS). Cells were harvested by centrifugation (4000 rpm, 15 min, 4 $^{\circ}$C) and stored at -80 $^{\circ}$C.

For purification, cell pellets were lysed in 30 mL wash buffer (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 10 % v/v glycerol, 20 mM imidazole, 1 mM DTT) supplemented with 1 v/v Igepal CA-630 and EDTA-free cOmplete protease inhibitor (Roche, Welwyn Garden City, UK) and incubated at 4 $^{\circ}$C for 15 min. Cell extracts were then sonicated (12 x 10 s on ice) and the insoluble fraction pelleted by centrifugation (14000 rpm, 20 min, 4 $^{\circ}$C). Nickel-NTA agarose beads (0.5 mL bed) were added to the soluble fraction and rotated (90 min, 4 $^{\circ}$C), before washing (3 x 10 mL wash buffer) and elution (3 mL wash buffer plus 400 mM imidazole). Proteins were buffer exchanged into protein storage buffer (25 mM MOPSO pH 7.2, 100 mM NaCl, 10% v/v glycerol, 1 mM DTT) using PD-10 columns (Cytiva, Little Chalfont, UK), with aliquots snap frozen ($N_{2(l)}$) and stored at -80 $^{\circ}$C. Expression and purification of *Ps*IDS3 (aa 31-384) was performed with the same methodology (37 $^{\circ}$C for 3 hrs post-induction).

Protein Mass Spectrometry. For assessment of tryptic peptides, protein aliquots (*Ll*TPS and *Ll*FPPS) were denatured with 5x Laemmli SDS sample buffer (Alfa Aesar, Thermo Fisher Scientific, Loughborough, UK) and incubation at 95°C for 5 minutes. Protein was loaded onto a 12% tris-glycine gel (BioRad, Watford, UK) and resolved by SDS-PAGE (160 V, 400 mA, 50 minutes). The gel was stained with GelCode™ Blue Safe Protein Stain (Thermo Fisher Scientific, Loughborough, UK) as per the manufacturer's instructions. Bands corresponding to the target proteins were excised, reduced (10 mM dithiothreitol) and alkylated (55 mM iodoacetamide) followed by addition of aqueous trypsin (0.5 µg, Promega, Southampton, UK) and incubation at 37°C for 18 hrs. Digestion was quenched by the addition of formic acid (1 µL, Honeywell, Fisher Scientific, Loughborough, UK), the supernatant centrifuged and subject to high-performance liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis.

Digests were analyzed using an Ultimate 3000 RSLCnano HPLC system fitted with a PepMap 300 C18 column (15 cm x 150 µm, 5 µm particle size, Thermo Fisher Scientific, Loughborough, UK) coupled to a Thermo LTQ Ultra Hybrid mass spectrometer (linear ion trap/FTICR MS) via a nano-ESI source equipped with a fused silica emitter tip (20 µm ID, MSWIL, Aarle-Rixtel, The Netherlands). A dual mobile phase gradient was used composed of 5:95:0.1 (A) and 95:5:0.1 (B) water:acetonitrile:formic acid with a flow rate of 400 µL per minute. MS analysis was performed in data-dependent acquisition mode whereby the three most abundant ions in the full scan were subject to collision induced dissociation fragmentation. MS/MS data was analysed using SearchGUI. In brief, RAW files were first converted to the compatible MZ5 using MSConvert (48), then subject to a search using X!Tandem (49) against a UniProt database (1,117,184 total sequences) followed by visualisation in Peptide Shaker (50). Identified peptides were mapped to the protein sequence (Fig. S2D and E) using MS Tools (51).

For assessment of protein native mass, *Ll*TPS protein was concentrated approximately two-fold using a Vivaspin centrifuge column (30K MWCO, Sartorius, Goettingen, Germany) and buffer exchanged into 50 mM ammonium acetate. Native mass spectrometry analysis was performed on a Waters Synapt G1 Quadrupole Time-of-Flight MS operating in positive ion mode. Samples were sprayed from pulled glass emitter tips (0.3 µm internal diameter) back filled with platinum wire. A 1.1 kV potential was applied to the tip with 0.1 mL per minute nitrogen for desolvation and minimal energies (25 V) applied to the trap and transfer cells.

