Manuscript

Click here to access/download;Manuscript;Horacek et al. Main Text.docx

# Evolutionary origin of terpenoid biosynthesis in termites

Natan Horáček[1,2], Ondřej Lukšan[1], Zarley Rebholz[3], Karel Harant[1], Radek Pohl[1], Lana Mutabdžija-Nedelcheva[1,4], Simon Hellemans[5], Daniel Jungwirth[1,6], Jan Křivánek[1], Anna Amirianová[1,7], Pavlína Kyjaková[1], Thomas Bourguignon[5], Dorothea Tholl[8], Robert Hanus[1,*], Jitka Štáfková[1,*]

[1] Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague, Czech Republic

[2] Department of Analytical Chemistry, Faculty of Science, Charles University, Prague, Czech Republic

[3] Department of Biology, Indiana University, Bloomington, IN, USA

[4] Department of Genetics and Microbiology, Faculty of Science, Charles University, Prague, Czech Republic

[5] Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan

[6] Department of Biochemistry, Faculty of Science, Charles University, Prague, Czech Republic

[7] Department of Biochemistry and Microbiology, University of Chemistry and Technology Prague, Prague, Czech Republic

[8] Department of Biological Sciences, Virginia Tech, Blacksburg, VA, USA

* Corresponding authors: jitka.stafkova@uochb.cas.cz, robert.hanus@uochb.cas.cz

**Abstract**

Termites produce the most diverse array of terpenoids among metazoans, comprising over 200 structures. However, their biosynthesis has not yet been elucidated. Here, we identify a gene family which arose through the duplication of geranylgeranyl pyrophosphate synthase in the common ancestor of Neoisoptera, the terpene-producing termite lineage. We functionally characterized several proteins from this family as terpene synthases generating biologically relevant sesqui- and diterpenes. These include the queen pheromone (3*R*,6*E*)-nerolidol in *Embiratermes neotenicus* and the precursor of polycyclic defensive diterpenes (*E*,*E*,*E*)-neocembrene in *Nasutitermes takasagoensis*. We explore transposable element-mediated genomic mechanisms and selection pressures acting in the evolution of this gene family and report an amino acid site crucial for cyclization capacity as well as the enantiospecificity of the characterized enzymes. We conclude that we have identified an enzyme family underlying the remarkable richness of termite terpenoids, which likely contributed to the ecological success of Neoisoptera.

37 **1. INTRODUCTION**

38 Termites hold the record among insects for the diversity of terpenoid secondary metabolites they
39 produce, with over 200 different monoterpenes, sesquiterpenes and diterpenes identified to date
40 [1,2]. The occurrence of terpenoids is restricted to the modern termite clade Neoisoptera and
41 coincides with the emergence of the frontal gland, a unique exocrine organ and a synapomorphy of
42 Neoisoptera. This gland is present in dispersing kings and queens, but most developed—and best
43 studied—in soldiers, the caste specialized in colony defense [3,4].

44 The supposed ancestral function of termite terpenoids was to serve as defensive compounds, released
45 in rich blends through the opening of the frontal gland on termite heads. Some of the defensive
46 terpenoids were later co-opted for chemical communication, in line with the observations of
47 pheromone evolution from defensive compounds in other insects [5]. Thus, terpenoids frequently act
48 as alarm pheromones secreted from the frontal gland of irritated soldiers [6,7], but may also be
49 synthesized in other exocrine organs and serve as sex pheromones and trail pheromones [8]. Finally,
50 terpenoids may play the role of queen pheromones produced by reigning queens to signal their
51 presence in the colony to the nestmates [9]. In contrast to the solid coverage of termite terpenoid
52 diversity, their biosynthesis and its evolution remain to be elucidated. Because terpenoids have been
53 recorded in the basal neoisopteran family Stylotermitidae [10] as well as in many other later diverging
54 lineages, it is reasonable to hypothesize that their biosynthesis has a single origin in the common
55 ancestor of Neoisoptera.

56 Terpenoid secondary metabolites usually originate from prenyl pyrophosphates, such as geranyl
57 pyrophosphate ($C_{10}$, GPP), farnesyl pyrophosphate ($C_{15}$, FPP), or geranylgeranyl pyrophosphate ($C_{20}$,
58 GGPP) giving rise to monoterpenes, sesquiterpenes, and diterpenes, respectively. The conversion of
59 prenyl pyrophosphates into terpenes is catalyzed by terpene synthases (TPS), which cleave off the
60 pyrophosphate and yield linear or cyclic terpene hydrocarbons, or the respective alcohols. Final
61 terpenoid products may then arise through further modifications (oxygenations, cyclizations) by
62 additional enzymes [11,12].

63 Like other animals, insects generally lack the "classical" TPS genes known from plants, bacteria, and
64 fungi [13] and rely on alternative mechanisms to obtain terpenoids. These include sequestering them
65 from the diet, acquiring them from microbial symbionts, using classical TPS genes gained through
66 horizontal gene transfer [14] or recruiting a completely unrelated enzyme family for TPS functions [15].
67 Nevertheless, the most common evolutionary route by which insects acquired the capacity for *de novo*
68 terpenoid biosynthesis is through duplication and neofunctionalization of isoprenyl diphosphate
69 synthase genes (IDS). This enzyme family is conserved in the primary metabolism across all domains of
70 life and catalyzes the elongation of prenyl pyrophosphates via the sequential condensation of shorter
71 pyrophosphates with isopentenyl pyrophosphate ($C_5$, IPP) [16,17]. Such origin of TPSs has so far been
72 documented in a handful of species across Coleoptera, Lepidoptera, Hemiptera, and Diptera, where
73 TPSs are paralogous either to farnesyl pyrophosphate synthases (FPPS [18–23]) or geranylgeranyl
74 pyrophosphate synthases (GGPPS [24,25]). Taken together, these findings indicate that TPSs originated
75 independently in multiple insect lineages through repeated IDS duplications and neofunctionalizations.
76 Here, we aim to elucidate the evolution of terpenoid biosynthesis in termites. Building on reports of
77 independent origins of TPSs from IDSs in several insect clades, as well as on the exclusive occurrence
78 of terpenoids in Neoisoptera, we hypothesize that terpenoid biosynthesis evolved in the common
79 ancestor of Neoisoptera through duplication/multiplication and neofunctionalization of IDSs. To
80 capture the structural and functional diversity of termite terpenoids, we selected two model species
81 from different subfamilies within Neoisoptera. Our primary model is the South American termite
82 *Embiratermes neotenicus* (Termitidae: Syntermitinae), in which we recently identified the
83 sesquiterpene alcohol (3*R*,6*E*)-nerolidol as the queen pheromone [9]. The second model is the
84 Japanese species *Nasutitermes takasagoensis* (Termitidae: Nasutitermitinae) belonging to a genus
85 whose soldiers produce a rich blend of monoterpenes and polycyclic oxygenated diterpenes derived
86 from the monocyclic precursor (3*E*,7*E*,11*E*)-neocembrene [1,2,26]. Previously, eight GGPPS homologs
87 have been reported as present and upregulated in the frontal gland of soldiers in this species. Although

88 GGPPS gene multiplicity has been tentatively linked to terpenoid production in termites [27,28], the
89 hypothesis has not been experimentally verified.
90 We report that GGPPS exists as a single-copy gene in basal termites, but its duplication or
91 multiplication is a ubiquitous feature of Neoisoptera, reaching up to fifteen paralogs in *E. neotenicus*.
92 This pervasive multiplicity strongly indicates that the initial duplication event occurred concurrently
93 with the frontal gland emergence in their common ancestor. We functionally characterize selected
94 GGPPS homologs from the two species and demonstrate that some of them act as TPS enzymes,
95 converting prenyl pyrophosphate substrates into terpenoids known from the respective species, or
96 biosynthetic precursors thereof. We conclude that we have identified a family of TPS enzymes
97 responsible for the remarkable diversity of terpenoid secondary metabolites in termites. Furthermore,
98 we bring evidence for possible involvement of transposable elements in shaping the GGPPS-like gene
99 family and outline how selection pressures may have shaped its evolutionary diversification. Finally,
100 we examine the structure-function relationship of termite TPS enzymes and identify an amino acid
101 crucial for their capacity of terpene cyclization and enantioselective biosynthesis.
102
103
104 **2. RESULTS**
105 **2.1 Terpenoid chemistry in *E. neotenicus* queens and *N. takasagoensis* soldiers**
106 To obtain a comprehensive image of the terpenoids produced by the two model termites, we first
107 analyzed body washes from selected castes of both species using gas chromatography coupled with
108 mass spectrometry (GC-MS). In *E. neotenicus*, substantial amounts of terpenes were only present in
109 neotenic queens; the dominant compound was (6*E*)-nerolidol, followed by (3*E*,7*E*,11*E*)-neocembrene,
110 along with minor or trace quantities of α- and β-farnesene, β-springene and two additional
111 unidentified diterpenes (Fig. 1A, Supplementary Fig. S1–S2).
112 In soldiers of *N. takasagoensis*, we detected a rich and abundant blend of terpenes comprising two
113 monoterpenes—(+)-α-pinene and (+)-limonene—and eight diterpenoids (Fig. 1B; Supplementary Figs.

114 S3–S5). Three of the most prominent diterpenoids, each bearing three or four oxygenated functional
115 groups, were fully characterized using nuclear magnetic resonance (NMR) as 14α-acetoxy-3β,12β-
116 dipropionoxy-kampa-6(7),8(9)-diene, 3α,9β,13α-tripropionoxy-trinervita-11(12),15(17)-diene, and
117 3α,9β,13α-tripropionoxy-11α(12β)-epoxy-trinervita-15(17)-ene (Supplementary Figs. S6–S8),
118 following termite diterpene nomenclature [29].

119

120 **Fig. 1. Terpenoid chemistry of the two model termite species**. **A.** Comprehensive gas chromatography
121 (GC×GC) analysis showing terpenes identified in the body washes of neotenic queens of *E. neotenicus*
122 and compared to 4$^{th}$ stage female nymphs. Chromatograms of other castes are provided in
123 Supplementary Fig. S1, mass spectra of the identified terpenes in Supplementary Fig. S2. **B.** GC
124 chromatogram of terpenes identified in body washes of *N. takasagoensis* soldiers. Comparison with a
125 chromatogram of worker body wash is shown in Supplementary Fig. S3, mass and NMR spectra of the
126 detected terpenes are shown in Supplementary Figs S4–S8. All chromatograms (A and B) show the
127 responses to diagnostic ions for mono-, sesqui- and diterpenes (*m/z* = 68 + 69 + 93 + 204 + 272).
128

129 **2.2 An FPPS gene duplication predates the occurrence of termites**

130 We did not retrieve any sequences similar to plant, fungal, or bacterial TPSs in the transcriptome
131 assemblies of *E. neotenicus* and *N. takasagoensis*, nor in the *E. neotenicus* genome, which we
132 assembled for this study (genome assembly statistics are provided in Supplementary Table S1). No
133 duplication was found in genes coding for subunits of decaprenyl diphosphate synthase, a long-chain
134 IDS. We then queried the datasets for short-chain IDS homologs. Our search confirmed the presence
135 of three previously reported FPPS sequences in *N. takasagoensis* transcriptome [27] and revealed an
136 FPPS tandem duplication in *E. neotenicus* genome. A phylogenetic analysis of FPPS homologs from
137 selected cockroach and termite species suggested that an FPPS gene duplication is present in all
138 Blattodea (Supplementary Fig. S9A). An inspection of amino acid sequence alignments showed that
139 the more conserved blattodean FPPS clade, containing one and two sequences of *E. neotenicus* and *N.*
140 *takasagoensis*, respectively, presents all main features typical of the ancestral FPPS enzymes, such as
141 the so-called first and second aspartate-rich motifs (FARM and SARM, sequence DDxxD) implicated in
142 binding of magnesium ions in the active site of the enzyme [30, 31]. We therefore designated this clade

143 as canonical FPPSs. By contrast, in the divergent clade which we refer to as FPPS-like, FARM and SARM
144 are disrupted (Supplementary Fig. S10). The distinction between FPPS and FPPS-like proteins was
145 confirmed by *in vitro* assays with recombinant EneoB from *E. neotenicus*, a member of the FPPS clade,
146 which presented activities typical of insect FPPS proteins. No IDS or TPS activity was detected with the
147 FPPS-like protein from the same species, EneoA (Supplementary Fig. S9B). The combination of
148 biochemical assays and phylogenetic analysis let us conclude that the FPPS-like homologs are not
149 relevant for terpenoid biosynthesis in termites.
150

151 **2.3 Neoisoptera possess an expanded family of GGPPS-like genes**

152 Next, we searched the *E. neotenicus* transcriptome and genome for GGPPS homologs and identified
153 multiple full-length GGPPS sequences. Interestingly, we also discovered numerous partial GGPPS
154 homologs lacking one or more exons and located in genomic regions extensively soft-masked by
155 RepeatMasker. An inspection of the unmasked genome revealed as many as 28 highly homologous
156 sequences with conserved gene organization (Fig. 2A–B, Supplementary Fig. S11), most of which were
157 expressed in one or more *E. neotenicus* castes (Fig. 2C). Based on gene integrity (six protein-coding
158 exons) and the presence of indels or nonsense mutations, we classified 15 of these GGPPS homologs
159 as genes and 13 as pseudogenes (Fig. 2A–B, Supplementary Fig. S11).
160 We then reconstructed the GGPPS nucleotide phylogeny in the context of publicly available sequences
161 from other insects, including two basal termites and three Neoisoptera (*Stylotermes halumicus*,
162 *E. neotenicus* and *N. takasagoensis*). GGPPS occurs as a single-copy gene in basal termites and other
163 insects, whereas multiple paralogs were found in Neoisoptera, including *S. halumicus* from the very
164 basal neoisopteran family Stylotermitidae (Fig. 2D). Neoisoptera possess one paralog clustering with
165 the single-copy genes of basal termites (denoted as clade I), presumably representing the canonical
166 GGPPS enzyme. The remaining paralogs form an expanded monophylum, further referred to as the
167 GGPPS-like gene family, comprising three sequences from *S. halumicus*, six from *N. takasagoensis* and

168 fourteen from *E. neotenicus*. The GGPPS-like clade resolves into three branches, denoted II–IV (Fig. 2B–
169 D).
170
171 **Fig. 2. Diversity, gene structure, expression and phylogeny of GGPPS homologs in termites. A.** GGPPs
172 homologs identified in the genome of *E. neotenicus* and their gene structure. Genes are marked with
173 black arrows, pseudogenes with grey arrows, regions masked by RepeatMasker at genome-wide scale
174 highlighted in pink. Sequences corresponding to CACTA TIR TE_00000081 element are represented by
175 green arrows also showing the TE orientation. **B.** Exon structure of GGPPS homologs. Alternative 5'
176 exons in EneoH and pgENC3949.2 located upstream of the canonical exon 1 are marked as "-1". **C.**
177 Expression profiles in *E. neotenicus* $4^{th}$ stage female nymphs, queens, workers, and soldiers determined
178 by mRNA-seq based on TPM normalized read counts (sum of counts for head and thorax + abdomen
179 samples). **D.** Phylogenetic tree reconstructed from protein-coding nucleotide sequences showing the
180 GGPPS-like gene family expansion in Neoisoptera. The tree was generated using the Neighbor-Joining
181 method with 10,000 bootstrap replicates. Neoisopteran gene abbreviations are set in bold. Accession
182 numbers of the studied sequences are provided in Supplementary Tables S3–S4.
183
184 We further verified the hypothesis that the GGPPS multiplication is specific to Neoisoptera by
185 phylogenetic analysis of GGPPS amino acid sequences including 7 basal species and 19 Neoisoptera
186 (Supplementary Fig. S12, Supplementary Table S5). The presumed canonical GGPPS branches
187 (paraphyletic clade I) comprised sequences from cockroaches, both basal termites, and Neoisoptera,
188 while the divergent GGPPS-like branch consisted exclusively of neoisopteran sequences and formed a
189 monophylum sister to neoisopteran GGPPSs. All Neoisoptera species were represented in the GGPPS-
190 like branch, subdivided into well-supported clades II–IV. Thus, the initial GGPPS duplication occurred
191 in the ancestor of Neoisoptera and was followed by subsequent duplication events.
192 Finally, we surveyed all publicly available sequences homologous to GGPPS across termites, including
193 partial TSA transcripts, and mapped them on a phylogenetic tree of termites together with the records
194 of terpenoid secondary metabolites in Neoisoptera (Fig. 3, Supplementary Table S6). We obtained a
195 clear picture of GGPPS evolution: while only GGPPS could be retrieved from the basal lineages
196 represented by 8 species from three families, at least some partial GGPPS-like sequences were
197 identified in all 46 included Neoisoptera from seven families and 12 subfamilies of Termitidae.
198

199 **Fig. 3. Evolution of terpenoid biosynthesis in termites.** Production of mono-, sesqui- and diterpenes,
200 and their secondary derivatives (containing complex polycycles and oxygenated moieties), presence of
201 GGPPS and GGPPS-like sequences, mapped on termite phylogeny of non-termitid families and termitid
202 subfamilies. Phylogeny adopted and simplified from Hellemans et al. [100], Hodotermopsidae and
203 Serritermitidae are excluded from the scheme due to the absence of relevant sequence data. Positive
204 records of terpenes denote previous identifications of any terpenoid secondary metabolite in at least
205 one species from the given (sub)family, based on [1,2], and the recent report of terpenes in
206 Stylotermitidae as the most basal Neoisoptera [10]. The presence or absence of GGPPS and GGPPS-
207 like sequences in individual termite species was determined based on GGPPS phylogeny and/or the
208 presence of motifs previously identified as characteristic for insect GGPPSs. Numbers show the
209 sampling size in each (sub)family. The termite species and relevant datasets are listed in
210 Supplementary Table S6.

