Duplication and soldier-specific expression of *geranylgeranyl diphosphate synthase* genes in a nasute termite *Nasutitermes takasagoensis*

Masaru Hojo[a,*], Shuji Shigenobu[b], Kiyoto Maekawa[c], Toru Miura[d], Gaku Tokuda[a]

[a] Tropical Biosphere Research Center, University of the Ryukyus, Okinawa 903-0213, Japan

[b] Core Research Facilities, National Institute for Basic Biology, Okazaki, 444-8585, Japan

[c] Graduate School of Science and Engineering, University of Toyama, Toyama 930-8555, Japan

[d] Misaki Marine Biological Station, School of Science, The University of Tokyo, Miura, Kanagawa 238‐0225, Japan

[*]Correspondence author: masaruhojo0911@gmail.com

Tel. +81-80-8097-1620

Running title: Duplication and soldier-specific expressions of *GGPPSs* in a nasute termite

ABSTRACT

In the evolutionarily-derived termite subfamily Nasutitermitinae (family Termitidae), soldiers defend their nestmates by discharging polycyclic diterpenes from a head projection called the "nasus." The diterpenes are synthesised in the frontal gland from the precursor geranylgeranyl diphosphate (GGPP), which is generally used for post-translational modification of proteins in animals. In this study, we constructed a comprehensive gene catalogue to search for genes involved in the diterpene biosynthesis by assembling RNA sequencing reads of *Nasutitermes takasagoensis*, identifying eight gene copies for *GGPP synthase* (*GGPPS*). The number of gene copies is much larger in contrast to other related insects. Gene cloning by reverse transcription-PCR and rapid amplification of cDNA ends confirmed that seven *GGPPS* genes (*NtGGPPS1 to NtGGPPS7*) have highly variable untranslated regions. Molecular phylogenetic analysis showed that the *NtGGPPS7* gene was grouped with homologs obtained from ancestral termites that have only a single copy of the gene, and the *NtGGPPS6* gene was grouped with homologs obtained from a basal lineage of termitids, in which soldiers do not synthesise diterpenes. As the sister group to this clade, furthermore, a monophyletic clade included all the other *NtGGPPS* genes (*NtGGPPS1* to *NtGGPPS5*). Expression analyses revealed that *NtGGPPS7* gene was expressed in all the examined castes and tissues, whereas all the other genes were expressed only in the soldier head. These results suggest that gene duplication followed by subfunctionalisation of the *GGPPS* genes might have accompanied the evolution of chemical defence in the nasute termite lineage.

Keywords: **chemical defence, exocrine gland, mevalonate pathway, nasus, Termitidae**

1. Introduction

In termites, a major group of eusocial insects, various castes perform specified tasks to maintain the integrity of their colonies. For example, the soldier caste is specialised to protect nestmates (Noirot, 1989; Roisin, 2000). The soldiers were the first altruistic caste and thought to be related to the origin of eusociality in termites (Roux and Korb, 2004; Tian and Zhou, 2014).

The defense by termite soldiers is classified into two types, i.e., physical and chemical defenses (Prestwich, 1984; Noirot and Darlington, 2000). In ancestral termite groups, soldiers physically defend their colony against invaders by biting using their sclerotised and enlarged mandibles. In contrast, in more derived termites, also known as Neoisoptera (Rhinotermitidae, Serritermitidae and Termitidae), soldiers possess frontal glands producing defensive substances, i.e., terpenoids and fatty acid-derived compounds (Prestwich, 1983b; Engel et al., 2009). Jirošová et al. (2017) firstly investigated metabolism to produce aliphatic nitroalkene (*E*)-1-nitropentadec-1-ene (NPD), which is the main component of defensive secretion in *Prorhinotermes simplex* (Rhinotermitidae), based on metabolomic and transcriptomic analyses. However, little is known about the biosynthesis process of defensive secretion in termites.