IDS activity assays. IDS activity was assessed by combining recombinant protein (2 µM) with allylic substrate (50 µM DMAPP/100 µM 4-methyl GPP) and 100 µM IPP in assay buffer (25 mM MOPSO pH 7.2, 10 mM $MgCl_2$, final volume 100 µl) and incubating for 1 hr at 30 °C. DMAPP (38426) was purchased from Merck (Gillingham, UK) and IPP (I-0050) from Echelon Biosciences (Salt Lake City, Utah, USA).

For assessment by reverse phase LC-MS, 20 mM sodium diphosphate was added to samples and 5 µl was injected onto a Waters Acquity Ultra Performance Liquid Chromatography system with a BEH C18 column (1.7 µM, 2.1 mm × 50 mm; Waters Corporation, Wilmslow, UK) coupled to a Waters Synapt G1 Quadrupole Time-of-Flight MS operating in negative electrospray ionisation mode. Mobile phase flow rate was maintained at 0.3 ml/minute, consisting of a binary gradient of 5 mM ammonium bicarbonate (solvent A) and acetonitrile (solvent B). Gradient parameters were as follows: starting at 10% v/v B, increasing to 65% B over 11.00 minutes, increasing to 95% B by 11.10 (0.9-minute hold to 12.00), before returning to 10% B by 12.10 (held until 14.00). Capillary voltage was -2.3kV, with a desolvation temperature of 500°C and gas flow rate of 600L/hr, and FPP was identified by its deprotonated ion (EIC *m/z* 381.2).

To analyse products by gas chromatography-mass spectrometry (GC-MS), enzyme assays were set up and incubated as described above, followed by addition of 20 U shrimp alkaline phosphatase (SAP, New England Biolabs, Ipswich, USA) and incubation for 4 hrs at 37 °C. Cyclohexane (200 µl) was added and assays were left overnight, before mixing the aqueous and organic layers, centrifugation (3000xg for 2 min) and discarding the aqueous layer. The organic layer was evaporated to ~50 µl under a stream of nitrogen and 5 µl was injected onto an Agilent-Jeol GC-MS (see below for details). Product identities were confirmed by comparison with MS spectra and retention times of known standards.

TPS activity assays. TPS activity was assessed by combining 2 µM protein with allylic substrate [50 µM GPP, (*E,E*)-FPP, (*Z,E*)-FPP, (*Z,Z*)-FPP or (*E,E,E*)-GGPP] in assay buffer (final volume 100 µl, 10 mM $MgCl_2$ was replaced with an equal concentration of either $MnCl_2$ or $CoCl_2$ in experiments designed to

test the efficacy of those ions), addition of 200 µl pentane and incubation at 30 $^{\circ}$C overnight. GPP (G6772), (*E,E*)-FPP (44722) and *(E,E,E)*-GGPP (G6025) were purchased from Merck (Gillingham, UK) and (*Z,E*)-FPP (I-0180) and (*Z,Z*)-FPP (I-0170) from Echelon Biosciences (Salt Lake City, Utah, USA). The aqueous and organic layers were mixed with a pipette, separated by centrifugation (3000xg for 2 min) and the aqueous layer discarded. The organic layer was evaporated to ~50 µl under a stream of nitrogen and assessed by GC-MS using either the Agilent-Jeol GC-MS or the Thermo ISQ 7000 GC-MS (see below for details). Product identities were confirmed by comparison with MS spectra and retention times of known standard compounds or pheromone extracts, where available. Where identification relied upon library spectrum fit only (i.e. helminthogermacrene and verticillenes-1 and -2) it is listed as tentative. IDS-TPS coupled assays were carried out similarly, but with incubation of 2 µM IDS protein with 100 µM 4-methyl GPP and 100 µM IPP for 1 hr at 30 $^{\circ}$C, before addition of 2 µM TPS enzyme and further incubation overnight overlaid with pentane.