211

212 **2.4 GGPPS and GGPPS-like genomic loci are rich in transposable elements (TE)**

213 In *E. neotenicus*, the observed extensive duplication in GGPPS and GGPPS-like genes occurs alongside
214 an increased frequency of low-complexity regions in the given loci, strongly indicating the involvement
215 of transposition events. As all the paralogs retained an unspliced gene structure with fully conserved
216 exon-intron boundaries, we focused on class II DNA TEs—elements that transpose directly as DNA
217 without splicing [33]. Upon running RepeatModeler at the whole-genome level, we observed multiple
218 DNA TE-like motifs in loci surrounding the GGPPS-like genes and pseudogenes, consistent with the
219 increased occurrence of masked regions. To improve prediction accuracy, we restricted our analysis to
220 27 contigs harbouring sequences which showed >80% homology to previously identified GGPPS-like
221 orthologs and at least two consecutive exons. An EDTA pipeline annotation revealed several TEs, one
222 of which—TE_00000081, classified as class II transposon from the CACTA TIR family—was found in all
223 but two loci. This TE was oriented in the opposite direction relative to the gene and was located in the
224 region surrounding exon 6 (Fig. 2A, Supplementary Fig. S13). Different genomic patterns were
225 observed in the FPPS locus which did not show any enrichment of low complexity regions or repeats.
226 No fragments of FPPS-like genes could be found elsewhere in the genome.

227

228 **2.5 Identification of proteins with GGPPS activities in *E. neotenicus* and *N. takasagoensis***

229 First, we functionally characterized the proteins from clade I expected to act as GGPPS enzymes (EneoJ
230 in *E. neotenicus* and NtGGPPS7 in *N. takasagoensis*). Indeed, both proteins showed IDS activity with
231 IPP in combination with all allylic substrates tested: they produced GGPP from IPP and FPP; IPP and
232 GPP generated both FPP and GGPP; upon incubation with IPP and DMAPP, only GPP was detected
233 (Supplementary Fig. S14). No TPS activity was observed in EneoJ or NtGGPPS7 for any of the substrates
234 tested (GPP, FPP, GGPP).

235

236 **2.6 Two *E. neotenicus* GGPPS-like proteins show TPS activity and generate queen terpenoids**

237 Among the fourteen *E. neotenicus* GGPPS-like paralogs, we selected EneoC and EneoE, highly
238 expressed in neotenic queens according to RNA-Seq analysis shown in Fig. 2C. Upon incubation of
239 EneoC with FPP, a single product was generated, identified by GC-MS and chiral GC as pure (3*R*,6*E*)-
240 nerolidol (Fig. 4A). Additionally, EneoC presented IDS activity by generating FPP upon incubation with
241 IPP and GPP (Fig. 4C). No other substrates were accepted by EneoC.

242 EneoE accepted FPP in TPS activity assays, generating racemic (6*E*)-nerolidol, and converted GGPP into
243 (3*E*,7*E*,11*E*)-neocembrene (Fig. 4B). The enzyme showed no IDS activity. Both (3*R*,6*E*)-nerolidol and
244 (3*E*,7*E*,11*E*)-neocembrene represent queen-specific terpenoids in *E. neotenicus* (Fig. 1A), the former
245 being known as *E. neotenicus* queen pheromone [9]. Thus, we consider EneoC an enantiospecific
246 nerolidol synthase and EneoE a neocembrene synthase.

247 Having identified EneoC as capable of producing *E. neotenicus* queen pheromone, we wanted to
248 confirm its upregulation in queens compared to nymphs. Indeed, in differential gene expression
249 analysis of RNA sequencing data from nymphs and queens, EneoC was by far the most upregulated of
250 all GGPPS-like genes in queen samples from thorax and abdomen ($\log_2$ fold change queen vs.
251 nymph = 9.9, padj = $1.8\times10^{-16}$) while also showing the highest transcript abundances (Supplementary
252 Fig. S15, Supplementary Table S7). Importantly, EneoC transcript upregulation was clearly reflected by
253 proteomic data. Among all detected IDS homologs, EneoC was present as the most abundant protein

254 and showed the strongest enrichment in queen abdomen and thorax compared to nymphs ($log_2$ fold
255 change queen vs. nymph = 10.6; Supplementary Fig. S15, Supplementary Table S8). In summary, both
256 transcriptomic and proteomic analyses supported the biological relevance of EneoC for queen
257 pheromone biosynthesis.
258
259 **Fig. 4. Functional characterization of GGPPS-like enzymes from *E. neotenicus* and *N. takasagoensis*.**
260 **A.** GC chromatograms of SPME samples obtained from the headspace upon EneoC, EneoE and
261 NtGGPPS6 incubation with FPP. Left chromatograms show separation on a semi-nonpolar column,
262 right chromatograms depict chiral separation of (6*E*)-nerolidol on a chiral cyclodextrin. Enzymatic
263 assays shown in yellow, controls without enzymes in black, standards of (6*E*)-nerolidol enantiomers in
264 grey. **B.** GC chromatograms of SPME samples obtained from the headspace upon EneoC, EneoE and
265 NtGGPPS6 incubation with GGPP. Enzymatic assays shown in green, controls without enzymes in black.
266 **C.** HPLC chromatogram demonstrating the IDS activity of EneoC. The purified enzyme was incubated
267 with IPP in combination with DMAPP, GPP, or FPP. Only the substrate combination giving rise to prenyl
268 pyrophosphate product (FPP) is shown (blue), along with standards (black). The chromatogram
269 visualizes the selected m/z 245.00, 313.06, 381.12, and 449.19.
270
271 **2.7 *N. takasagoensis* GGPPS-like protein NtGGPPS6 is a neocembrene synthase**
272 All six GGPPS-like genes were reported as upregulated in *N. takasagoensis* soldier heads, the
273 biosynthetic site of terpenoid defensive compounds [27]. We expressed these in the *S. cerevisiae* strain
274 JWY501 which is specifically engineered to accumulate GGPP. In the NtGGPPS6-expressing strain, we
275 detected high amounts of (3*E*,7*E*,11*E*)-neocembrene, while no significant terpenoid production was
276 observed in any other strain (Supplementary Fig. S16). Subsequently, we expressed the six GGPPS-like
277 genes in *E. coli* and used the purified proteins in IDS and TPS assays. Only recombinant NtGGPPS6
278 showed any activity *in vitro*, producing a mixture of (3*R*,6*E*)- and (3*S*,6*E*)-nerolidol enantiomers from
279 FPP (Fig. 4A), and (3*E*,7*E*,11*E*)-neocembrene from GGPP (Fig. 4B). No IDS activity was observed. As only
280 monoterpenes and diterpenes are present in the soldier extract, we assume that *in vivo*, NtGGPPS6
281 does not generate nerolidol, and acts instead as a neocembrene synthase producing the precursor of
282 polycyclic diterpenes.
283

284 **2.8 Sequence motifs within termite GGPPs gene family and selection analysis**

285 We examined whether the divergence of GGPPS-like genes involves substitutions within motifs typical
286 of GGPPS proteins. We defined such insect GGPPS motifs (IGMs) based on patterns generated for
287 canonical insect GGPPS sequences by MEME, and we combined them with the IPP-binding motif
288 prediction used in our earlier study [34]. We observed that only FARM and IGM 3 motifs are conserved
289 in termite GGPPS-like proteins (Supplementary Fig. S17).

290 To understand how selection pressures may have influenced the evolution of TPS function in
291 Neoisoptera, we performed selection analysis on a phylogeny of selected full-length termite GGPPS
292 and GGPPS-like coding sequences (Fig. 5A, Supplementary Table S9). The HyPhy RELAX analysis showed
293 that selection pressures have overall intensified along branches in the GGPPS-like clade containing the
294 neoisopteran TPSs ($k = 11.1$, $p < 10^{-4}$). The partitioned descriptive model fitted to each branch class
295 indicated an overall increase in the intensity of diversifying selection along GGPPS-like clade branches
296 compared to the canonical GGPPS reference branches (Supplementary Table S10, $\omega = 22.72$ vs. 3.58).
297 Additionally, a lower degree of constraint from purifying selection acting on the GGPPS-like vs.
298 canonical GGPPS clades can be inferred from the substantial decrease in the proportion of codons
299 under negative selection (Supplementary Table S10, 33.34% in GGPPS-like sequences vs. 97.06% in
300 canonical GGPPSs) and a concurrent emergence of a category of neutrally evolving codons in the test
301 branch class (Supplementary Table S10, 64.45% vs. 0).

302

303 **Fig. 5. Selection analysis of termite GGPPS homologs. A.** Maximum-likelihood phylogenetic tree based
304 on codon alignment of sequences listed in Table S9, reconstructed using IQTree and substitution model
305 JTT+F+I+G4, ultrafast bootstrap values (1000 replicates) shown as node support. RELAX analysis was
306 performed with GGPPS sequences (clade I, reference branches) and GGPPS-like clades II–IV (yellow
307 highlighting, test branches) showing increased intensity of positive selection in GGPPS-like clades. The
308 arrows point to internal branches along which positive selection has likely occurred according to
309 aBSREL analysis. Ancestral sequences were reconstructed for nodes 1–3 (Supplementary Table S10).
310 Functionally characterized extant sequences relevant for the selection analysis are set in bold.
311 **B.** Structure model of ancestral sequence at node 2 with amino acid residues colored with respect to
312 Bayes Empirical Bayes posterior probabilities of the site having experienced positive selection (0.5–1,
313 sites with BEB values > 0.8 are labelled). Active site cavity is filled with grey mesh, $Mg^{2+}$ ions shown as
314 green spheres. **C.** Overlay of modelled structures of NtGGPPS6 (green) and ancestral sequence 2

315 (orange) focusing on the amino acid sites highlighted in B. **D.** Abridged alignment of amino acid
316 sequences of extant GGPPS and TPS sequences in target species and ancestral sequences 1–3 showing
317 residues at insect GGPPS motifs 1–6 (IGMs), FARM and SARM. Amino acids identical to GGPPS
318 sequences are shaded grey. BEB values at individual sites for the branch 2—3 are shown as in B, sites
319 in node 2 sequence with BEB values > 0.8 are shaded yellow.
320

321 We next tested all branches within the termite GGPPS-like clade for evidence of positive selection
322 affecting the diversification of the clade by applying the aBSREL (adaptive branch-site random effects
323 likelihood) procedure in HyPhy to our dataset. Three internal branches were significantly likely to have
324 experienced positive diversifying selection (Holm-Bonferroni-corrected p-values < 0.05). Two of these
325 occur within the GGPPS-like clade II containing EneoC and NtGGPPS6. We designated the internal
326 nodes delimiting these branches 1, 2, and 3 (Fig. 5A), and used PAML to reconstruct their ancestral
327 sequences (Supplementary Table S11). PAML was also employed to detect specific amino acid sites
328 affected by positive selection (ω > 1) using the Bayes empirical Bayes (BEB) method. For the branch
329 connecting nodes 2 and 3, posterior probability higher than 0.9 was recorded in four sites
330 corresponding to the following putative amino acid changes: P23S, Q26G, S59G, N90S (Fig. 5C). Two
331 additional sites (Q21K, S60I) showed BEB values higher than 0.8. These sites were retained almost
332 completely in both NtGGPPS6 (K21, S23, G26, G59, G60, S90) and EneoC (K21, S23, G26, C59, I60, S90).
333 No site with a BEB value exceeding 0.9 was detected on the branch leading to node 2. A structure
334 model of ancestral sequence 2 was then generated using AlphaFold 3 and compared to NtGGPPS6
335 (Fig. 5B), showing that all listed sites are located near the active site cavity opening. Three of these
336 (Q21K, P23S, Q26G) lie on a loop adjacent to the opening, a region with greater conformational
337 flexibility compared to the prevailing α-helical secondary structure.
338

339 **2.9 Structural features underlying cyclization capacity and enantioselectivity of termite TPS**

340 We modelled structures of proteins with known activity (EneoC, EneoE, EneoJ, NtGGPPS6, and
341 NtGGPPS7), and mapped the IGM motifs onto the structures. Overall, all protein structures aligned
342 very well as exemplified by EneoJ and EneoC models (Fig. 6A–C). Interestingly, a distinct pre-cyclization

343 GGPP substrate conformation was modelled with EneoE protein, contrasting with a rather linear GGPP
344 positioning in the active site of NtGGPPS6 (Fig. 6D). To evaluate the proximity and potential significance
345 of individual amino acid changes for protein enzymatic activities, we focused on the less divergent
346 clade II GGPPS-like TPS enzymes (EneoC, NtGGPPS6). Specifically, we searched for variable amino acids
347 located near the FARM and SARM motifs within the highly polar part of the active site where the
348 isoprenoid carbocation is formed and transferred to an IPP molecule in IDS enzymes, taking IGM motif
349 substitutions into account. We noticed an interesting site within IGM 4, a motif located in helix 6,
350 approximately 8 Å from FARM (Fig. 6C–D). Thr152 is conserved in NtGGPPS6 and all IDS enzymes as a
351 part of the KT motif (Fig. 6A), and replaced by isoleucine in EneoC. We hypothesized that Thr152 might
352 be important for carbocation stabilization which would aid GGPP cyclization into neocembrene by
353 NtGGPPS6. Indeed, the substitution of this amino acid in the T152I mutant of NtGGPPS6 abolished
354 neocembrene production from GGPP and instead lead to the formation of the linear diterpene alcohol
355 geranyllinalool, a $C_{20}$ homolog of nerolidol. In addition, upon incubation of T152I NtGGPPS6 with FPP,
356 we observed a significant shift in the enantioselectivity of (6*E*)-nerolidol biosynthesis. While the wild
357 type produced 65% of the *R* enantiomer, this proportion increased to 91% in the mutant enzyme, thus
358 nearing the full enantioselectivity of EneoC (Fig. 6B, Supplementary Table S12).

359

360 **Fig. 6. Comparison of overall structures and active sites of termite GGPPS and GGPPS-like proteins.**
361 **A.** Location of insect GGPPS motifs (IGMs) and aspartate-rich motifs (FARM and SARM) in GGPPS and
362 GGPPS-like sequences; IGM 4 motif alignment in functionally characterized termite proteins and the
363 human HsGGPPS, position 152 highlighted. **B.** TPS activities of T152I NtGGPPS6 mutant and the native
364 protein (WT) documented by GC chromatograms of SPME samples. Left, (6*E*)-nerolidol generation
365 from FPP and the plot of (3*R*,6*E*)-nerolidol enantiomer proportion determined by chiral analysis (mean
366 values ± SD, source data in Supplementary Table S12). Right, (6*E*,10*E*)-geranyllinalool and (*E*,*E*,*E*)-
367 neocembrene generation from GGPP. **C.** Aligned structural models of EneoJ (GGPPS, green) and EneoC
368 (nerolidol synthase, orange). Colored regions correspond to IGM 1–3 and IGM 5 (light blue), IGM 4
369 (pink), FARM (yellow), and SARM (dark blue). $Mg^{2+}$ ions are shown as green spheres. The active site
370 cavity is represented by a grey mesh. **D.** Active site cavity with fitted substrates in cyan (FPP in EneoC,
371 GGPP in NtGGPPS6 and EneoE) and amino acids close to the substrate in dark grey. $Mg^{2+}$ ions are
372 shown as green spheres.

373

374 **3. DISCUSSION**

375 A novel biosynthetic capacity emerged in termites of the modern lineage Neoisoptera, enabling them
376 to produce a remarkably broad scope of terpenoid secondary metabolites. The goal of this study was
377 to unveil the evolutionary background of this innovation by identifying the responsible terpene
378 synthases. We provide evidence that termite terpenoids are biosynthesized by a family of GGPPS-like
379 enzymes, which evolved by a duplication of a GGPPS gene and a subsequent expansion accompanied
380 by the gain of TPS function. The identification of termite TPS enzymes as members of a GGPPS-like
381 gene family adds another instance to the list of independent acquisitions of TPS function via the
382 duplication and neofunctionalization of ancestral IDS (FPPS or GGPPS) genes previously reported in
383 several lineages of insects and in fungi [18–25,35]. In fact, the classical TPS gene family in plants and
384 microorganisms has itself been hypothesized to have arisen from an IDS ancestor [13,36], possibly
385 making its origin the most ancient of the recurring IDS neofunctionalization events.