Nasute termites belong to the subfamily Nasutitermitinae (family Termitidae) which is the most derived and diversified group of termites, containing approximately 20% of described termite species (Krishna et al., 2013). The nasute soldier possesses a frontal projection called a ‘nasus’, from which frontal gland secretions are discharged as defensive substances, and their defensive behaviours are considered the most sophisticated among termites (Deligne et al., 1981; Prestwich, 1983a). The defensive secretion which is the mixture of viscous diterpenes and volatile monoterpenes (Prestwich,

1988) is sticky and thus effective for entangling enemies (Eisner et al., 1976), since it contains high concentrations of cembrene-derived, hydrogen-bonded, and dome-shaped polycyclic diterpenes in a monoterpene solvent (Deligne et al., 1981; Prestwich, 1984). The common biosynthetic pathway of diterpene is conserved among animals as an essential metabolic pathway, although functional studies of these chemicals were lacking in animals. The frontal-gland reservoir which is surrounded by a single-layered secretory epithelium composed of Class 1 cells, occupies a large part of soldier head, allowing soldiers to store a large amount of defensive secretion (Noirot and Quennedey, 1974; Costa-Leonardo and De Salvo, 1987; Quennedey, 1984). Tracer experiments using radio-labelled markers have demonstrated that *Nasutitermes* soldiers synthesise diterpenes *de novo* without any metabolic assistance from workers (Singh and Prestwich, 1986; Prestwich et al., 1981). Coupled with the anatomical observations of the frontal gland described above, these imply that diterpenes are synthesised in the Class 1 cells (Prestwich, 1979a; Deligne et al., 1981; Quennedey, 1984).

All isoprenoids are synthesised from a five-carbon precursor isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) (Cane, 1999). The IPP synthesis pathway in animals, known as the mevalonate pathway, begins with acetyl coenzyme A (acetyl-CoA) (Goldstein and Brown, 1990; Kuzuyama, 2002). Geranylgeranyl diphosphate (GGPP), a common precursor of diterpene, is synthesised by GGPP synthase (GGPPS), which catalyses the condensation of farnesyl diphosphate (FPP) with IPP (Ogura and Koyama, 1998; Wang and Ohnuma, 2000; Croteau et al., 2000) (Supporting Information Fig. S1). GGPP is generally used for the protein prenylation in animals, which is important for protein–protein interactions and protein–membrane interactions (Zhang and Casey, 1996). Therefore, the GGPP synthesis pathway

used for protein prenylation is thought to be ubiquitous in any cell types in termites. After the GGPP synthesis, via cembrene-like intermediates through cyclisation steps, hydroxylation and/or oxidation by oxidoreductase such as cytochrome P450 (CYP) generates the characteristic diterpene secretion (Prestwich, 1979a).

GGPPS involved in the protein prenylation is generally encoded by a single gene copy in insects and mammals (Ericsson et al., 1998; Lai et al., 1998; Kainou et al., 1999; Kuzuguchi et al., 1999; Zhao et al., 2000). Conversely, in *N. takasagoensis*, three *GGPPS* genes and some other partial *GGPPS* sequences have been identified based on Sanger sequencing and 454 pyrosequencing (Hojo et al., 2007, 2012; Hayashi et al., 2013). Therefore, we hypothesised that gene duplication followed by subfunctionalisation or neofunctionalisation of the *GGPPS* genes should be important for the evolution of diterpene biosynthesis in nasute termites.

In a nasute termite *N. takasagoensis*, the genes involved in the diterpene biosynthesis were identified based on a comprehensive gene catalogue constructed by *de novo* assembly of RNA sequencing (RNA-seq) reads. Then, quantitative real-time polymerase chain reaction (qRT-PCR) was performed to characterise the expression patterns of genes related to the diterpene synthesis in various castes. Based on the obtained results, the evolution of chemical defensive strategies in termites is discussed.

2. Materials and methods

2.1. Termite sample

Nests of *N. takasagoensis* were collected from Iriomote Island and Ishigaki Island, Okinawa, Japan. The nests were kept in an air-conditioned room maintained at approximately 25°C until use.