Synthesis of (±)-4-methylgeranyl diphosphate. (±)-4-methylgeraniol was purchased from Enamine (Kyiv, Ukraine). ^{1}H NMR (400 MHz, CDCl$_3$) δ 5.44 (t, $^{3}J_{H,H}$ = 6.8 Hz, 1H), δ 5.07 (t, $^{3}J_{H,H}$ = 7.2 Hz, 1H), 4.19 (d, $^{3}J_{H,H}$ = 6.7 Hz, 2H), 2.14 (m, 1H), 2.05 (m, 2H), 1.71 (s, 3H), 1.64 (s, 3H), 1.62 (s, 3H), 1.01 (d, $^{3}J_{H,H}$ = 6.6 Hz, 3H); ^{13}C NMR (101 MHz, CDCl$_3$) δ 144.0, 132.0, 123.6, 123.1, 59.5, 42.9, 33.5, 25.8, 19.0, 17.8, 13.4 (Fig. S14). (±)-4-methylgeranyl diphosphate was synthesized from the alcohol via the tosylate following a literature method (52) to yield 20 mg of product. ^{1}H NMR (500 MHz, D$_2$O) δ 5.38 (t, $^{3}J_{H,H}$ = 6.7 Hz, 1H), δ 5.09 (t, $^{3}J_{H,H}$ = 7.1 Hz, 1H), 4.40 (m, 2H), 2.13 (q, 7.0 Hz, 1H), 2.01 (m, 2H), 1.60 (s, 3H), 1.68 (s, 3H), 1.53 (s, 3H), 0.92 (d, $^{3}J_{H,H}$ = 6.8 Hz, 3H); ^{13}C NMR (125 MHz, D$_2$O) δ 142.3, 133.9, 123.3, 119.3, 62.5, 42.4, 32.8, 24.9, 18.4, 17.1, 12.7 (Fig. S15); ^{31}P NMR (162 MHz, D$_2$O) δ -6.74 (d, $^{2}J_{P,P}$ = 21.7 Hz), -10.43 (d, $^{2}J_{P,P}$ = 21.3 Hz) (Fig. S16).

Gas chromatography-mass spectrometry. GC-MS was performed on either an Agilent 7890B GC equipped with a DB-5ms column (30m x 0.25 mm x 0.25 µm, Agilent, Cheadle, UK) coupled to a Jeol AccuTOF GCx system (Welwyn Garden City, UK), or a Thermo ISQ 7000 GC-MS system equipped with a Trace Gold TG17-MS column (30m x 0.25 mm x 0.25 µM, Thermo Fisher Scientific, Loughborough, UK). For both instruments, the inlet temperature was set to either 150 $^{\circ}$C or 200 $^{\circ}$C, operated in splitless mode. For the Agilent-Jeol GC-MS, the GC oven temperature was held at 35 °C for 4 min, then an increase of 10 °C/min to 300 °C followed by a 6 min hold, and the helium carrier gas flow rate was 1 ml/min. The MS was operated in positive electron ionization mode and the mass range scanned from m/z 40-500. For the Themo GC-MS the oven program started with a 3 min isothermal hold at 35 °C, followed by an increase of 10 °C/min to 260 °C and a 5 min hold. The helium carrier gas flow rate was 1.5 ml/min. The MS was operated in positive electron ionization mode and the mass range scanned from m/z 40-450. Both mass spectrometers operated with and 8 min solvent delay, reduced to 5 min for monoterpene analysis. Kovats retention indices were estimated using an alkane C$_8$-C$_{20}$ standard (Merck, Gillingham, UK). Authentic standards/extracts were as follows: E-β-farnesene (provided by Pickett), geraniol (Merck, Gillingham, UK), bisabolene (mix of isomers, Thermo Fisher Scientific, Loughborough, UK), cedarwood Himalayan oil (himachalenes, Nikura, London, UK), *L. longipalpis* pheromone extract ((*S*)-9-methylgermacrene-B, sobralene, see below).

Insect collection. *L. longipalpis* flies were collected in the Alfa settlement (latitude -6.424452570, longitude -38.9214688), near Ico, Ceará, Brazil during the period from 30.05.2023 to 01.06.2023. Traps, of the CDC type, were placed near farms with chickens and goats from 5:00 pm to 6:00 am each night. In total, 577 sandfly specimens were captured (381 males and 196 females). Collected insects were placed in vials containing hexane (HPLC grade, Merck, Darmstadt, Germany) using 2 mL of solvent for males and 1 mL for females. Both samples were filtered and the hexane solution stored at -20 °C.

Data sharing statement

Mass spectrometry data created during this research are openly available from the University of Nottingham data repository at DOI: 10.17639/nott.7364.
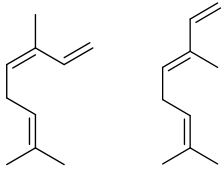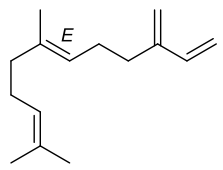
Acknowledgements

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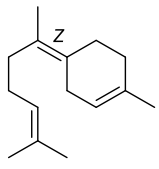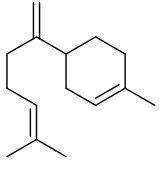
1. J. Gershenzon, N. Dudareva, The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **3**, 408–414 (2007).