386 Our data show that GGPPS-like family evolved in the common ancestor of Neoisoptera, suggesting its
387 concurrent origin with the terpene-producing frontal gland, a key neoisopteran synapomorphy. A
388 broader examination of published data shows that the evolutionary origin of terpene synthases in the
389 ancestor of Neoisoptera aligns with the presence of terpenoid secondary metabolites across termite
390 phylogeny, as both features are first recorded in the basal neoisopteran genus *Stylotermes* (Fig. 3). The
391 GGPPS-like genes remained present in the genomes of all later branching clades over roughly 100
392 million years of their diversification (according to dating by Buček et al. [37]), even though in some
393 extant lineages the terpenoids themselves seem to be missing. A trend toward higher terpenoid
394 complexity is apparent in later branching lineages of Neoisoptera, presumably reflecting the
395 involvement of additional modifying enzymes acting on TPS-derived skeletons (Fig. 3). For example,
396 the oxygenated polycyclic diterpenes such as the three highly complex structures identified here in
397 *N. takasagoensis* (Fig. 1B) are expected to arise via subsequent modifications of the TPS-produced
398 (3*E*,7*E*,11*E*)-neocembrene [1,26]. The most intuitive candidates for further cyclizations and

399 oxygenations are cytochromes P450, some of which are reported as conspicuously upregulated in the
400 frontal glands of *Nasutitermes* soldiers [12,27].
401 In eusocial insects, a gene family expansion accompanied by a caste-biased expression pattern of newly
402 emerging paralogs, as observed here in the case of GGPPS, represents a feature often associated with
403 caste diversification [38–41]. Multiple examples have been accumulated showing that duplications and
404 neofunctionalizations of gene sets and their differential use in individual castes drive social insect
405 phenotypic plasticity, e.g., in genes involved in endocrine signalling, chemosensation, metabolism or
406 digestion [28,42–45].
407 The emergence of a gene family and its further expansion can be mediated by different mechanisms,
408 which are not necessarily shared in IDS-like gene families in individual insect lineages. For example, the
409 genomic organization of pan-Blattodean FPPS-like genes in termites is suggestive of origin by tandem
410 duplication, which also appears to have been a major driver of FPPS-like gene family expansion in stink
411 bugs [23]. Alternatively, gene duplication can be mediated by transposable elements [46], examples
412 including non-TPS genes involved in terpenoid biosynthetic pathways in roses [47]. We provide
413 indications for the involvement of transposable elements (TEs) in shaping the expansion of GGPPS-like
414 gene family in *E. neotenicus*. A single TE, TE_00000081, was consistently associated with focal loci in
415 *E. neotenicus*, and sequences homologous to this element are also present in GGPPS transcripts from
416 *N. takasagoensis* and in the *Reticulitermes speratus* GGPPS genomic locus (Supplementary Fig. S13).
417 We thus propose that GGPPS gene duplications in Neoisoptera were frequently concurrent with
418 transposition events. Transposable element-mediated gene duplication would also conform with the
419 observed presence or absence of mitochondrial targeting sequences (MTS) in termite GGPPS/GGPPS-
420 like proteins (Supplementary Fig. S13): the MTS might have been acquired during an early transposition
421 event of the GGPPS coding sequence, explaining its prevailing presence in GGPPS-like proteins and its
422 absence in GGPPSs.
423 We have identified three TPS enzymes which all produce (6*E*)-nerolidol from FPP *in vitro*. Out of these,
424 only EneoC is an enantiospecific TPS generating pure (3*R*,6*E*)-nerolidol, which acts as a queen

425 pheromone in *E. neotenicus*. *EneoC* gene has been found as highly upregulated in queens at transcript
426 and protein levels, both in an inter-caste comparison and relative to other GGPPS-like genes,
427 supporting its biological role in queen pheromone production. Interestingly, to our knowledge, EneoC
428 is the only known terpene synthase producing exclusively the *R* enantiomer of nerolidol. The only other
429 enzyme reported to generate enantiomerically pure (3*R*,6*E*)-nerolidol is a maize enzyme which,
430 however, also produces additional sesquiterpenes [48]. The other two TPSs which we functionally
431 characterized, NtGGPPS6 and EneoE, presumably act as (3*E*,7*E*,11*E*)-neocembrene synthases *in vivo*,
432 generating a precursor of defensive compounds in *N. takasagoensis* and a queen specific volatile in
433 *E. neotenicus*. In addition to TPSs, we have also identified canonical GGPPs in both *E. neotenicus* and
434 *N. takasagoensis* corresponding to typical insect class III GGPPS enzymes [49–51].
435 To obtain a comprehensive view of terpenoid biosynthesis in termites, it would be required to identify
436 the enzymes responsible for production of monoterpenes. Therefore, one of the selected models was
437 *N. takasagoensis*, whose defensive compounds are rich in pinene and limonene. To our dismay, none
438 of the recombinant *N. takasagoensis* enzymes showed monoterpene synthase activity. Recognizing
439 the abundance of GGPPS-like paralogs in Neoisoptera and the complexity of their correct annotation,
440 we propose that a successful TPS activity screen requires a high-quality candidate list based on
441 manually annotated genomic and transcriptomic sequence data.
442 Genetic novelty provides the substrate for evolutionary processes, and our analyses suggest that
443 selection has shaped the diversification of the Neoisopteran GGPPS-like gene family. We show that the
444 intensity of diversifying selection was higher along branches in the GGPPS-like clades compared to the
445 GGPPS clade and that the GGPPS-like clades are less constrained by purifying selection, both of which
446 facilitates specialized adaptation in general [38]. Furthermore, the basal branches within GGPPS-like
447 clade II, which contains TPS enzymes EneoC and NtGGPPS6, likely experienced episodic positive
448 diversifying selection. Positive selection-driven diversification of terpenoid biosynthesis has been
449 reported in plants [52–54] and we are presenting here an analogous phenomenon in termites. In an
450 early branch of the same clade II, we have identified six amino acids sites with evidence of positive

451 selection corresponding to the following substitutions along the branch: P23S, Q26G, S59G, N90S,
452 Q21K and S60I (Fig. 5B–D). Ser59 and Ser60 are associated with substrate and product chain-length
453 specificity in type III GGPPS [55,56]. In fungal GGPPS-like TPSs, the S60C mutation shifts product
454 specificity from the alcohol (*E*)-nerolidol to (*E*,*E*)-α-farnesene [35]. Gln21 typical for insect GGPPS
455 enzymes is changed in termite TPSs (Q21K) and retained in both TPSs of the butterfly *Heliconius*
456 *melpomene* [24]. At a corresponding position, fungal GGPPS sequences of *Melamspora* share Asn,
457 while an N21I change is present in GGPPS-like TPSs from the same species [35]. Thus, in two out of
458 three lineages with confirmed GGPPS-like TPS enzymes (termites, *Melamspora* fungi), the GGPPS-TPS
459 functional shift is accompanied by an amino acid change in this position.
460 We then focused on features linked to GGPP cyclization activity. The introduction of a T152I mutation
461 in the neocembrene synthase NtGGPPS6 lead to the formation of the linear diterpene alcohol
462 geranyllinalool from GGPP instead of the cyclic neocembrene. Thus, Thr152 is a core amino acid
463 involved in GGPP cyclization. Our results suggest that after the lysis of GGPP, the arising highly reactive
464 carbocation is stabilized by the hydroxyl group of Thr152 at C-1. This intermediate is then attacked by
465 the nucleophilic double bond at C-14, followed by proton elimination from C-16.
466 The NtGGPPS6 T152I mutant further let us identify the importance of Ile152 in shifting the ratio of
467 nerolidol enantiomers synthesized from FPP in favour of the *R* enantiomer. The farnesyl carbocation
468 backbone is stabilized by isoleucine in an orientation suitable for enantiospecific nucleophilic attack by
469 water at C-3. It appears that isoleucine found at a corresponding position in EneoC contributes to the
470 full enantioselectivity of this enzyme giving rise to pure (3*R*,6*E*)-nerolidol reported from queens [9,57].
471 Interestingly, EneoE, which also generates neocembrene, lacks Thr152. We suggest that in EneoE, the
472 substrate cyclization proceeds by a different mechanism. The active site is rich in aliphatic amino acids,
473 enabling GGPP pre-folding prior to pyrophosphate cleavage (see Fig. 6D). The fact that EneoE produces
474 a completely racemic mixture of nerolidol when supplied with FPP also points to distinct reaction
475 mechanisms in the two enzymes.

476 With this study, we outline the evolutionary background of terpenoid biosynthesis in termites, and our
477 findings invite further exploration of several related topics. Firstly, to gain mechanistic understanding
478 of the evolutionary transition from IDS to TPS enzymes, confirming our hypothesis based on AlphaFold
479 models using crystal structures is required, followed by a thorough mutational analysis of candidate
480 amino acids. Furthermore, we propose that the high diterpenoid diversity recorded in Neoisoptera is
481 due not only to terpene synthases, but also to other types of currently unidentified enzymes which
482 modify the basic terpene skeleton. The presence of monoterpenes in termites raises questions about
483 the identity of the monoterpene synthase involved. In addition, the characterization of the enzyme
484 responsible for producing the queen pheromone (3*R*,6*E*)-nerolidol in *E. neotenicus* queens opens
485 important questions about its regulation and localization at both tissue and subcellular levels, which
486 remain unknown [9].
487
488
489 **5. MATERIALS AND METHODS**
490 **5.1 Origin of termites**
491 Colonies of *Nasutitermes takasagoensis* (Nawa, 1911) and *Embiratermes neotenicus* (Holmgren, 1906)
492 were used in this study. *N. takasagoensis* colony was collected at West beach, Ishigaki island, Okinawa
493 Prefecture, Japan (N24°25.886' E124°04.280') in September 2021 and kept thereafter at Okinawa
494 Institute of Science and Technology, Japan. Samples for chemical analyses were collected in January
495 2022 and processed at IOCB Prague.
496 Colonies of *E. neotenicus* were collected in French Guiana between 2022 and 2024 at two forest
497 localities along the Road to Petit Saut (N5°04.250' W52°58.770'–N5°04.650' W53°01.360') and one
498 locality 13 km south-west of Kourou (N5°06.440' W52°45.521').
499
500

501 **5.2 Chemical analyses**

502 **5.2.1 Chemical analyses of *N. takasagoensis***

503 Entire bodies of 500 *N. takasagoensis* soldiers were extracted overnight in 5 mL of hexane. 1 µL aliquot
504 of soldier extract was analyzed using a TRACE 1310 gas chromatograph coupled to an ISQ LT
505 quadrupole mass spectrometer. The instrument was equipped with an Rxi-5MS column (30 m ×
506 0.25 mm × 0.25 µm). A split/splitless injector was operated in split mode with a 10:1 split ratio, and
507 the injector temperature was set to 250 °C. The carrier gas flow rate was maintained at 1.5 mL×min$^{-1}$.
508 The oven temperature program was as follows: the initial temperature was set to 50 °C and held for
509 1 min, followed by a ramp of 8 °C×min$^{-1}$ to a final temperature of 320 °C, which was held for 5 min.
510 Mass spectra were acquired at an ionization energy of 70 eV, with a solvent delay of 4 min. The transfer
511 line temperature was set to 260 °C.

512

513 **5.2.2 Chemical analyses of *E. neotenicus***

514 In *E. neotenicus*, we sampled 3 neotenic queens, 10 workers, 10 soldiers, and 10 female nymphs (4$^{th}$
515 stage), and extracted them with 300 µL of hexane for 10 min. One primary king was extracted with
516 100 µL of hexane for 10 min. A 1 µL aliquot of each extract from individual castes was analyzed using
517 comprehensive gas chromatography coupled with mass spectrometry (GC×GC-MS). The analysis was
518 performed on an Agilent gas chromatograph equipped with a secondary oven, quad-jet thermal
519 modulator, and a Gerstel multipurpose sampler used as an autosampler. A LECO Pegasus BT 4D time-
520 of-flight mass spectrometer was used as the detector. The injection was carried out using a
521 split/splitless injector operated in splitless mode, with the injector temperature set to 250 °C. The
522 chromatographic setup consisted of two columns: in the first dimension, an Rxi-5MS column (30 m ×
523 0.25 mm × 0.25 µm) and in the second dimension, an Rxi-17Sil column (1.5 m × 0.25 mm × 0.25 µm)
524 was used. The oven temperature program was as follows: the initial temperature was set to 50 °C and
525 held for 1 min, followed by a ramp of 8 °C×min$^{-1}$ to a final temperature of 320 °C, which was held for
526 5 min. The modulation period was set to 3 seconds, and the transfer line temperature was maintained

527 at 250 °C. A 15 °C offset was used for the secondary oven and a 15 °C offset relative to the secondary
528 oven was used for the modulator.
529
530 **5.2.3 Preparative gas chromatography and NMR**
531 Unknown polycyclic diterpenes from soldier extracts of *N. takasagoensis* were identified using a
532 combination of preparative gas chromatography (prep-GC) and NMR spectroscopy. The prep-GC
533 system consisted of an Agilent GC connected to a Gerstel preparative fraction collector (PFC) and
534 equipped with a Gerstel program temperature vaporization (PTV) inlet. A 5 µL aliquot of the sample
535 was injected onto an Rxi-5MS column (30 m × 0.53 mm × 1.50 µm). PTV temperature was set to 250 °C
536 for 0.1 min, then ramped at 700 °C×min$^{-1}$ to 320 °C held for 4 min; then the inlet was cooled down at
537 100 °C×min$^{-1}$ back to 250 °C. Column flow was set to 1.4 mL×min$^{-1}$. After 1.5 min the inlet pure flow
538 was increased to 50 mL×min$^{-1}$ and held for 3 min. The GC temperature program was set as follows: the
539 starting temperature was set to 200 °C and held for 1 min, then ramped to 320 °C by 15 °C×min$^{-1}$ and
540 held for 3 min. The PFC temperature was set to 320 °C. Fraction collection was performed using coiled
541 glass traps (1 µL capacity) cooled to 0 °C with liquid nitrogen. Each trap was washed with 500 µL of
542 dichloromethane. The collected samples were then evaporated under a stream of nitrogen and re-
543 diluted with deuterated dichloromethane. Six traps were connected to the effluent splitter; however,
544 only three fractions were collected with sufficient purity and quantity for subsequent NMR analysis.
545 NMR spectra were measured on Bruker Avance III HD 600 MHz (600.1 MHz for $^{1}$H spectra and 150.9
546 MHz for $^{13}$C spectra) equipped with 1.7 mm TCI MicroCryoProbe. Identification was based on H,H-
547 COSY, H,C-HSQC, H,C-HMBC, and H,H-NOESY 2D spectra.
548
549 **5.2.4 Chiral chromatography**
550 For the chiral separation of pinene, a 1 µL aliquot of *N. takasagoensis* soldier extract was analyzed
551 using an HP 6850 gas chromatograph equipped with a flame ionization detector (FID) and an HP-Chiral
552 20β GC column (30 m, 0.25 mm, 0.25 µm) with helium as the carrier gas. The split/splitless injector was

553 operated in splitless mode, and the injector temperature was set to 230 °C. The carrier gas flow rate
554 was maintained at 1.5 mL×min$^{-1}$. The detector temperature was set to 220 °C, with an $H_2$ flow of
555 40 mL×min$^{-1}$ and an air flow of 400 mL×min$^{-1}$. The temperature program was as follows: an initial
556 temperature of 40 °C held for 20 min, followed by a ramp of 1 °C×min$^{-1}$ to 100 °C, then a secondary
557 ramp of 30 °C×min$^{-1}$ to 240 °C, which was held for 10 min. The same setup was used for chiral analysis
558 of limonene, differing only in the temperature program which was set as follows: starting temperature
559 was held for 1 min and then ramped 3 C°/min to 130 °C followed by second temperature ramp
560 30°C/min to 240 which was held for 10 min. Standards of (+)- and (–)-α-pinene and (+)- and (–)-
561 limonene were used for identification.
562 For nerolidol analysis, the headspace in the vial from the TPS assay was sampled using a
563 polydimethylsiloxane-coated solid-phase microextraction (SPME) fiber (red fiber, 100 µm; Supelco) for
564 the duration of incubation, followed by desorption in a split/splitless injector operated in splitless
565 mode. The instrument setup for this analysis was the same as for the monoterpene analysis, except
566 for the temperature program: 80 °C held for 1 min, followed by a ramp of 2 °C×min$^{-1}$ to 200 °C, which
567 was then held for 10 min. For quantitation of nerolidol enantiomers in WT and T152I mutant
568 NtGGPPS6, the peak area of (3*R*,6*E*)-nerolidol was divided by sum of peak areas of both enantiomers.
569 The difference in resulting ratios in wild type and mutant enzymes was tested by a two-tailed t-test.
570 Each experiment was done in triplicate.
571
572 **5.3 DNA and RNA isolation, library preparation, sequencing**
573 **5.3.1 DNA isolation and sequencing of *E. neotenicus***
574 High molecular weight DNA for Oxford Nanopore sequencing was isolated by organic extraction from
575 4 male and 8 female nymphs (whole bodies without digestive tube) of *E. neotenicus*. Deep-frozen tissue
576 samples were homogenized in 600 µL of lysis buffer (10 mM Tris-HCl, 400 mM NaCl, 100 mM EDTA,
577 1% SDS and 1 mg×mL$^{-1}$ Proteinase K, pH 8.0) with a PTFE pestle. After overnight incubation at 37 °C,
578 DNA was extracted from aqueous phase by two consecutive extractions with 650 µL