115

116 *2.2. RNA extraction and cDNA library construction*

117 For RNA-seq analysis, total RNA was extracted from 50 mg of heads derived from
118 each of the following castes: soldiers, presoldiers, minor workers and male alates (winged
119 adults) in the same colony. The heads of soldiers were divided into two parts (Fig. S2):
120 one comprising the Class 1 cell layer of the frontal gland (library 1), which is thought to
121 synthesise diterpene defensive secretions (Deligne et al., 1981; Quennedey, 1984), and
122 another comprising the remaining tissues (library 2), including Class 3 gland cells, the
123 brain and the suboesophageal ganglion (Fig. S2). For the heads of presoldiers (library 3),
124 minor workers (library 4) and male alates (library 5), whole tissues were used for the
125 library preparation. For comparison, we also examined the head tissues (including the
126 frontal gland Class 1 cell layer) of soldiers from another colony (library 6). RNA
127 extraction was performed using an RNeasy mini Kit with DNase treatment according to
128 the manufacturer's instruction (Qiagen, Hilden, Germany). The amounts of RNA and
129 DNA contamination in each sample were quantified using a Qubit fluorometer (Thermo
130 Fisher Scientific, Waltham, MA, USA). The quality of RNA was validated using an
131 Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

132

133 *2.3. RNA sequencing*

134 Following the procedure described by Hayashi et al. (2017), we used TruSeq RNA
135 sample preparation kits (Version 2) (Illumina, San Diego, CA, USA) to generate a
136 complementary DNA (cDNA) library from 1 μg of total RNA from each caste or tissue.
137 The cDNA libraries were individually indexed for multiplex sequencing before being

138 pooled and sequenced on a HiSeq 2000 platform (Illumina) with 101 bp paired-end
139 reading.
140
141 *2.4.* De novo *transcriptome assembly and expression analysis*
142 We trimmed the low-quality ends (<QV30) and adapter sequences using the cutadapt
143 programme (Martin, 2011), whereas the reads shorter than 50 bp were discarded. The
144 filtered reads obtained from the five cDNA libraries derived from the same termite colony
145 were pooled and subjected to *de novo* transcriptome assembly using the Trinity software
146 (Grabherr et al., 2011). Open reading frames (ORFs) were predicted using the
147 TransDecoder programme (Trinity package). The predicted ORF sequences were
148 grouped using the cd-hit-est programme (Li and Godzik, 2006) with a minimum identity
149 of 97%, from which we obtained a non-redundant (nr) coding sequence set. A sequence
150 similarity search was performed against the NCBI's nr protein database (ver. June 2018)
151 with BLASTp using predicted proteins as queries.
152 When we evaluated the Trinity assembly, we noticed that the contigs derived from
153 different *GGPPS* gene paralogs were erroneously assembled together; this was probably
154 because the *GGPPS* gene paralogs are so similar to each other that Trinity assembling
155 with a k-mer size = 25 collapsed paralogs sharing sequences longer than 25 base pairs.
156 To overcome this issue, we used the velvet programme (ver. 1.2.07) (Zerbino and Birney,
157 2008) for assembling *GGPPS* gene paralogs because velvet programme allows a flexible
158 setting of the parameter of k-mer size as longer values. We collected the Illumina reads
159 derived from *GGPPS* gene paralogs and attempted assembling using the velvet
160 programme with various k-mer sizes ranging from 75 to 99. We found that the *GGPPS*