2. 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).

3. E. David Morgan, *Biosynthesis in Insects, Advanced Edition* (Royal Society of Chemistry, 2010).

4. M. Rohmer, M. Knani, P. Simonin, B. Sutter, H. Sahm, Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochem. J.* **295**, 517–524 (1993).

5. J. Gershenzon, R. B. Croteau, "Terpenoid Biosynthesis: The Basic Pathway and Formation of Monoterpenes, Sesquiterpenes, and Diterpenes" in *Lipid Metabolism in Plants*, T. S. Moore, Ed. (CRC Press, 1993), pp. 339–388.

6. E. Vranová, D. Coman, W. Gruissem, Network Analysis of the MVA and MEP Pathways for Isoprenoid Synthesis. *Annu. Rev. Plant Biol.* **64**, 665–700 (2013).

7. D. Tholl, Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr. Opin. Plant Biol.* **9**, 297–304 (2006).

8. 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. *The Plant J.* **66**, 212–229 (2011).

9. F. Beran, *et al.*, Novel family of terpene synthases evolved from *trans* -isoprenyl diphosphate synthases in a flea beetle. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2922–2927 (2016).

10. Z. Rebholz, *et al.*, Emergence of terpene chemical communication in insects: Evolutionary recruitment of isoprenoid metabolism. *Protein Sci.* **32** (2023).

11. R. Lainson, R. Ward, J. Shaw, Experimental transmission of *Leishmania chagasi*, causative agent of neotropical visceral leishmaniasis, by the sandfly *Lutzomyia longipalpis*. *Nature* **266**, 628–630 (1977).

12.   H. J. C. de Vries, H. D. Schallig, Cutaneous Leishmaniasis: A 2022 Updated Narrative Review into Diagnosis and Management Developments. *Am. J. Clin. Dermatol.* **23**, 823–840 (2022).

13.   S. Scarpini, *et al.*, Visceral Leishmaniasis: Epidemiology, Diagnosis, and Treatment Regimens in Different Geographical Areas with a Focus on Pediatrics. *Microorganisms* **10**, 1887 (2022).

14.   J. G. C. Hamilton, "Chapter 13: Sand fly sex/aggregation pheromones" in *Sensory Ecology of Disease Vectors*, (Brill | Wageningen Academic, 2022), pp. 349–371.

15.   C. N. Spiegel, *et al.*, The Lutzomyia longipalpis complex: a brief natural history of aggregation-sex pheromone communication. *Parasit. Vectors* **9**, 580 (2016).

16.   J. G. C. Hamilton, G. W. Dawson, J. A. Pickett, 9-Methylgermacrene-B; proposed structure for novel homosesquiterpene from the sex pheromone glands of *Lutzomyia longipalpis* (Diptera: Psychodidae) from Lapinha, Brazil. *J. Chem. Ecol.* **22**, 1477–1491 (1996).

17.   J. G. C. Hamilton, H. C. Ibbotson, A. M. Hooper, J. A. Pickett, 9-Methylgermacrene-B is confirmed as the sex pheromone of the sandfly *Lutzomyia longipalpis* from Lapinha, Brazil, and the absolute stereochemistry defined as S. *Chem. Commun.* 2335–2336 (1999).

18.   J. G. C. Hamilton, G. W. Dawson, J. A. Pickett, 3-Methyl-α-himachalene: Proposed structure for novel homosesquiterpene sex pheromone of *Lutzomyia longipalpis* (Diptera: Psychodidae) from Jacobina, Brazil. *J. Chem. Ecol.* **22**, 2331–2340 (1996).

19.   J. G. C. Hamilton, A. M. Hooper, J. A. Pickett, K. Mori, S. Sano, 3-Methyl-α-himachalene is confirmed, and the relative stereochemistry defined, by synthesis as the sex pheromone of the sandfly *Lutzomyia longipalpis* from Jacobina, Brazil. *Chem. Commun.* 355–356 (1999).

20.   M. J. Palframan, K. K. Bandi, J. G. C. Hamilton, G. Pattenden, Sobralene, a new sex-aggregation pheromone and likely shunt metabolite of the taxadiene synthase cascade, produced by a member of the sand fly *Lutzomyia longipalpis* species complex. *Tetrahedron Lett.* **59**, 1921–1923 (2018).