579 phenol:chloroform:isoamyl alcohol (25:24:1) and a reextraction with 600 µL chloroform:isoamyl
580 alcohol (24:1). Phases were separated by centrifugation (5 min, 14,000 g, 4 °C). The DNA was then
581 precipitated by addition of NaCl and ethanol to final concentration of 0.2 M and 70% respectively,
582 incubated for 1 hour at 4 °C, washed with 70% ethanol and resuspended in nuclease-free water. The
583 purity and quantity of isolated DNA was assessed with Nanodrop ND-1000 spectrophotometer and
584 Qubit 4 fluorometer using the dsDNA BR Assay Kit (Thermo Fisher Scientific), the integrity was checked
585 on 0.5% agarose gel in TAE buffer. The samples were used for Oxford Nanopore sequencing analysis
586 on one PromethION flowcell at Novogene UK.
587
588 **5.3.2 RNA isolation and sequencing**
589 *E. neotenicus* kings, neotenic queens, soldiers, workers and female nymphs were sampled for RNA
590 isolation. Collected individuals were cold-anesthetized and dissected to remove the digestive tube and,
591 when appropriate, to sample distinct tissues (heads and thorax + abdomen). Samples were stored in
592 RNA later (Thermo Fisher Scientific). Samples were homogenized in TRI Reagent (Sigma-Aldrich) with
593 a PTFE pestle following the manufacturer's protocol. The precipitated RNA was resuspended in DEPC-
594 treated $ddH_2O$. RNA quality and quantity were inspected on Nanodrop ND-1000 UV/VIS
595 spectrophotometer, Qubit 4 fluorometer with the RNA HS Assay Kit (all Thermo Fisher Scientific), and
596 Agilent 2100 Bioanalyzer using the RNA 6000 Nano Kit (Agilent Technologies). The samples were
597 shipped to Novogene UK, where non-stranded poly(A)-selected RNA library preparations and
598 sequencing analysis on Illumina NovaSeq platform with output of 20 million of 2×150 bp paired-end
599 reads were conducted.
600
601 **5.4 Bioinformatics and protein structure prediction**
602 **5.4.1 *E. neotenicus* draft genome assembly and annotation**
603 The quality of Oxford Nanopore sequencing data was evaluated using fastQC v0.12.1 and NanoPlot
604 v1.30.1, and residual sequencing adapters were removed using Porechop v0.2.4. Draft *E. neotenicus*

605 genome was assembled using Flye v2.9.4 [58] with 5 subsequent polishing iterations, and the
606 estimated genome size was 0.9 Gb, based on an in-house draft genome assembly of a related termite
607 species *Prorhinotermes simplex* [59]. High-quality Illumina RNA data were mapped to the newly
608 assembled draft genome with Segemehl aligner v0.3.4 [60] and used for an additional round of
609 assembly polishing with Pilon algorithm v1.24 [61]. In a final step, all contigs shorter than 1 kb were
610 removed, and assembly quality was evaluated with Quast v5.0.2 [62], BUSCO v5.4.4 using the
611 insecta_odb10 HMM library [63], and Merqury using Meryl v1.3 [64]. The final draft genome with a
612 total size of 1.38 Gb consisted of 4,512 polished scaffolds with an N50 of 3.2 Mb, and covered 99.4%
613 (97.4% single-copy, 2.0% duplicated, 0.4% fragmented, 0.2% missing) from a total of 1,367
614 benchmarking universal single-copy orthologs (BUSCO) [63] of insecta_odb10 lineage. The
615 identification of low-complexity genomic regions was performed using the EDTA pipeline v2.2.0 [65]
616 incorporating both RepeatModeler TE prediction and RepeatMasker for genome masking. The masked
617 genome was then used for annotation with BRAKER v3.0.8 gene prediction tool [66] using full RNA-seq
618 data from caste-specific transcriptomic analysis mapped with Segemehl and annotated genes from
619 published isopteran assemblies of *Zootermopsis nevadensis*, *Cryptotermes secundus*, and
620 *Reticulitermes speratus*. Gene annotation provided in total 26,292 gene models covering 92.8% insect
621 BUSCOs. The genome assembly statistics are summarized in Supplementary Table S1.
622
623 **5.4.2 Transcriptome assembly**
624 The *E. neotenicus* transcriptome was assembled *de novo* with RNA-seq data representing heads and
625 body cavities from kings, queens, workers, soldiers and female nymphs. The data were stripped of
626 artefacts using rCorrector [67] and Trimmomatic v0.32 [68], reads corresponding to ribosomal RNA
627 were filtered out based on mapping to the blacklist from SILVA database [69] using bowtie2 v2.3.4.1
628 algorithm [70]. The transcriptome was assembled using rnaSPAdes algorithm [71] and the assembly
629 quality was evaluated using rnaQUAST [72] and BUSCO [63]. Open reading frames (ORFs) with at least
630 100 amino acid residues in size were predicted using Transdecoder

631 (https://github.com/TransDecoder). The SPAdes assembly provided roughly 500,000 unfiltered
632 contigs, covering 98.3% of 1,367 insect BUSCO genes (28.4% single-copy, 69.9% duplicated, and 0.8%
633 fragmented). A similar approach was applied for the *de novo* assembly of *S. halumicus* transcriptome
634 from SRX6715957 data downloaded from NCBI SRA archive and for *de novo* assemblies of
635 transcriptomes of *Spinitermes trispinosus* and *Inquilinitermes inquilinus* from our in-house data
636 deposited at NCBI as SRA within the BioProject PRJNA918850.
637
638 **5.4.3 IDS gene discovery**
639 The prediction of *E. neotenicus* FPPS- and GGPPS-like sequences was carried out in a series of BLAST
640 searches [73] against transcriptome assembly and predicted ORFs using orthologous sequences of
641 three related species (*Z. nevadensis*, *C. secundus*, and *N. takasagoensis*) available in public
642 repositories. The same approach was used for *S. halumicus*, *S. trispinosus*, and *I. inquilinus*. Candidate
643 *E. neotenicus* transcript sequences were used for another round of BLAST searches in BRAKER-
644 annotated gene models. Manual annotation of GGPPS-like sequences was performed by a combination
645 of multiple local alignments in unmasked genome including transcripts that were found in
646 transcriptome assembly but not recognized in predicted gene models, and a detailed inspection of
647 exon junctions in RNA-seq data mapped on unmasked genome using IGV genomic viewer [74].
648
649 **5.4.4 *E. neotenicus* expression analysis**
650 Trimmed RNA-seq data from the heads and thoraces + abdomens of *E. neotenicus* nymphs and
651 neotenic queens were mapped to the annotated genome using STAR v2.7.10b aligner [75]. Read-count
652 matrices were generated with the featureCounts algorithm from rSubread package [76] and
653 normalized according to the TPM (Transcripts Per Kilobase Million) method. Differential gene
654 expression was inspected in pairwise analyses comparing castes and body parts using R packages
655 DESeq2 and edgeR [77,78].
656

657 **5.4.5 Transposable elements prediction in the genome of *E. neotenicus***

658 The prediction of low-complexity genomic regions and annotation of suspect transposable elements
659 (TE) in GGPPS-like loci was performed using the EDTA pipeline first on a genome-wide level, later on a
660 subset of 27 genomic contigs selected based on local alignments of protein-coding sequences of
661 GGPPS-like genes and pseudogenes. Only contigs with full or partial overlaps showing homology over
662 80% and containing at least 2 consecutive exon-like regions were selected in the panel and used for a
663 focused prediction of transposable elements in the second round of the EDTA pipeline.

664

665 **5.4.6 Phylogenetic analyses and prediction tools**

666 Publicly available sequences of FPPS and GGPPS homologs were retrieved from NCBI. Online blastp
667 and tblastn searches of nr (non-redundant protein sequences) and TSA databases limited to Blattodea
668 were run with *Z. nevadensis* FPPS and GGPPS amino acid sequences as queries (XP_021942300,
669 XP_021933767). The same searches were also run with *Z. nevadensis* decaprenyl-diphosphate
670 synthase subunit 1 as query (XP_021928794). Local blastp and tblastn searches were used for
671 searching *R. speratus* datasets deposited at figshare [28], our in-house datasets from *E. neotenicus*,
672 *S. trispinosus*, and *I. inquilinus*, and the in-house assembled transcriptome from *S. halumicus*. For
673 phylogenetic analyses, amino acid sequences with a minimum length of 150 amino acids and
674 comprising both FARM and SARM motifs were selected. Unless otherwise stated, sequence alignment
675 was performed using Clustal Omega [79] run from SeaView (v4.7) [80], and the resulting alignments
676 were inspected and trimmed manually in AliView [81]. Phylogenetic trees were reconstructed with IQ-
677 TREE 2 [82] using the ModelFinder option and setting the bootstrap replicate number to 1,000 unless
678 stated otherwise.

679 We used MEME [83] (v5.5.7) to detect insect-specific GGPPS motifs (IGMs) in a dataset comprising
680 functionally characterized insect GGPPSs, several single-copy GGPPSs from a diverse selection of insect
681 lineages, and Blattodean sequences from the ancestral GGPPS clade. Regular expressions were
682 generated for the motifs and their boundaries were set as defined for IPP-binding motifs (IBM) by

683 Rebholz et al. [34]. The presence or absence of GGPPS and GGPPS-like sequences in individual termite
684 species was determined based on a phylogeny for sequences ≥ 150 amino acids that included FARM
685 and SARM motifs. In species with shorter sequences and/or not comprising FARM and SARM motifs,
686 we used the motifs previously identified as characteristic for insect GGPPSs to distinguish between
687 GGPPS and GGPPS-like sequences.
688 Putative MTSs were predicted with MitoFates [84] in newly reported genes from *E. neotenicus*. In
689 addition, we checked the completeness of annotated ORFs in publicly available *N. takasagoensis*
690 GGPPS and GGPPS-like mRNA sequences. In two cases (NtGGPPS1B, NtGGPPS3B), we applied MTS
691 predictions to putative proteins coded by ORFs starting with upstream ATG codons not considered in
692 the published annotation. In two other cases (NtGGPPS2B, NtGGPPS4), the ORF was found to be
693 positioned at the very beginning of the transcript. Therefore, it was impossible to assess the
694 completeness of these ORFs and to consider MTS prediction relevant for these proteins.
695 All modelled sequences were trimmed based on alignment with *Homo sapiens* GGPPS sequence
696 (UniProt accession O95749) and were modelled as a single subunit with three $Mg^{2+}$ ions, although it
697 can be assumed that *in vivo*, the enzymes exist as dimers or hexamers [85]. Structure models for EneoC,
698 EneoE, NtGGPPS6, and for the ancestral sequence at node 2 were generated using Alphafold 3.0
699 provided by e-INFRA project. The GGPP substrate was included in models of NtGGPPS6, EneoE, and
700 the node 2, the FPP substrate in case of EneoC. NtGGPPS7 and EneoJ were modelled by homology-
701 based modelling using modeller 10.6 software [86] with structure of *Homo sapiens* GGPPS (PDB Entry
702 2Q80) as a template. The active site cavity was identified and highlighted using KVfinder-web [87]
703 implementing parKVFinder v 1.2.0 [88]. We probed whole structures with probe size in 1.4 Å and probe
704 size out 4 Å. Structures were aligned and processed using UCSF ChimeraX 1.6.1 [89].
705
706 **5.5 Quantitative proteomics**
707 Heads and thoraces + abdomens (with digestive tubes removed) from four nymphs and four neotenic
708 queens of *E. neotenicus* were subjected to label-free quantitative (LFQ) mass spectrometry (MS)

709 analysis. Each sample was homogenized in 500 µL of homogenization buffer consisting of 100 mM
710 triethylammonium bicarbonate buffer (TEAB; Sigma-Aldrich) and 2% sodium deoxycholate (SDC;
711 Sigma-Aldrich) in a 1.5-mL Eppendorf tube using FastPrep-24 homogenizer (MP Biomedicals). The
712 homogenate was centrifuged to remove cell debris and the supernatant was used for analysis. Further
713 sample processing and MS analysis were performed as previously described [90]. Briefly, protein
714 concentration was determined using a BCA protein assay kit (Thermo Fisher Scientific). Cysteines were
715 reduced with Tris(2-carboxyethyl) phosphine hydrochloride and blocked with methyl
716 methanethiosulfonate. The samples were digested with porcine trypsin. Then, the samples were
717 acidified with trifluoroacetic acid to a final concentration of 1%. SDC was removed by extraction with
718 ethylacetate and subsequently with hexane. Peptides were desalted using a Michrom C18 column.
719 Dried peptides were resuspended in 25 µL of water containing 2% acetonitrile and 0.1% trifluoroacetic
720 acid. Peptides were separated by nanoliquid chromatography (nanoLC) on a Dionex Ultimate 3000
721 system and analyzed on an Orbitrap Fusion Tribrid mass spectrometer (both Thermo Fisher Scientific).
722 The data were evaluated with MaxQuant version 2.2.0.0 using LFQ algorithms [91,92] and the
723 Andromeda search engine [93]. The key criteria used in the data analysis were a false discovery rate
724 (FDR) of 0.01 for proteins and peptides, a minimum length of seven amino acids, a fixed modification
725 (methylthio), and variable modifications of N-terminal protein acetylation and methionine oxidation.
726 The data were locally evaluated against selected databases of nonredundant protein sequences that
727 were downloaded from NCBI on February 13, 2023. The in-house database of *E. neotenicus* protein
728 sequences was based on protein models generated by BRAKER (see *E. neotenicus* draft genome
729 assembly and annotation) complemented by a set of manually annotated IDS genes. The local database
730 is available along with the raw data and result files in the PRIDE repository of the ProteomeXchange
731 Consortium (see Data availability). The data were processed in Perseus version 2.0.7.0 [94]. Results
732 representing standard contaminants, reverse sequences (decoys), and sequences identified only with
733 modified peptides were discarded. The dataset containing proteins with at least two valid LFQ values
734 in at least one experimental group was further analyzed. The dataset was $\log_2$ transformed. Missing

735 LFQ values were replaced from the normal distribution (width 0.3; downshift 1.8) and histograms
736 before and after the data imputation were examined. A heatmap with hierarchical clustering was used
737 to check the uniformity of the sample proteomes between the two subgroups. Significant differences
738 were determined by a permutation-based two-sided t-test with an error-corrected p-value (FDR = 0.05;
739 S0 = 0.1). The number of randomizations in the analysis was 1,000.
740
741 **5.6 Heterologous expression**
742 Coding sequences of candidate TPS and IDS proteins were codon-optimized for expression in
743 *Saccharomyces cerevisiae*, synthesized, and cloned into MCS-1 of pESC-URA by Genscript. The resulting
744 plasmids coded for recombinant proteins with N-terminal FLAG-tag and were used for expression in
745 *S. cerevisiae*. For expression of recombinant proteins with N-terminal hexahistidine tag in *Escherichia*
746 *coli*, the coding sequences were recloned in MCS-1 of pRSFDuet-1 (Novagen). Proteins in which
747 mitochondrial targeting peptides were predicted by MitoFates were expressed in the presumed
748 cleaved form. Amino acid sequences of all expressed proteins are listed in Supplementary Table S13.
749
750 **5.6.1 Expression in *S. cerevisiae* and metabolite detection**
751 The JWY501 strain of *S. cerevisiae* was used for all yeast experiments [95]. Prior to transformation, the
752 JWY501 strain was grown in a complete medium (YPD) containing 20 g×$L^{-1}$ Tryptone, 10 g×$L^{-1}$ Yeast
753 Extract (both from Duchefa Biochemie), and 20 g×$L^{-1}$ glucose. Selective media (SCE) used for cultivation
754 of transformed JWY501 contained 1.92 g×$L^{-1}$ Yeast Synthetic Drop-out Medium Supplements without
755 Uracil, 6.7 g×$L^{-1}$ Yeast Nitrogen Base without amino acids (both from Sigma-Aldrich), and 20 g×$L^{-1}$
756 glucose. Selective media used for expression had the same composition as SCE, except that 20 g×$L^{-1}$
757 galactose was used instead of glucose to induce expression of the gene of interest.
758 Yeast transformations were performed using the lithium acetate PEG method [96]. Shortly, cells were
759 pre-cultured overnight in YPD medium at 30 °C with shaking at 200 rpm. The following day, the
760 overnight culture was diluted to an optical density (OD600) of 0.1 in fresh YPD medium and grown at

761 30 °C with shaking at 200 rpm until reaching an OD600 of approximately 0.5. Cells were harvested by
762 centrifugation at 3,000g for 5 min, washed twice in 25 mL of sterile $H_2O$, and pelleted again under the
763 same conditions. The resulting cell pellet was resuspended in 1 mL of sterile water, centrifuged at
764 13,000g for 30 seconds, and resuspended in 1 mL of sterile water. For each transformation, 100 µL of
765 resuspended cells were mixed with 240 µL of 50% (w/v) PEG 3350, 36 µL of 1 M LiAc, 100 ng of plasmid
766 DNA, and 100 µg of denatured salmon sperm DNA (all from Sigma-Aldrich) resulting in a final volume
767 of 460 µL. The transformation mixture was gently mixed, incubated at 42 °C for 40 min, and centrifuged
768 at 13,000g for 30 s. Cells were washed with sterile water, plated onto a selective agar medium, and
769 incubated at 30 °C for 3–5 days to allow for colony formation.
770 Cells carrying the plasmid of interest were pre-cultured in a selective medium (YNB - URA + 2% glucose)
771 for 1 day at 30 °C with shaking at 200 rpm. The following day, the culture was diluted to an initial
772 optical density (OD600) of 0.3 in fresh selective medium and grown at 30 °C with shaking until the
773 OD600 reached approximately 1. At this point, cells were harvested by centrifugation, resuspended in
774 YNB - URA + 2% galactose, and incubated to induce expression of the gene of interest. Cultures were
775 sampled and extracted with hexane either on day 1 after induction (1 mL sample + 1 mL hexane) or on
776 day 7 after induction (100 µL sample + 1 mL hexane). The organic phase was analyzed by GC×GC-MS.
777 A 1 µL aliquot of hexane yeast extract was analyzed using an Agilent comprehensive gas
778 chromatograph equipped with a secondary oven and a quad-jet thermal modulator, coupled with a
779 LECO Pegasus 4D Time-of-Flight (TOF) mass spectrometer. The injection was performed using a
780 split/splitless injector in splitless mode, with the injector temperature set to 250 °C. A 10 °C offset was
781 used for the secondary oven and a 10 °C offset relative to the secondary oven was used for the
782 modulator. The primary column used was ZB-5MS (30 m × 0.25 mm × 0.25 µm), and the secondary
783 column was Rxi-17Sil (1.980 m × 0.25 mm × 0.25 µm). Helium was used as the carrier gas, with a flow
784 rate of 1 mL×min$^{-1}$. The temperature program started at 50 °C, held for 1 min, then ramped at 8 °C×min$^{-}$
785 $^{1}$ to 320 °C and was held for 5 min. The transfer line temperature was set to 260 °C. The electron

786 ionization (EI) ion source operated at 70 eV, with the ion source temperature maintained at 220 °C. A
787 solvent delay of 400 s was applied.
788
789 **5.6.2 Expression in *E. coli* and protein purification**
790 For protein purification, pRSFDuet-1 plasmids were used for transformation of *E. coli* Lemo21(DE3).
791 Bacteria were grown in 2.5 L Tunair baffled shake flasks (IBI Scientific) filled with 0.4 L auto-induction
792 medium, composed of 15 g×$L^{-1}$ yeast extract, 10 g×$L^{-1}$ peptone, 10 g×$L^{-1}$ glycerol, 2 g×$L^{-1}$ D-lactose,
793 0.5 g×$L^{-1}$ D-glucose, 100 mM $KH_2PO_4$, pH 7.5, 50 mM NaCl, 25 mM $(NH_4)_2SO_4$ and 3 mM $MgSO_4$. The
794 culture (starting OD600 ~ 0.01, total culture volume 2–2.4 L) was supplemented with 30 mg×$L^{-1}$
795 kanamycin, 30 mg×$L^{-1}$ chloramphenicol and 0.1 mM L-rhamnose, and a drop of Antifoam A concentrate
796 (Merck) was added to each flask. Cultures were shaken for 6 hours at 37 °C, 250 RPM (capped with
797 aluminium foil lids), and further for 16–20 hours at 17 °C (capped with cellulose paper tissues). The
798 cells were harvested by centrifugation and stored at -80 °C.
799 After thawing, the cells were suspended with 10 volumes of lysis buffer (100 g×$L^{-1}$ glycerol, 0.3 M NaCl,
800 50 mM $KH_2PO_4$, pH 6.7, 10 mM imidazole, 2 mM $MgSO_4$) using a tissue homogenizer. The suspension
801 was supplemented with 0.5 mg×$mL^{-1}$ chicken egg lysozyme, 2,000 KU×$L^{-1}$ DNase I (Merck), and 1 mM
802 PMSF and incubated while rolling. After 1 hour, 10 mM β-mercaptoethanol (ME) was added and
803 incubation while rolling continued for another 2 hours at 4 °C. Cells were broken by four passages
804 through Avestin Emulsiflex C3 and the resulting crude lysate was clarified by consecutive
805 centrifugations (30,000g, 5 and 10 min). The supernatant was filtered through 0.6 µm pore glass filter
806 and pump-loaded at 5 mL×$min^{-1}$ onto HisTrap Fast Flow 5 mL column (Cytiva) using ÄKTA start or ÄKTA
807 pure chromatograph (Cytiva). The column was subsequently washed with 250 mL of lysis buffer with
808 5 mM ME and the protein was eluted by a four-step gradient of 10% elution buffer (EB; i.e. lysis buffer
809 with 5 mM ME and 500 mM imidazole), 20% EB, 30% EB, and 100% EB. Protein concentration was
810 estimated using a calculated extinction coefficient at 280 nm.
811

812 **5.7 TPS and IDS *in vitro* assays**

813 For enzyme activity assays, the reaction mixture in a 2 mL glass vial consisted of 50 µM substrate and
814 roughly 5 µM recombinant protein in 10 mM 3-(N-morpholino)propanesulfonic acid (MOPS), 5 mM
815 $MgCl_2$, 1 mM dithiothreitol, 10% glycerol, pH 7.0, in a final volume of 100 µL. The substrates used in
816 TPS assays were GPP, (*E*,*E*)-FPP, or (*E*,*E*,*E*)-GGPP (Merck or Cayman Chemicals). For IDS assays, the
817 substrates included a mixture of 50 µM IPP with 50 µM DMAPP, GPP, or (*E*,*E*)-FPP (Merck or Cayman
818 Chemicals). The reactions were incubated at 30 °C for 1 hour. In TPS assays, the headspace in the vial
819 was sampled by a polydimethylsiloxane-coated SPME fiber for the duration of incubation (red fiber,
820 100 µm; Supelco).

821 The fiber containing the sample was desorbed and analyzed using a TRACE 1310 gas chromatograph
822 coupled to an ISQ LT quadrupole mass spectrometer. An SPME liner was used in the inlet, with the
823 desorption temperature set to 250 °C. The fiber was desorbed in the inlet for 5 min to ensure complete
824 sample desorption. The GC and MS methods were identical to those described previously.

825 Isoprenoid pyrophosphates were analyzed using a Vanquish Horizon HPLC system coupled with an
826 Orbitrap ID-X mass spectrometer. The separation was performed on an ACCUCORE VANQUISH C18+
827 column (1.5 µm, 150 × 2.1 mm). As mobile phase A, 10 mM ammonium bicarbonate with 0.1%
828 ammonium hydroxide was used. Acetonitrile with 0.1% ammonium hydroxide was used as mobile
829 phase B. The column flow rate was set to 0.35 $mL \times min^{-1}$. The system was equilibrated with 95% mobile
830 phase A and 5% mobile phase B before the chromatographic run. The gradient was programmed as
831 follows: 5% B for the first 5 min, ramp to 100% B over 5.5 min, 100% B held for 7.3 min. The column
832 temperature was maintained at 40 °C throughout the analysis. Electrospray ionization (ESI) in negative
833 mode was used with an ionization voltage of -3,000 V applied from 0.2 min. The gas settings were as
834 follows: Sheath gas, 50 Arb; Auxiliary gas, 10 Arb; Sweep gas, 1 Arb. Additional parameters included an
835 ion transfer tube temperature of 325 °C and a vaporizer temperature of 350 °C.

836

837

838 **5.8 Selection analysis and ancestral sequence reconstruction**

839 To investigate the role of selection pressures in the evolution of terpene synthase genes in termites
840 from canonical GGPPS progenitors, we utilized the packages HyPhy (v2.5.29) and PAML (v4.10.7) to
841 perform selection analysis [97,98]. Full-length GGPPS and GGPPS-like coding sequences from termites
842 were selected from both terpene-releasing Neoisoptera and early branching termite clades without
843 terpenoid production. To generate a tree for a selection analysis and ancestral sequence
844 reconstruction, translated coding sequences were first aligned using ClustalOmega (v1.2.3) [79] under
845 default parameters and trimmed using Trimal (v1.5) with a gap threshold of 0.5. The alignment was
846 used to infer a maximum likelihood protein phylogeny with IQTree2 (UFboot x1000) using the
847 substitution model JTT+F+I+G4 selected by ModelFinder and the singleton GGPPS of the cockroach
848 *Blatella germanica* as an outgroup. To generate a codon alignment, nucleotide coding sequences were
849 aligned using ClustalOmega (v1.2.3) implemented in the Translation Align function of Geneious
850 software. To test for clade-wide differences in the selection regimes acting on canonical GGPPS and
851 the expanded GGPPS-like gene family, an analysis was conducted using HyPhy RELAX [99]. Termite
852 canonical GGPPS clades were designated as the reference branches and the GGPPS-like clade
853 containing TPSs characterized in this study as the test branches. dN/dS ratios (ω) of specific branches
854 were estimated and changes in selection between test and reference branches were tested by fitting
855 a selection intensity parameter (k) to the data. Statistical evidence for a change in selection intensity
856 in the test branches was then detected by comparison of the likelihood of the fitted k to the that of a
857 null model in which no change in selection intensity has occurred (k = 1). To test for evidence of
858 episodic diversifying selection in the evolution of the termite GGPPS-like gene family, an analysis was
859 conducted using HyPhy aBSREL [100] implemented in the HyPhy package testing all branches in the
860 GGPPS-like clade. Branches predicted to be under positive diversifying selection using HyPhy aBSREL
861 were again tested for positive selection using branch-site tests implemented using the CODEML
862 program from PAML [98], where the likelihoods of three-ratio foreground models (allowing positive
863 selection on a proportion of sites on the foreground branches) were tested against the likelihood of

tree-wide two-ratio null models (no positive selection). The Bayes Empirical Bayes method was also employed in branch-site tests to predict codons under positive selection pressure on different foreground branches. Significance for all selection analysis was assessed using Holm-Bonferroni corrected p-values with a significance threshold of 0.05.

Ancestral termite TPS sequences were reconstructed in PAML, using the above phylogeny and codon alignment as input. A stretch of 7 amino acids (LLNQTDI) was found to be present in all ancestral sequences across the tree, obviously reflecting an EneoO-specific insertion which did not occur in any other extant sequence. Therefore, we omitted these amino acids from ancestral sequences which are presented in Supplementary Table S11. Accordingly, the ancestral sequence B used for modelling the protein structure lacks this insertion.

**6. Data availability**

*E. neotenicus* genomic data (draft genome assembly fasta, *in silico* annotated gene and protein models, annotation files) are deposited in the ASEP repository of the Czech Academy of Sciences (dataset ID 0640983, https://doi.org/10.57680/asep.0640983). Full gene sequences of *E. neotenicus* FPPSs and GGPPSs are available in NCBI GenBank records PX512703-19. The transcriptomes of *I. inquilinus* and *S. trispinosus* are provided as NCBI TSA files under GenBank IDs GKIC00000000.1 and GLKH00000000.1. Caste-specific RNA-seq data used for differential gene expression analysis and GGPPS expression profiling in *E. neotenicus* are available as NCBI SRR archives (Biosamples SAMN53173298-SAMN53173329). The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (dataset PXD070647). Supplementary tables list, among others, sequences used in phylogenetic analyses and sequences of expressed proteins. Complete chromatographic data are available on request from the corresponding authors.

## 7. Acknowledgments

We would like to thank the Keasling lab (Joint BioEnergy Institute, University of California, Berkeley) for providing the *S. cerevisiae* strain JWY501. We are grateful to Esra Kaymak (Okinawa Institute of Science and Technology, Japan) for assistance with samples of *N. takasagoensis*, and to former IOCB interns Nanyun Zhang and Guillame Masson for participation in experiments.

## 8. Conflict of interests

The authors declare no competing interests.

## 9. References

1. Gössinger E. Chemistry of the secondary metabolites of termites. Prog Chem Org Nat Prod. 2019; 109: 1-384.
2. Šobotník J, Jirošová A, Hanus R. Chemical warfare in termites. J Insect Physiol., 2010; 56: 1012-1021.
3. Noirot, CH. Glands and secretions. In Krishna K, Weesner FM, editors. Biology of termites. New York: Academic Press; 1969. pp 89-123.
4. Costa-Leonardo AM, da Silva IB, Laranjo LT. Termite exocrine systems: a review of current knowledge. Entomol Exp Appl. 2023; 171: 325-342.
5. Stökl J, Steiger S. Evolutionary origin of insect pheromones. Curr Opin Insect Sci. 2017; 24: 36-42.
6. Dolejšová K, Krasulová J, Kutalová K, Hanus R. Chemical alarm in the termite *Termitogeton planus* (Rhinotermitidae). J Chem Ecol. 2014; 40: 1269-1276.
7. Šobotník J, Hanus R, Kalinová B, Piskorski R, Cvačka J, Bourguignon T, Roisin Y. (*E*,*E*)-α-farnesene, an alarm pheromone of the *termite Prorhinotermes canalifrons*. J Chem Ecol. 2008; 34: 478-486.
8. Bordereau C, Pasteels JM. Pheromones and chemical ecology of dispersal and foraging in termites. In: Bignell D, Roisin Y, Lo N, editors. Biology of termites: A modern synthesis. Dordrecht: Springer; 2010. pp. 279-320.
9. Dolejšová K, Křivánek J, Štáfková J, Horáček N, Havlíčková J, Roy V., et al. Identification of a queen primer pheromone in higher termites. Commun Biol. 2022; 5: 1165.
10. Thakur H , Agarwal S, Buček A, Hradecký J, Sehadová H, Mathur V, et al. Defensive glands in Stylotermitidae (Blattodea, Isoptera). Arthropod Struct Dev. 2024; 79: 101346.

11. Ashour M, Wink M, Gershenzon J. Biochemistry of terpenoids: Monoterpenes, sesquiterpenes and diterpenes. In: Wink M, editor. Biochemistry of plant secondary metabolism. Wiley Blackwell; 2010. pp. 258-303.

12. Weitzel C, Simonsen HT. Cytochrome P450-enzymes involved in the biosynthesis of mono- and sesquiterpenes. Phytochem Rev. 2015; 14: 7-24.

13. Jiang SY, Jin J, Sarojam R, Ramachandran S. A comprehensive survey on the terpene synthase gene family provides new insight into its evolutionary patterns. Genome Biol Evol. 2019; 11: 2078-2098.

14. Chen X, Urban JM, Wurlitzer J, Wei X, Han J, O'Connor SE, et al. Canonical terpene synthases in arthropods: Intraphylum gene transfer. Proc Natl Acad Sci USA. 2024; 121: e2413007121.

15. Köllner TG, David A, Luck K, Beran F, Kunert G, Zhou JJ, et al. Biosynthesis of iridoid sex pheromones in aphids. Proc Natl Acad Sci USA, 2022; 119: e2211254119.

16. Ludwiczuk A, Skalicka-Woźniak K, Georgiev MI. Pharmacognosy: Fundamentals, applications and strategy. In: Heinrich M, Sticher O, editors. Pharmacognosy: Fundamentals, applications and strategy. Amsterdam: Elsevier Inc.; 2017. pp. 233-266.

17. Tholl D, Rebholz Z, Morozov AV, O'Maille PE. Terpene synthases and pathways in animals: enzymology and structural evolution in the biosynthesis of volatile infochemicals. Nat Prod Rep. 2023; 40: 766793.

18. Beran F, Rahfeld P, Luck K, Nagel R, Vogel H, Wielsch N, et al. Novel family of terpene synthases evolved from trans-isoprenyl diphosphate synthases in a flea beetle. Proc Natl Acad Sci USA. 2016; 113: 2922-2927.

19. Ducker C, Baines C, Guy J, Euzébio Goulart Santana A, Pickett JA, Oldham NJ, et al. A diterpene synthase from the sandfly *Lutzomyia longipalpis* produces the pheromone sobralene. Proc Natl Acad Sci USA. 2024; 121: e2322453121.

20. Gilg AB, Tittiger C., Blomquist GJ. Unique animal prenyltransferase with monoterpene synthase activity. Naturwissenschaften. 2009; 96: 731-735.

21. Lancaster J, Lehner B, Khrimian A, Muchlinski A, Luck K, Köllner TG, et al. An IDS-Type sesquiterpene synthase produces the pheromone precursor (*Z*)-α-bisabolene in *Nezara viridula*. J Chem Ecol. 2019; 45: 187-197.

22. Lancaster J, Khrimian A, Young S, Lehner B, Luck K, Wallingford A, 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 USA. 2018; 115: E8634-E8641.

23. Rebholz Z, Lancaster J, Larose H, Khrimian A, Luck K, Sparks ME, *et al.* Ancient origin and conserved gene function in terpene pheromone and defense evolution of stink bugs and hemipteran insects. Insect Biochem Mol Biol. 2022; 32: 103879.

24. Darragh K, Orteu A, Black D, Byers KJRP, Szczerbowski D, Warren IA, et al. A novel terpene synthase controls differences in anti-aphrodisiac pheromone production between closely related *Heliconius* butterflies. PLoS Biol. 2021; 19: e3001022.

25. Zhou Y, Liu X, Yang Z. Characterization of terpene synthase from tea green leafhopper being involved in formation of geraniol in tea (*Camellia sinensis*) leaves and potential effect of geraniol on insect-derived endobacteria. Biomolecules. 2019; 9: 808.

962 26. Prestwich GD, Jones RW, Collins MS. Terpene biosynthesis by nasute termite soldiers (Isoptera:
963 Nasutitermitinae). Insect Biochem. 1981; 11: 331-336.

964 27. Hojo M, Shigenobu S, Maekawa K, Miura T, Tokuda G. Duplication and soldier-specific expression
965 of geranylgeranyl diphosphate synthase genes in a nasute termite *Nasutitermes takasagoensis*.
966 Insect Biochem Mol Biol. 2019; 111: 103177.