21.   J. G. C. Hamilton, R. D. C. Maingon, B. Alexander, R. D. Ward, R. P. Brazil, Analysis of the sex pheromone extract of individual male *Lutzomyia longipalpis* sandflies from six regions in Brazil. *Med. Vet. Entomol.* **19**, 480–488 (2005).

22.   A. M. Hooper, J.-B. Farcet, N. P. Mulholland, J. A. Pickett, Synthesis of 9-methylgermacrene B, racemate of the sex pheromone of *Lutzomyia longipalpis* (Lapinha), from the renewable resource, Geranium macrorrhizum essential oil. *Green Chem.* **8**, 513 (2006).

23.   D. P. Bray, K. K. Bandi, R. P. Brazil, A. G. Oliveira, J. G. C. Hamilton, Synthetic Sex Pheromone Attracts the Leishmaniasis Vector *Lutzomyia longipalpis* (Diptera: Psychodidae) to Traps in the Field. *J. Med. Entomol.* **46**, 428–434 (2009).

24.   M. A. González, *et al.*, Synthetic sex-aggregation pheromone of *Lutzomyia longipalpis*, the South American sand fly vector of *Leishmania infantum*, attracts males and females over long-distance. *PLoS Negl. Trop. Dis.* **14**, e0008798 (2020).

25.   D. P. Bray, *et al.*, Synthetic Sex Pheromone in a Long-Lasting Lure Attracts the Visceral Leishmaniasis Vector, *Lutzomyia longipalpis*, for up to 12 Weeks in Brazil. *PLoS Negl. Trop. Dis.* **8**, e2723 (2014).

26. R. Gonçalves, *et al.*, Community deployment of a synthetic pheromone of the sand fly Lutzomyia longipalpis co-located with insecticide reduces vector abundance in treated and neighbouring untreated houses: Implications for control of *Leishmania infantum*. *PLoS Negl. Trop. Dis.* **15**, e0009080 (2021).

27. F. Labbé, *et al.*, Genomic analysis of two phlebotomine sand fly vectors of Leishmania from the New and Old World. *PLoS Negl. Trop. Dis.* **17**, e0010862 (2023).

28. N. González-Caballero, J. G. Valenzuela, J. M. Ribeiro, P. Cuervo, R. P. Brazil, Transcriptome exploration of the sex pheromone gland of *Lutzomyia longipalpis* (Diptera: Psychodidae: Phlebotominae). *Parasit. Vectors* **6**, 56 (2013).

29. S. Frick, *et al.*, Metal ions control product specificity of isoprenyl diphosphate synthases in the insect terpenoid pathway. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4194–4199 (2013).

30. F. Ecker, A. Vattekkatte, W. Boland, M. Groll, Metal-dependent enzyme symmetry guides the biosynthetic flux of terpene precursors. *Nat. Chem.* **15**, 1188–1195 (2023).

31. P. Schrepfer, *et al.*, Identification of amino acid networks governing catalysis in the closed complex of class I terpene synthases. *Proc. Natl. Acad. Sci. U.S.A.* **113** E958-E967 (2016).

32. A. B. Gilg, C. Tittiger, G. J. Blomquist, Unique animal prenyltransferase with monoterpene synthase activity. *Naturwissenschaft.* **96**, 731–735 (2009).

33. R. B. Tesh, H. Guzman, M. L. Wilson, *Trans*-beta-farnesene as a Feeding Stimulant for the Sand Fly *Lutzomyia longipalpis* (Diptera: Psychodidae). *J. Med. Entomol.* **29**, 226–231 (1992).

34. M. A. C. Gutierrez, *et al.*, DNA barcoding of *Lutzomyia longipalpis* species complex (Diptera: Psychodidae), suggests the existence of 8 candidate species. *Acta. Trop.* **221**, 105983 (2021).

35. Z. Rebholz, *et al.*, Ancient origin and conserved gene function in terpene pheromone and defense evolution of stink bugs and hemipteran insects. *Insect Biochem. Mol. Biol.* **152**, 103879 (2023).

36. J. Lancaster, *et al.*, An IDS-Type Sesquiterpene Synthase Produces the Pheromone Precursor (Z)-α-Bisabolene in *Nezara viridula*. *J. Chem. Ecol.* **45**, 187–197 (2019).