967 28. Shigenobu S, Hayashi Y, Watanabe D, Tokuda G, Hojo MY, Toga K, et al. Genomic and
968 transcriptomic analyses of the subterranean termite *Reticulitermes speratus*: Gene duplication
969 facilitates social evolution. Proc Natl Acad Sci USA. 2022; 119: e2110361119.

970 29. Prestwich GD, Vrkoč J. Standard nomenclature for termite diterpenes. Sociobiology. 1979; 4: 139-
971 140.

972 30. Ashby MN, Edwards PA. Elucidation of the deficiency in two yeast coenzyme Q mutants.
973 Characterization of the structural gene encoding hexaprenyl pyrophosphate synthetase. J Biol
974 Chem. 1990; 265: 13157-13164.

975 31. Tarshis LC, Sacchettini JC, Yan M, Poulter CD. Crystal structure of recombinant farnesyl
976 diphosphate synthase at 2.6-Å resolution. Biochemistry. 1994; 33: 10871-10877.

977 32. Hellemans S, Rocha MM, Wang ML, Arias JR, Aanen DK, Bagneres AG, et al. Genomic data provide
978 insights into the classification of extant termites. Nat Commun. 2024; 15: 6724.

979 33. Wells JN, Feschotte C. A field guide to eukaryotic transposable elements. Annu Rev Genet. 2020;
980 54: 539-561.

981 34. Rebholz Z, Shewade L, Kaler K, Larose H, Schubot F, Tholl D, et al. Emergence of terpene chemical
982 communication in insects: Evolutionary recruitment of isoprenoid metabolism. Protein Sci. 2023;
983 32: e4634.

984 35. Wei G, Eberl F, Chen X, Zhang C, Unsicker SB, Köllner TG, et al. Evolution of isoprenyl diphosphate
985 synthase-like terpene synthases in fungi. Sci Rep. 2020; 10: 14944.

986 36. Lesburg CA, Zhai G, Cane DE, Christianson DW. Crystal structure of pentalenene synthase:
987 Mechanistic insights on terpenoid cyclization reactions in biology. Science. 1997; 277: 1820-1824.

988 37. Buček A, Šobotník J, He S, Shi M, McMahon DP, Holmes EC, et al. Evolution of termite symbiosis
989 informed by transcriptome-based phylogenies. Curr Biol. 2019; 29: 3728-3734.e4.

990 38. Gadagkar R. The evolution of caste polymorphism in social insects genetic release followed by
991 diversifying evolution. J Genet. 1997; 76: 167-179.

992 39. Korb J. Genes underlying reproductive division of labor in termites, with comparisons to social
993 Hymenoptera. Front Ecol Evol. 2016; 4: 182024.

994 40. Chau LM, Goodisman MAD. Gene duplication and the evolution of phenotypic diversity in insect
995 societies. Evolution. 2017; 71: 2871-2884.

996 41. Fujiwara K, Oka T, Ashihara R, Miyazaki S, Toyoda A, Lo N, et al. Genomic features underlying the
997 origin of sociality and the diversification of caste systems in termites. Mol Biol Evol. 2025; 42:
998 msaf284.

42. Corona M, Libbrecht R, Wurm Y, Riba-Grognuz O, Studer RA, Keller L. Vitellogenin underwent subfunctionalization to acquire caste and behavioral specific expression in the Harvester Ant *Pogonomyrmex barbatus*. PLoS Genet. 2013; 9: e1003730.

43. Zhou X, Rokas A, Berger SL, Liebig J, Ray A, Zwiebel LJ. Chemoreceptor evolution in Hymenoptera and its implications for the evolution of eusociality. Genome Biol Evol. 2015; 7: 2407-2416.

44. McKenzie SK, Fetter-Pruneda I, Ruta V., Kronauer DJC. Transcriptomics and neuroanatomy of the clonal raider ant implicate an expanded clade of odorant receptors in chemical communication. Proc Natl Acad Sci USA. 2016; 113: 14091-14096.

45. Smýkal V, Pivarči M, Provazník J, Bazalová O, Jedlička P, Lukšan O, et al. Complex evolution of insect insulin receptors and homologous decoy receptors, and functional significance of their multiplicity. Mol Biol Evol. 2020; 37: 1775-1789.

46. Cerbin S, Jiang N. Duplication of host genes by transposable elements. Curr Opin Genet Dev. 2018; 49: 63-69.

47. Conart C, Saclier N, Foucher F, Goubert C, Rius-Bony A, Paramita SN, et al. Duplication and specialization of NUDX1 in Rosaceae led to geraniol production in rose petals. Mol Biol Evol. 2022; 39: msac002.

48. Schnee C, Köllner TG, Gershenzon J, Degenhardt J. The maize gene terpene synthase 1 encodes a sesquiterpene synthase catalyzing the formation of (*E*)-β-farnesene, (*E*)-nerolidol, and (*E*,*E*)-farnesol after herbivore damage. Plant Physiol. 2002; 130: 2049-2060.

49. Barbar A, Couture M, Sen SE, Béliveau C, Nisole A, Bipfubusa M, et al. Cloning, expression and characterization of an insect geranylgeranyl diphosphate synthase from *Choristoneura fumiferana*. Insect Biochem Mol Biol. 2013; 43: 947-958.

50. Frick S, Nagel R, Schmidt A, Bodemann RR, Rahfeld P, Pauls G, et al. Metal ions control product specificity of isoprenyl diphosphate synthases in the insect terpenoid pathway. Proc Natl Acad Sci USA. 2013; 110: 4194-4199.

51. Ducker C, French S, Pathak M, Taylor H, Sainter A, Askem W, et al. Characterisation of geranylgeranyl diphosphate synthase from the sandfly *Lutzomyia longipalpis*. Insect Biochem Mol Biol. 2023; 161: 104001.

52. Chen H, Li G, Köllner TG, Jia Q, Gershenzon J, Chen F. Positive Darwinian selection is a driving force for the diversification of terpenoid biosynthesis in the genus *Oryza*. BMC Plant Biol. 2014; 14: 1-12.

53. Chaw SM, Liu YC, Wu YW, Wang HY, Lin CYI, Wu CS, et al. Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution. Nat Plants. 2019; 5: 63-73.

54. Han X, Zhang J, Han S, Chong SL, Meng G, Song M, et al. The chromosome-scale genome of *Phoebe bournei* reveals contrasting fates of terpene synthase (TPS)-a and TPS-b subfamilies. Plant Commun. 2022; 3: 100410.

55. Ohnuma SI, Hirooka K, Ohto C, Nishino T. Conversion from archaeal geranylgeranyl diphosphate synthase to farnesyl diphosphate synthase: Two amino acids before the first aspartate-rich motif solely determine eukaryotic farnesyl diphosphate synthase activity. J Biol Chem. 1997; 272: 5192-5198.

1039 56. Zhang H, Li Z-X. A type-III insect geranylgeranyl diphosphate synthase with a novel catalytic
1040 property. Protein Pept Lett. 2014; 21: 615-623.

1041 57. Havlíčková J, Dolejšová K, Tichý M, Vrkoslav V, Kalinová B, Kyjaková P, Hanus R. (3*R*,6*E*)-nerolidol,
1042 a fertility-related volatile secreted by the queens of higher termites (Termitidae: Syntermitinae).
1043 Z Naturforsch C J Biosci. 2019; 74: 251-264.

1044 58. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs.
1045 Nat Biotechnol. 2019; 37: 540-546.

1046 59. Koubová J, Pangrácová M, Jankásek M, Lukšan O, Jehlík T, Brabcová J, et al. Long-lived termite
1047 kings and queens activate telomerase in somatic organs. Proc Biol Sci. 2021; 288: 20210511.

1048 60. Otto C, Stadler PF, Hoffmann S. Lacking alignments? The next-generation sequencing mapper
1049 segemehl revisited. Bioinformatics. 2014; 30: 1837-1843.

1050 61. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An integrated tool
1051 for comprehensive microbial variant detection and genome assembly improvement. PLoS One.
1052 2014; 9: e112963.

1053 62. Gurevich A, Saveliev V, Vyahhi N, Tesler, G. QUAST: Quality assessment tool for genome
1054 assemblies. Bioinformatics. 2013; 29: 1072-1075.

1055 63. Seppey M, Manni M, Zdobnov EM. BUSCO: Assessing genome assembly and annotation
1056 completeness. Methods Mol Biol.2019; 1962: 227-245.

1057 64. Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: Reference-free quality, completeness, and
1058 phasing assessment for genome assemblies. Genome Biol. 2020; 21: 245.

1059 65. Su W, Ou S, Hufford MB, Peterson T. A tutorial of EDTA: Extensive de novo TE annotator. Methods
1060 Mol Biol. 2021; 2250: 55-67.

1061 66. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-Genome Annotation with BRAKER.
1062 Methods Mol Biol. 2019; 1962: 65-95.

1063 67. Song L, Florea L. Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads.
1064 Gigascience. 2015; 4: 48.

1065 68. Bolger AM, Lohse M, Usadel B. Genome analysis Trimmomatic: a flexible trimmer for Illumina
1066 sequence data. Bioinformatics. 2014; 30: 2114-2120.

1067 69. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P*, et al.* The SILVA ribosomal RNA gene
1068 database project: Improved data processing and web-based tools. Nucleic Acids Res. 2013; 41:
1069 D590-D596.

1070 70. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9: 357-
1071 359.

1072 71. Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. RnaSPAdes: A de novo transcriptome
1073 assembler and its application to RNA-Seq data. Gigascience. 2019; 8: giz100.

1074 72. Bushmanova E, Antipov D, Lapidus A, Suvorov V, Prjibelski AD. rnaQUAST: A quality assessment
1075 tool for de novo transcriptome assemblies. Bioinformatics. 2016; 32: 2210-2212.

1076 73. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+:
1077 Architecture and applications. BMC Bioinform. 2009; 10: 421.

1078 74. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP.
1079 Integrative genomics viewer. Nat Biotechnol. 201; 29: 24-26.

1080 75. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-
1081 seq aligner. Bioinformatics. 2013; 29: 15-21.

1082 76. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for
1083 alignment and quantification of RNA sequencing reads. Nucleic Acids Res. 2019; 47: e47.

1084 77. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression
1085 analysis of digital gene expression data. Bioinformatics. 2009; 26: 139-140.

1086 78. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq
1087 data with DESeq2. Genome Biol. 2014; 15: 550.

1088 79. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences.
1089 Prot Sci. 2018; 27: 135-145.

1090 80. Gouy M, Guindon S, Gascuel O. Sea view version 4: A multiplatform graphical user interface for
1091 sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010; 27: 221-224.

1092 81. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets.
1093 Bioinformatics. 2014; 30: 3276-3278.

1094 82. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE
1095 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol.
1096 2020; 37: 1530-1534.

1097 83. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015; 43: W39-W49.

1098 84. Fukasawa Y, Tsuji J, Fu SC, Tomii K, Horton P, Imai K. MitoFates: Improved prediction of
1099 mitochondrial targeting sequences and their cleavage sites. Mol Cell Proteom. 2015; 14: 1113-
1100 1126.

1101 85. Kavanagh KL, Dunford JE, Bunkoczi G, Russell RGG, Oppermann U. The crystal structure of human
1102 geranylgeranyl pyrophosphate synthase reveals a novel hexameric arrangement and inhibitory
1103 product binding. J Biol Chem. 2006; 281: 22004-22012.

1104 86. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol.
1105 1993; 234: 779-815.

1106 87. Guerra JVS, Ribeiro-Filho HV, Pereira JGC, Lopes-de-Oliveira PS. KVFinder-web: a web-based
1107 application for detecting and characterizing biomolecular cavities. Nucleic Acids Res. 2023; 51:
1108 W289-W297.

1109 88. Guerra JVS, Ribeiro Filho HV, Bortot LO, Honorato RV, de Carvalho Pereira JG, Lopes-de-Oliveira
1110 PS. ParKVFinder: A thread-level parallel approach in biomolecular cavity detection. SoftwareX.
1111 2020; 12: W289-W297.

1112 89. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE. UCSF ChimeraX:
1113 Tools for structure building and analysis. Protein Sci. 2023; 32: e4792.

1114 90. Štáfková J, Rada P, Meloni D, Žárský V, Smutná T, Zimmann N, et al. Dynamic secretome of
1115 *Trichomonas vaginalis*: Case study of β-amylases. Mol Cell Proteom. 2018; 17: 304-320.

1116 91. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range
1117 mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008; 26: 1367-1372.

1118 92. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free
1119 quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ.
1120 Mol Cell Proteom. 2014; 13: 2513-2526.

1121 93. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: A peptide search
1122 engine integrated into the MaxQuant environment. J Proteome Res. 2011; 10: 1794-1805.

1123 94. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational
1124 platform for comprehensive analysis of (prote)omics data. Nat Methods. 2016; 13: 731-740.

1125 95. Wong J, de Rond T, D'Espaux L, van der Horst C, Dev I, Rios-Solis L, et al. High-titer production of
1126 lathyrane diterpenoids from sugar by engineered *Saccharomyces cerevisiae*. Metab Eng. 2018; 45:
1127 142-148.

1128 96. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG
1129 method. Nat Protoc. 2007; 2: 31-34.

1130 97. Kosakovsky Pond SL., Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, et al. HyPhy 2.5 - A
1131 customizable platform for evolutionary hypothesis testing using phylogenies. Mol Biol Evol. 2020;
1132 37: 295.

1133 98. Álvarez-Carretero S, Kapli P, Yang Z. Beginner's guide on the use of PAML to detect positive
1134 selection. Mol Biol Evol. 2023; 40: msad041.

1135 99. Wertheim JO, Murrell B, Smith MD, Pond SLK, Scheffler K. RELAX: Detecting relaxed selection in a
1136 phylogenetic framework. Mol Biol Evol. 2015; 32: 820-832.

1137 100. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: An
1138 adaptive branch-site random effects model for efficient detection of episodic diversifying
1139 selection. Mol Biol Evol.2015; 32: 1342-1353.

1140

1141

1142 **Supplementary Fig. S1. Representative GC×GC chromatograms of body washes from**
1143 **different *E. neotenicus* castes.** The chromatograms show a retention window where terpenoids
1144 were detected. 1, β-farnesene; 2, α-farnesene; 3, (6*E*)-nerolidol; 4, unidentified diterpene; 5, β-
1145 springene; 6, (*E*,*E*,*E*)-neocemberene, 7, unidentified diterpene. The chromatogram visualizes the
1146 responses to diagnostic ions for mono-, sesqui- and diterpenes (*m*/*z* = 68 + 69 + 93 + 204 + 272).
1147 Mass spectra of the identified terpenes are shown below in Supplementary Fig. S2.

1148

1149 **Supplementary Fig. S2. Mass spectra of the terpenes identified in the body washes of**
1150 **different *E. neotenicus* castes.** Compound numbering refers to that in Fig. 1 and Supplementary Fig.
1151 S1. For compound 7 (unidentified diterpene), we could not obtain a reliable mass spectrum.

1152

1153 **Supplementary Fig. S3. Representative GC chromatogram of body washes from**
1154 ***N. takasagoensis* soldiers and workers.** The chromatogram visualizes the responses to diagnostic

ions for mono-, sesqui- and diterpenes (*m/z* = 68 + 69 + 93 + 204 + 272). Mass spectra and NMR spectra of the identified terpenes are shown below in Supplementary Figs S4–S8.

**Supplementary Fig. S4. Identification of compound a from *N. takasagoensis* soldiers as (+)-α-pinene**. **A.** Chiral chromatogram comparing the retention of compound **a** with standards of (-) and (+)-α-pinene. GC-FID was equipped with cyclodextrin chiral column (HP CHIRAL β, 30m, 0.25mm, 0.25µm, 19091G-B233E). One µL of sample was injected in splitless mode following this temperature program: 40 °C held for 20 minutes, ramped 1 °C/min to 100 °C, ramp at 30 °C/min to 240 °C, held for 20 minutes. **B.** Mass spectrum of compound **a**.

**Supplementary Fig. S5. Mass spectrum and chiral analysis of compound b from *N. takasagoensis* soldiers identified as (+)-limonene**.

**Supplementary Fig. S6, part I. Identification of compound c from *N. takasagoensis* soldiers as 14a-acetoxy-3b,12b-dipropionoxy-kampa-6(7),8(9)-diene. A.** Mass spectrum. **B.** Assigned $^{1}$H and $^{13}$C NMR spectra.

**Supplementary Fig. S6, part II. Identification of compound c from *N. takasagoensis* soldiers as 14a-acetoxy-3b,12b-dipropionoxy-kampa-6(7),8(9)-diene. C.** $^{1}$H NMR spectra. **D.** APT NMR spectrum.

**Supplementary Fig. S7, part I. Identification of compound d from *N. takasagoensis* soldiers as 3α,9β,13α-tripropionoxy-trinervita-11(12),15(17)-diene**. **A.** Mass spectrum. **B.** Assigned $^{1}$H and $^{13}$C NMR spectra.

**Supplementary Fig. S7, part II. Identification of compound d from *N. takasagoensis* soldiers as 3α,9β,13α-tripropionoxy-trinervita-11(12),15(17)-diene**. **C.** $^{1}$H NMR spectra. **D.** APT NMR spectrum.

**Supplementary Fig. S8, part I. Identification of compound e from *N. takasagoensis* soldiers as 3α, 9β,13α-tripropionoxy-11α(12β)-epoxy-trinervita-15(17)-ene**. **A.** Mass spectrum. **B.** Assigned $^{1}$H and $^{13}$C NMR spectra.

**Supplementary Fig. S8, part II. Identification of compound e from *N. takasagoensis* soldiers as 3α, 9β,13α-tripropionoxy-11α(12β)-epoxy-trinervita-15(17)-ene**. **C.** $^{1}$H NMR spectra. **D.** APT NMR spectrum.

**Supplementary Fig. S9. Phylogeny and function of FPPS homologs in termites**. **A.** Phylogenetic tree inferred from amino acid sequences of FPPS homologs identified in Blattodea (five termite and two cockroach species) and two other polyneopteran orders, Orthoptera and Phasmatodea. The maximum likelihood tree was reconstructed with IQ-TREE 2 using JTT+G4 model; bootstrap values (500 replicates) greater than 50 are shown as node support. Arrow marks the FPPS duplication predating the diversification of Blattodea. Asterisks highlight the two paralogous sequences EneoA and EneoB from *Embiratermes neotenicus* studied with respect to their function. Accession numbers are provided in Supplementary Table S2. **B.** HPLC chromatogram demonstrating the IDS

activity of EneoB. The purified enzyme was incubated with isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), IPP and geranyl pyrophosphate (GPP), or IPP and farnesyl pyrophosphate (FPP). Only the substrate combinations giving rise to any prenyl pyrophosphate products are shown. The chromatograms visualize the selected m/z 245.00, 313.06, 381.12, and 449.19. No TPS activity was recorded in EneoB and no TPS or IDS activity was observed in EneoA.

**Supplementary Fig. S10. Alignment of IDS catalytic motifs in FPPS and FPPS-like amino acid sequences.** Motif boundaries and sequence regular expressions are displayed as defined by Rebholz et al. (2023) for insect IDS sequences. Residues in consensus with the regular expression are shaded blue. Termite species are displayed in bold text. IBM, IPP-binding motif. FARM, first aspartate-rich motif. SARM, second aspartate-rich motif.

**Supplementary Fig. S11. Genomic landscape surrounding the *E. neotenicus* GGPPS-like genes and pseudogenes.** The figure shows gene organization of full or partial genomic contigs (boldfaced, starting with prefix ENC and contig size in brackets) represented by horizontal line with position coordinates at both sides. Exons are numbered and shown as black (genes), grey (pseudogenes) or red boxes (chimeric duplicates), regions corresponding to genes and pseudogenes are marked with black or grey arrows while regions masked by RepeatMasker at genome-wide scale are shown as pink boxes. Sequences homologous to CACTA TIR TE_00000081 DNA transposon predicted with the EDTA pipeline are marked with green arrows. Blue scale bar represents 5 kb.

**Supplementary Fig. S12. Maximum-likelihood phylogenetic tree inferred from alignment of Blattodea GGPPS amino acid sequences.** The tree was reconstructed using IQ-TREE with the JTT+F+R4 substitution model; bootstrap values (1,000 replicates) greater than 50 are shown as node support. The tree was rooted with a dragonfly GGPPS sequence (*Ischnura elegans*). See Supplementary Table S5 for source sequences.

**Supplementary Fig. S13. Sequence conservation of the CACTA TIR DNA transposable element likely involved in GGPS-like gene multiplication in Neoisoptera. A.** Sequences corresponding to CACTA TIR TE_00000081 predicted with RepeatModeler (green) and homologous sequences found in the proximity of GGPPS-like genes and pseudogenes in *E. neotenicus* (yellow) are shown as arrows, conserved regions are represented by dark color. Red triangles above sequence diagrams represent indels relative to the CACTA TIR TE_00000081 reference. Orthologs from related species *N. takasagoensis* (Ntak) and *R. speratus* (Rspe) found by BLAST search against the complete NCBI nucleotide collection restricted to Blattodea are included. **B.** The sequence of CACTA TIR TE_00000081 predicted with RepeatModeler, latin numbers I-III are used to specify three regions with various levels of conservation.

**Supplementary Fig. S14. HPLC chromatograms demonstrating the IDS activity in canonical GGPPSs EneoJ (yellow) and NtGGPPS7 (green).** The purified enzymes were incubated with IPP and DMAPP, IPP and GPP, or IPP and FPP. The chromatograms visualize the selected m/z 245.00, 313.06, 381.12, and 449.19. The products resulting from IDS activity are underlined with black lines.

1241 **Supplementary Fig. S15. Expression of IDS and IDS-like genes in *E. neotenicus* queens and**
1242 **nymphs at the level of transcript (A) and protein (B). A.** Differential expression analysis
1243 comparing transcript abundances of IDS and IDS-like genes in heads or thoraces and abdomens of
1244 neotenic queens and queen-destined 4$^{th}$ stage female nymphs. Mean $log_2$ TPM values represent
1245 three biological replicates, colors of connecting lines correspond to $log_2$ fold changes between the two
1246 castes. Only genes which showed significant regulation between the two castes in DESEq analysis are
1247 presented (padj < 0.05). Expression data for all IDS/IDS-like genes are listed in Supplementary Table
1248 S7. **B.** Quantitative proteomic analysis of IDS/IDS-like proteins in thoraces and abdomens of queen-
1249 destined 4$^{th}$ stage female nymphs and neotenic queens. Mean $log_2$ intensity values represent four
1250 biological replicates. Colors correspond to $log_2$ fold changes between the two castes. Only proteins
1251 which conform to the permutation-based FDR 0.05 are shown; points with no connecting lines
1252 represent proteins which were only detected in samples from queen thoraces and abdomens. Results
1253 of proteomic analysis for all IDS/IDS-like proteins from all tissues are listed in Supplementary Table
1254 S8.

1255

1256 **Supplementary Fig. S16. Screening of enzymatic activity of GGPPS-like enzymes from**
1257 ***N. takasagoensis.*** Sequences were expressed in *S. cerevisiae* strain JWY501. One mL of culture
1258 was extracted with 1mL of hexane and 1 µL aliquot of hexane extract was measured by
1259 comprehensive gas chromatography mass spectrometry. No specific products were detected
1260 in any of the strains apart from neocembrene (a) in the NtGGPPS6-expressing strain.

1261

1262 **Supplementary Fig. S17. Alignment of IDS catalytic motifs in GGPPS and GGPPS-like**
1263 **protein sequences.** Boundaries of motifs are displayed as defined by Rebholz et al. (2023). Insect
1264 GGPPS-specific regular expression for each motif was generated based on the canonical insect GGPPS
1265 sequences displayed in the alignment using MEME (v5.5.7). Residues in consensus with the regular
1266 expression are shaded blue. Termite species are printed in bold text, amino acid residues for confirmed
1267 TPS proteins are set in bold. MTS, presence or absence of a mitochondrial targeting peptide as
1268 predicted by MitoFates (v1.2). IGM, insect GGPPS motif. FARM, first aspartate-rich motif. SARM,
1269 second aspartate-rich motif.

1270

1271 **Table S1. General statistics of *E. neotenicus* draft genome assembly and in silico annotated**
1272 **gene models.**

1273

1274 **Table S2. Source of FPPS and FPPS-like sequences identified in Blattodea and**
1275 **polyneopteran outgroups.** Related to Fig. S9.

1276

1277 **Table S3. Source of GGPPS and GGPPS-like nucleotide sequences in selected termites**
1278 **and non-termite hexapod outgroups.** Related to Fig. 2.

1279

1280 **Table S4. *Stylotermes halumicus* GGPPS and GGPPS-like nucleotide sequences assembled**
1281 **de novo from the SRA data deposited as SRX6715957.** Related to Fig. 2.

1282

1283 **Table S5. Source of GGPPS and GGPPS-like sequences in selected termites and non-**
1284 **termite outgroups used for protein phylogeny.** Related to Supplementary Fig. 12.

1285

1286 **Table S6. Source of GGPPS and GGPPS-like sequences identified in 56 species of**
1287 **termites.** Related to Fig. 3.

1288

1289 **Table S7. Differential expression analysis on IDS and IDS-like genes in female nymphs and**
1290 **queens of *E. neotenicus*.** Related to Fig. S15.

1291

1292 **Table S8. Quantitative proteomics analysis on IDS and IDS-like proteins in female**
1293 **nymphs and queens of *E. neotenicus*.** Related to Fig. S15.

1294

1295 **Table S9. Source of GGPPS and GGPPS-like sequences used for the selection analysis.**
1296 Related to Fig. 5.

1297

1298 **Table S10. RELAX analysis of selection intensity in GGPPS and GGPPS-like clades.**

1299

1300 **Table S11. Reconstructed ancestral amino acid sequences at nodes 1, 2, and 3.** Related to
1301 Fig. 5.

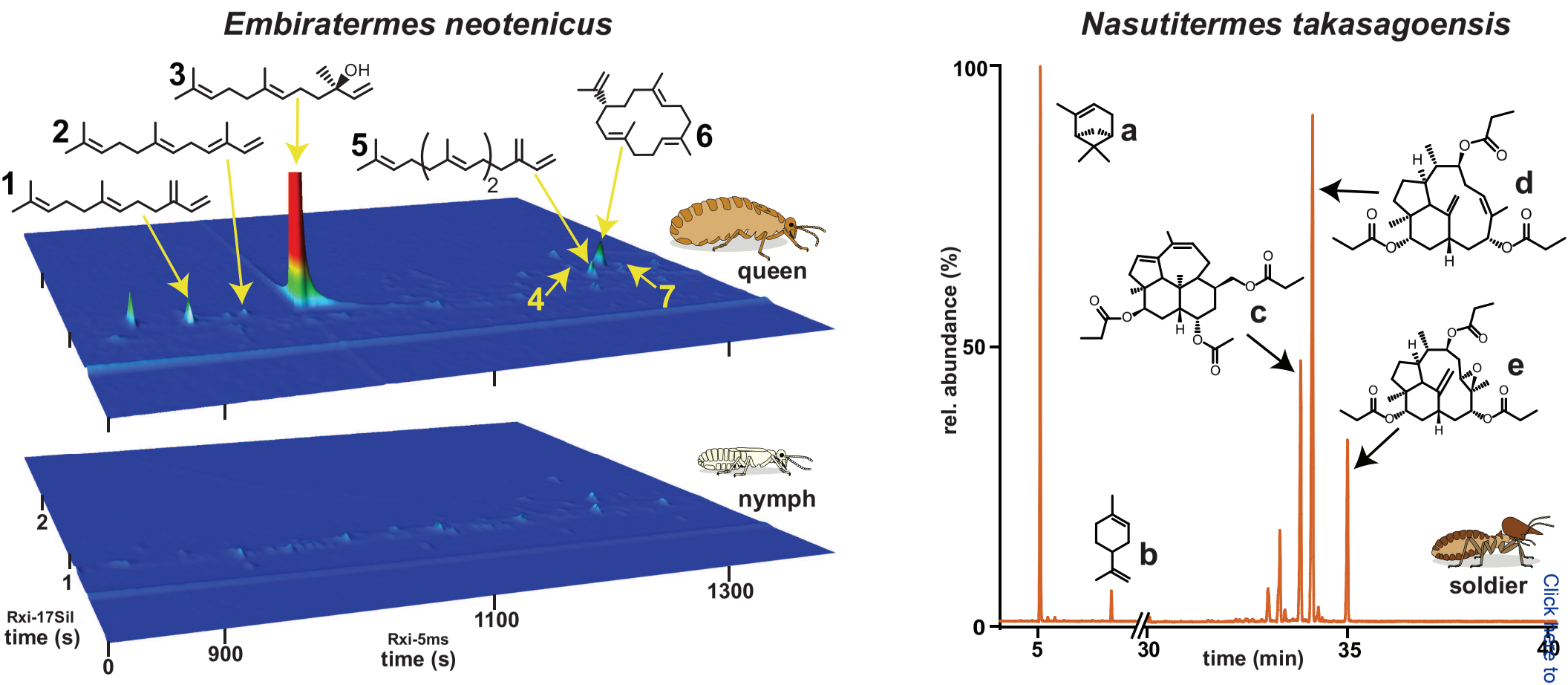
1302

1303 **Table S12. Proportion of *R* and *S* enantiomers of (6*E*)-nerolidol upon incubation of wild**
1304 **type NtGGPPS6 and T152I mutant with FPP.** Related to Fig. 6.

1305

1306 **Table S13. Amino acid sequences of recombinant proteins.**

Figure 1




**A**

| | | queen | king | nymph | soldier | worker |
|---|---|---|---|---|---|---|
| 1 | β-farnesene | + | | | | |
| 2 | α-farnesene | tr | | | | |
| 3 | (3*R*,6*E*)-nerolidol | +++ | | | | |
| 4 | unidentified diterpene | tr | | | | |
| 5 | β-springene | tr | tr | tr | tr | tr |
| 6 | (*E*,*E*,*E*)-neocembrene | ++ | tr | | | |
| 7 | unidentified diterpene | tr | tr | tr | tr | tr |

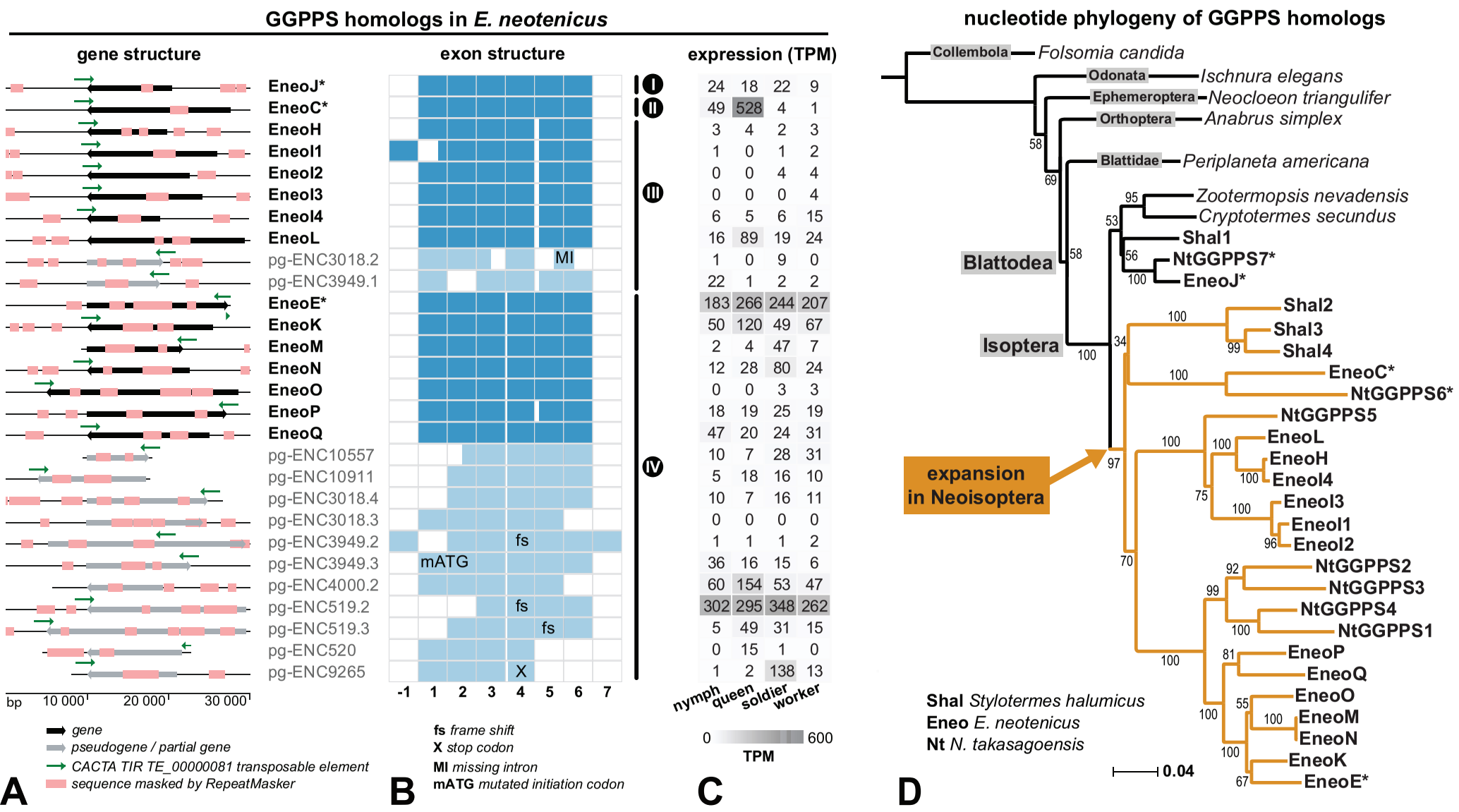
**B**

| | | soldier |
|---|---|---|
| a | (+)-α-pinene | +++ |
| b | limonene | + |
| c | 14α-acetoxy-3β,12β-dipropionoxy-kampa-6(7),8(9)-diene | ++ |
| d | 3α,9β,13α-tripropionoxy-trinervita-11(12),15(17)-diene | +++ |
| e | 3α, 9β,13α-tripropionoxy-11α(12β)-epoxy-trinervita-15(17)-ene | ++ |

Scale: +++, >60%; ++, 30–60%; +, 1–30%; tr, <1% (relative to max. peak)