37. M. A. C. Gutierrez, *et al.*, DNA barcoding of *Lutzomyia longipalpis* species complex (Diptera: Psychodidae), suggests the existence of 8 candidate species. *Acta. Trop.* **221**, 105983 (2021).

38. C. Ducker, *et al.*, Characterisation of geranylgeranyl diphosphate synthase from the sandfly *Lutzomyia longipalpis*. *Insect Biochem. Mol. Biol.* **161**, 104001 (2023).

39. R. P. Lane, D. de Souza Bernardes, Histology and ultrastructure of pheromone secreting glands in males of the phlebotomine sandfly *Lutzomyia longipalpis*. *Ann. Trop. Med. Parasitol.* **84**, 53–61 (1990).

40. A. J. Birch, W. V. Brown, J. E. T. Corrie, B. P. Moore, Neocembrene-A, a termite trail pheromone. *J. Chem. Soc. Perkin 1*, 2653 (1972).

41. J. P. Edwards, J. Chambers, Identification and source of a queen-specific chemical in the pharaoh's ant, *Monomorium pharaonis* (L.). *J. Chem. Ecol.* **10**, 1731–1747 (1984).

42. R. Zhao, L. Lu, Q. Shi, J. Chen, Y. He, Volatile Terpenes and Terpenoids from Workers and Queens of *Monomorium chinense* (Hymenoptera: Formicidae). *Molecules* **23**, 2838 (2018).

43.   M. J. Palframan, K. K. Bandi, J. G. C. Hamilton, G. Pattenden, Acid-Catalysed Rearrangement of the Sandfly Pheromone Sobralene to Verticillenes, Consolidating its Relationship inter alia to the Taxanes and Phomactins. *Synlett.* **30**, 1899–1903 (2019).

44.   M. J. Palframan, G. Pattenden, The verticillenes. Pivotal intermediates in the biosynthesis of the taxanes and the phomactins. *Nat. Prod. Rep.* **36**, 108–121 (2019).

45.   C. J. Hayes, M. J. Palframan, G. Pattenden, The Impact of Macrocycle Conformation on the Taxadiene-Forming Carbocation Cascade: Insight Gained from Sobralene, a Recently Discovered Verticillene Isomer. *J. Org. Chem.* **85**, 4507–4514 (2020).

46.   J. J. Almagro Armenteros, *et al.*, Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* **2**, e201900429 (2019).

47.   H. Bannai, Y. Tamada, O. Maruyama, K. Nakai, S. Miyano, Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* **18**, 298–305 (2002).

48.   M. C. Chambers, *et al.*, A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).

49.   R. Craig, R. C. Beavis, TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467 (2004).

50.   Y. M. Farag, C. Horro, M. Vaudel, H. Barsnes, PeptideShaker Online: A User-Friendly Web-Based Framework for the Identification of Mass Spectrometry-Based Proteomics Data. *J. Proteome Res.* **20**, 5419–5423 (2021).

51.   D. Kavan, P. Man, MSTools—Web based application for visualization and presentation of HXMS data. *Int. J. Mass Spectrom.* **302**, 53–58 (2011).

52.   A. Hou, L. Lauterbach, J. S. Dickschat, Enzymatic Synthesis of Methylated Terpene Analogues Using the Plasticity of Bacterial Terpene Synthases. *Chem. Eur. J.* **26**, 2178–2182 (2020).

Table 1. Terpene synthase (TPS) activity of *Ll*TPS from *L. longipalpis* when incubated with a panel of isoprenyl diphosphate substrates in the presence of Mg^{2+}. The major product from each substrate is indicated in bold and the structure shown.

Substrate	Terpene product	
Geranyl diphosphate	β-Myrcene Limonene **(Z)-/(E)-β-Ocimene** Linalool Geraniol	
(*E,E*)-Farnesyl diphosphate	**(E)-β-Farnesene** β- Bisabolene (Z)-γ-Bisabolene (E)-α-Bisabolene	
(*Z,E*)-Farnesyl diphosphate	β- Bisabolene (Z)-α-Bisabolene **(Z)-γ-Bisabolene** (E)-α-Bisabolene	
(*Z,Z*)-Farnesyl diphosphate	**β- Bisabolene** Helminthogermacrene? (E)-α-Bisabolene (E)-γ-Bisabolene	
(*E,E,E*)-Geranylgeranyl diphosphate	Diterpene 1 Cembrene A **Sobralene** Verticillene-1 Verticillene-2 Diterpene 2	

A)

DMAPP
1

IPP
2

IDS -HOPP

GPP
3

IDS + **2** -HOPP

FPP
4

IDS + **2** -HOPP

GGPP
5

FPPS

GGPPS

B)

TPS

- OPP

-H⊕

Terpenes

C)

9-Methylgermacrene-B
6

3-Methyl-α-himachalene
7

Sobralene
8

A)

```
PsTPS1   VLMDYTVTHDDPYFLSSAVLPLYFYKAVEESDKLTEENIKRACLMSWAYRTLETSQIIVD
LlTPS    RMLRYHNE--NIF---GGILYLDLFKALAPPSRQTDEYYQLVVYLDLIVHMGNIALVFLD
LlFPPS   RVLQYNAQGGKKN---RGVTTAQTFELIAAKSDLTEENRRLAIYLGWCVELLQALFLLMD
DmFPPS   QVLQYNVPRGKKN---RGILTVLTYKNLVPTQDLTPENIKLAQYLGWCVEMLQSFFIISD
          ::  *    .    .:      ::  .   * *   *   :.      :    ::  *

PsTPS1   DILDKSEVRYNKPAWYKKDGVSMELTILDSHYLATGAYMVLTKRLAGHPCCLDILDLYAE
LlTPS    DLLDGDDIRFKKTCWYKIVGYGA--AANDAVIIVSLFIVLKNHFSHLDCYSKLSEVLAE
LlFPPS   DIMDGSTTRRGQLCWYKQEDVQLS-AVNDGLMIEAGIYQILKKHFSHLDCYVKLLELFHE
DmFPPS   DVMDNSTTRRGQPCWHKVENVGLT-AINDALMIENAMYAILKKHFSHLDCYVALMELFHE
         *::* .   *   :.*:*   .      :   *.  : . : :*.:.::    *    : ::  *

PsTPS1   EMFVMIIAQYMD-------IKKLDLKDFQKLVRHRFDKALYVFNGSARSGLYLANVRDRE
LlTPS    SYFHISLGQQLDTKFASASFDDFTEENYDILLMSK---SSCTFAIIIRGILCLAGVEDLT
LlFPPS   MTFITTMGQHLDTRSANVDILNFTMELYKTIVLNK--TAWYSFYMPVAAAMHLAGYKDDE
DmFPPS   ITYITTCGQSLDQLNSNRCVSEFTMENYKAIVENK--TAYYSFYLPFALALHLAGYKDAE
          :    .* :*      ..  :: :.  ::  .   :    *         : **.   *

PsTPS1   THDCMKKFSVPMSRFFQVQNDFSGVFEEESKFQNSCPDIVNGRNSWLVTTALKMANPAQR
LlTPS    VIKKATDFFNDLGRLAQSQNDFWDCFDYMGSEVKTGTDISTRKCSWLSVQCIKLATPQQR
LlFPPS   AFRQARTILLEIGYFFQIQDDFLDCYGDPEHTGKIGTDIQDGKCTWLAVVFLQRATAAQK
DmFPPS   AFRQSKTILLEMGNFFQVQDDFLDCFGNPEVTGKIGTDIQDNKCSWLAVVAMQRANVEQK
         .      :    :. : * *:**  . :        :   .**   . :**  . ::  *.  *.
```

B)

```
              ┌── DmFPPS
          81 ─┤
         99   ├── AaFPPS
              └── LlFPPS        FPPS
      99 ─── PsIDS1
      62 ─── PsIDS3
   61 ─── IpGPPS/TPS
   59 ─── LlTPS
                  ┌── NzTPS1
             100 ─┤
                  └── HhTPS2
        100 ─────── MhTPS       FPPS-like TPS
                  ┌── NzTPS2
             100 ─┤
              83  └── HhTPS1
             ┌── PsTPS1
        81 ──┤
        94   ├── PsTPS2
             ├── PsTPS3
        100 ─┴── PsTPS4
                  ┌── HmTPS1
             99 ──┤
                  └── HmTPS2    GGPPS-like TPS
        100 ─────── MoTPS
             99 ──┬── LlGGPPS   GGPPS
             90    └── DmGGPPS

0.50
```

(E,E,E)-GGPP **5**

Cembrene A **9**

Cyclization

C11-H to C7 transfer

Sobralene **8**

Verticillene **10**

Verticillene **11**

Taxadiene synthase

Taxadiene **12**