Click here to access/download;Figure;FIGURE 1.eps

Figure 2




Click here to access/download;Figure;FIGURE 2.eps

Figure 3

Click here to access/download;Figure;FIGURE 3.eps

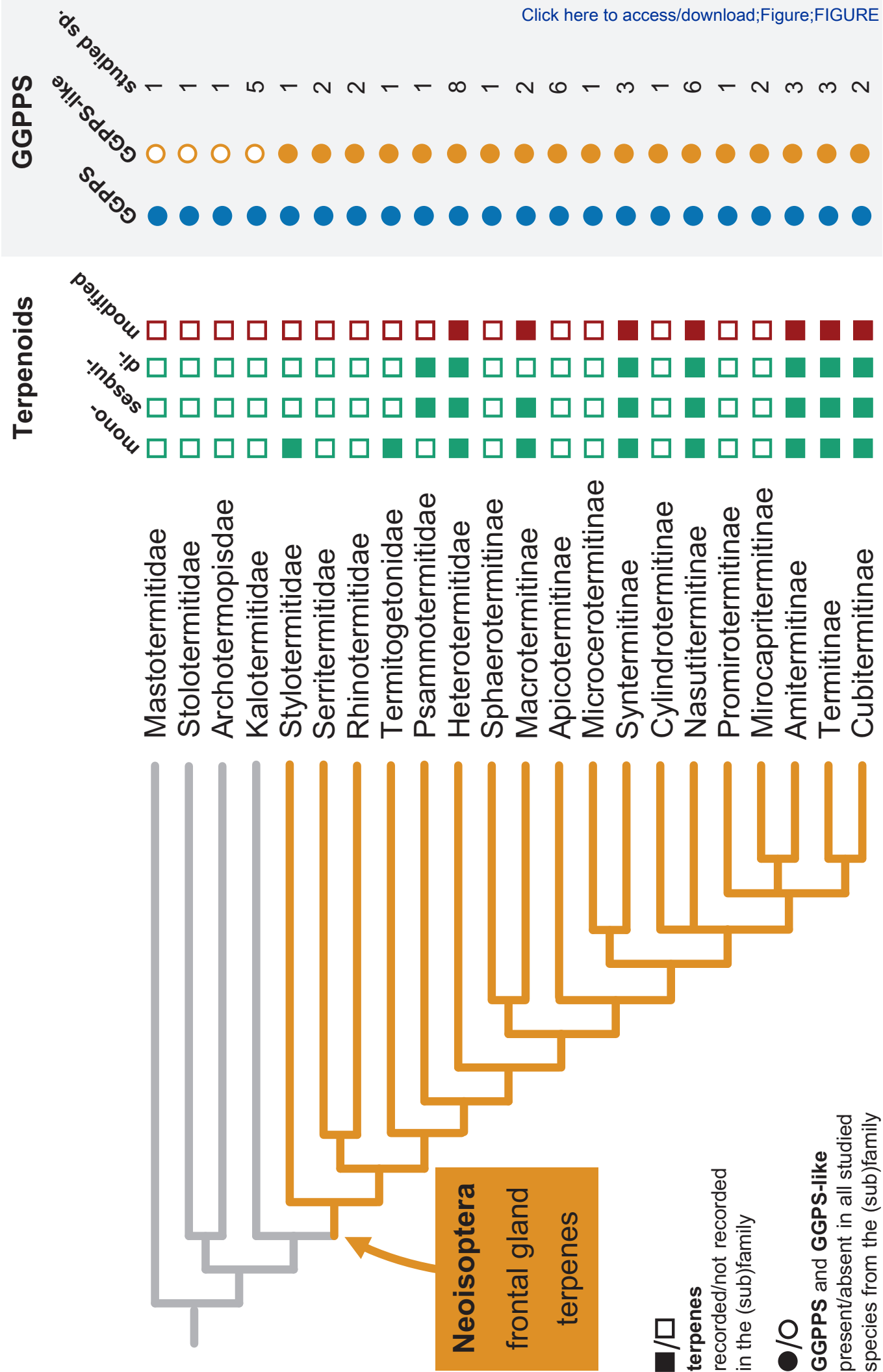


Figure 4

Click here to access/download;Figure;FIGURE 4.eps

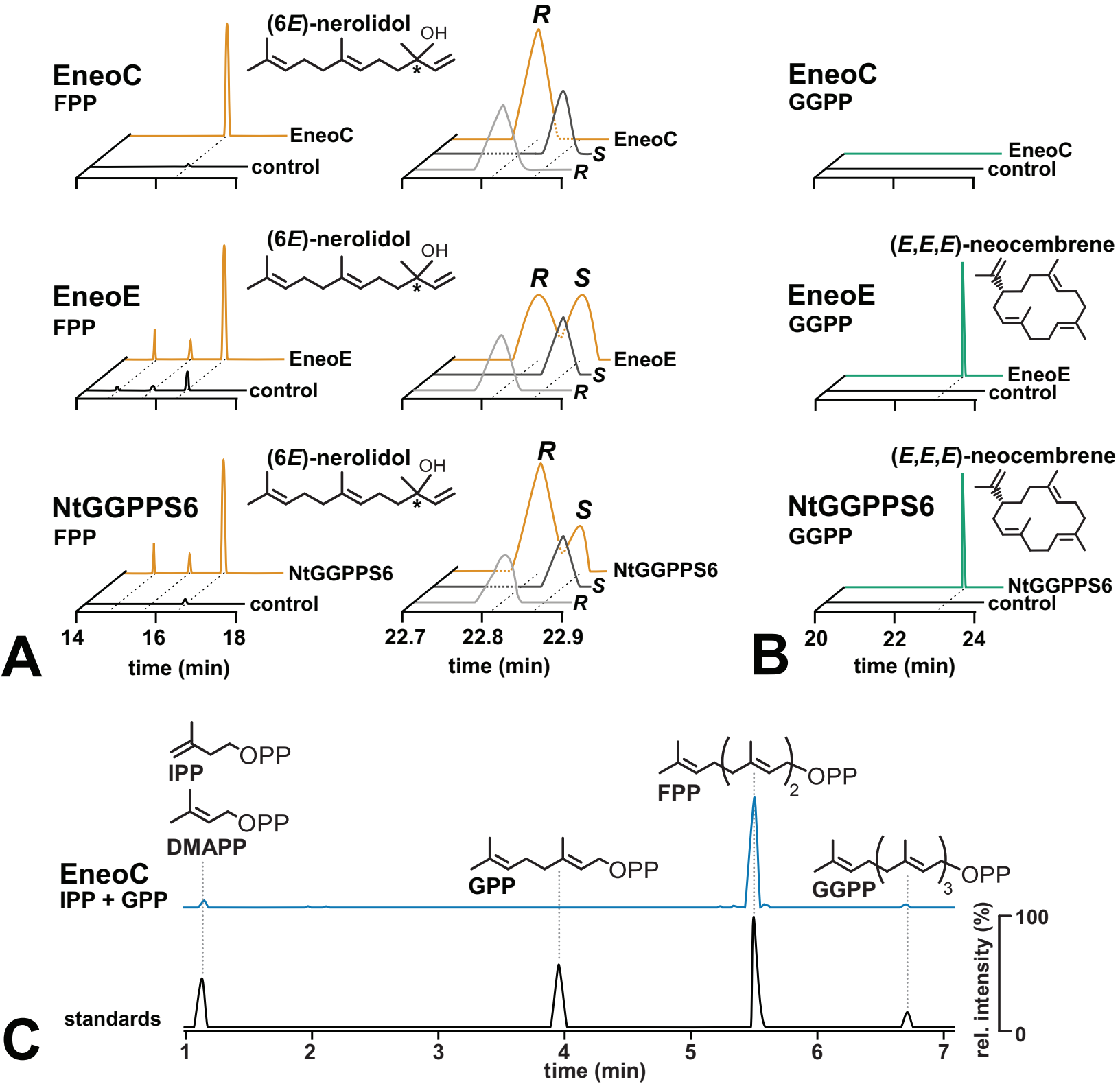
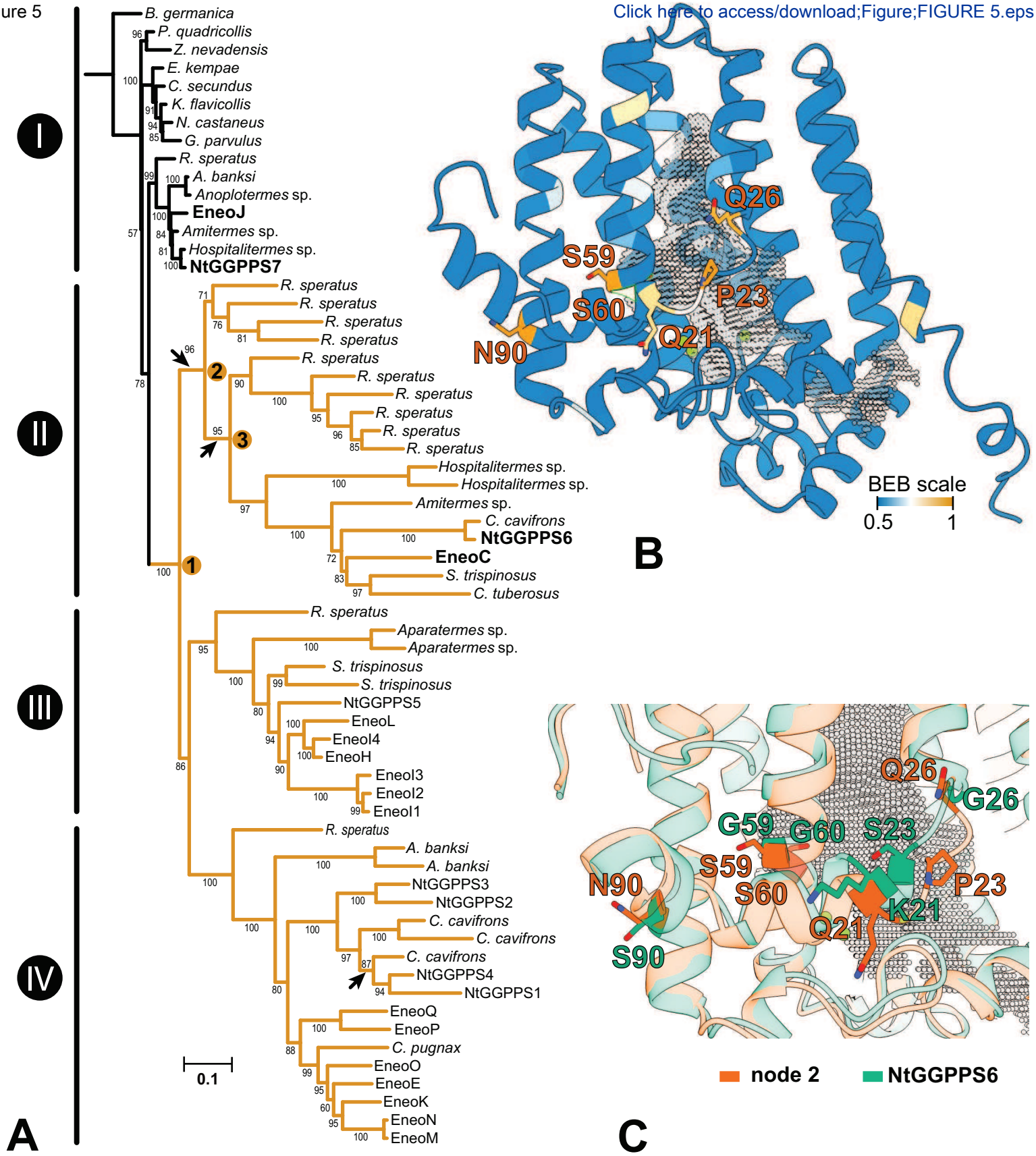


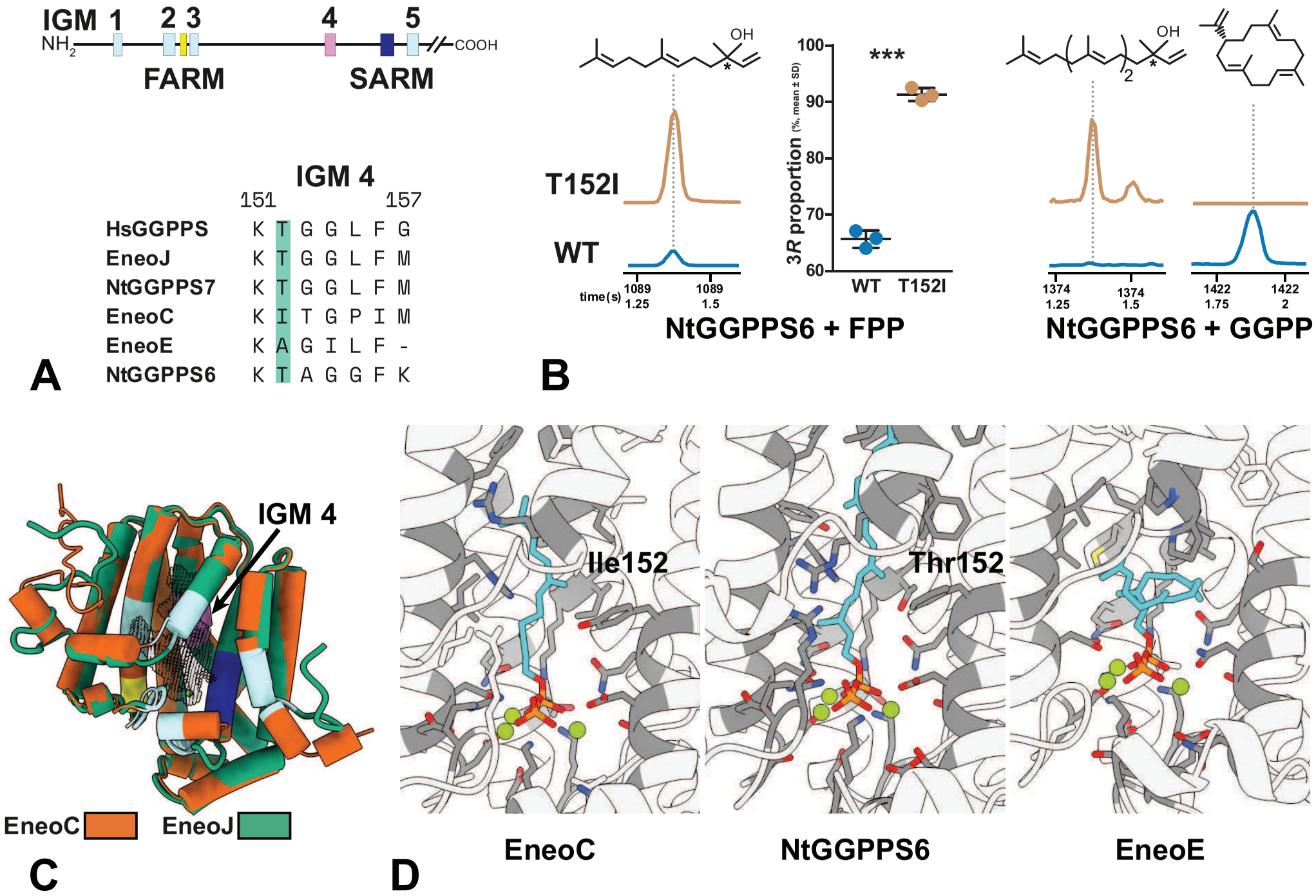
Figure 5

Click here to access/download;Figure;FIGURE 5.eps




D

| | Protein | 21 | 24 | 29 | 54 | 61 | 64 | 68 | 70 | 75 | 90 | 151 | 157 | 184 | 192 | 200 | 207 | 291 | 296 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | IGM-1 | | IGM-2 | | FARM | | IGM-3 | | | IGM-4 | | SARM | | IGM-5 | | IGM-6 | |
| IDS | EneoJ | Q V P | G K Q I R A | | Q M L H N S S L | | D D I E D | | S I L R R G | | N | K T G G L F M | | F Q I R D D Y C N | | E N K S Y C E D | | L L S W K R | |
| IDS | NtGGPPS7 | Q V P | G K Q I R A | | Q M L H N S S L | | D D I Q D | | S I L R R G | | N | K T G G L F M | | F Q I R D D Y C N | | E N K S Y C E D | | L L S W K R | |
| | node 1 | Q V P | G K Q I R A | | Q M L H N S S L | | D D I Q D | | S I L R R G | | N | K T G A L F M | | F Q I R D D Y C N | | E D K G Y C E D | | L L N W K R | |
| | node 2 | Q V P | G K Q I R A | | K M L H N S S L | | D D I Q D | | S I L R R G | | N | K T G A A F M | | F Q I R D D Y C N | | Q D K S Y C E D | | L L N W K K | |
| | node 3 | K A S | G K G I R A | | K M L H N G I L | | D D I Q D | | S I L R R G | | S | K T G A A F M | | F Q I R D D Y C N | | Q D K S Y C E D | | L L N W K K | |
| TPS | EneoC | K G S | T K G F R A | | Q M I H V C I I | | D D I Q D | | S S L R R G | | S | K I T G P I M | | Y Q I R D D Y C N | | Q K K G Y G E D | | M L N W K R | |
| TPS | NtGGPPS6 | K A S | C V G I R A | | E I Y H V G G L | | D D I Q D | | S F M R R G | | S | K T A G G F K | | F Q I R D D Y C S | | Q K K G Y L E D | | M L N W K N | |
| | BEB scale | | | | | | | | | | | | | | | | | | |




Click here to access/download;Figure;FIGURE 6.eps