161 ORF assembly produced by a parameter of k-mer size = 95 was the best, and we used it
162 for the downstream analyses.
163 For differential expression analysis, short reads were mapped to the reference ORF
164 sequences using Bowtie2 software (Langmead and Salzberg, 2012). Transcript
165 abundances were estimated using eXpress software (Roberts and Pachter, 2013). To
166 adjust the library sizes and skewed expressions of transcripts, we normalised the
167 estimated abundance values using the trimmed mean of M values (TMM) normalisation
168 method (Robinson and Oshlack, 2010). Differential expression analysis was conducted
169 using the edgeR statistical package (Robinson et al., 2010).
170
171 *2.5. Identification of genes involved in diterpene biosynthesis and cDNA cloning*
172 We used the tBLASTx, with the nucleotide sequences of *Acyrthosiphon pisum* as
173 queries, to search the reference ORF sequences of *N. takasagoensis* for the genes involved
174 in diterpene biosynthesis. For confirming the accurate assembly, we determined cDNA
175 sequences, including the untranslated region (UTR) of the genes of interest, as follows.
176 Total RNA was extracted from soldier heads as described above. First-stranded cDNA
177 was synthesised using SuperScript III First-Strand Synthesis SuperMix (Thermo Fisher
178 Scientific). Each cDNA sequence was obtained using RT-PCR and rapid amplification of
179 cDNA ends (RACE) with gene-specific primers (Table S1). PCR products were purified
180 and nucleotide sequences were determined with a Big Dye Terminator v3.1 Cycle
181 Sequencing Kit using a 3130xl Genetic Analyzer (Thermo Fisher Scientific).
182
183 *2.6. Amino acid sequences alignment, phylogenetic analysis and protein structure*
184 *prediction*

185 Using amino acid sequences of *N. takasagoensis* as queries, we used tBLASTx to
186 search for homologous genes of the GGPP synthesis pathway genes in closely related
187 insect species, such as *Blattella germanica* Linnaeus, *Zootermopsis nevadensis* Hagen,
188 *Cryptotermes secundus* Hill, and *Macrotermes natalensis* Haviland, whose genome
189 sequences were published (Terrapon et al., 2014; Poulsen et al., 2014; Harrison et al.,
190 2018). We used MEGA7 software (Kumar et al., 2016) to perform sequence alignments
191 and construct a phylogenetic tree of the *GGPPS* genes obtained from termites and other
192 insects and organisms. Sequences were aligned using the MUSCLE algorithm.
193 Phylogenetic relationships were inferred using the Maximum Likelihood method based
194 on an LG model (Le and Gascuel, 2008). We obtained initial phylogenetic trees for the
195 heuristic search by applying the Neighbour-Joining method (Saitou and Nei, 1987). To
196 infer the structural differences among GGPPSs, the protein structures were predicted
197 based on the deduced amino acid sequences using Phyre2 web site (Kelley et al., 2015)
198 with *Z. nevadensis* GGPPS amino acid sequences (XP_021933767) as a reference.
199
200 *2.7. Relative expression analysis by qRT-PCR*
201 RNA samples were separately extracted from head and body tissues of the four
202 terminal castes (i.e. male soldier, female major worker, male alate and female alate). Each
203 biological sample was replicated from four different colonies. We determined the quality
204 and quantity of the extracted RNA through spectroscopic measurements at 230, 260 and
205 280 nm using the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). The
206 total RNA was transcribed using the PrimeScript RT reagent Kit with gDNA Eraser
207 (Takara Bio, Shiga, Japan), according to the manufacturer's instructions. The qRT-PCR
208 procedure was performed using a Fast SYBR Green Master Mix (Thermo Fisher

Scientific) and a StepOnePlus Real-Time PCR system (Thermo Fisher Scientific). We also evaluated the suitability of three reference genes [i.e. *beta-actin* (AB501107), *EF1a* (AB501108) and *glyceraldehyde 3-phosphate dehydrogenase* (LC440333, determined in this study using Sanger sequencing)]. These reference genes were used to determine the endogenous control of constitutive expression using the following software: geNorm (Vandesompele et al., 2002), BestKeeper (Pfaffl et al., 2004) and NormFinder (Andersen et al., 2004). Primers for target genes were designed considering each open reading frame using Primer Express software (Thermo Fisher Scientific) (forward and reverse primers are underlined and double-underlined, respectively in Fig. S3). Sequences and optimum concentrations of primers are presented in Table S1. The thermal cycle programme comprised 20 s at 95°C followed by 50 cycles of 3 s at 95°C and 30 s at 60°C. The production of single products was confirmed by dissociation curve analysis. The thermal cycle programme in the dissociation stage comprised 15 s at 95°C, 1 min at 60°C and a slow ramp up of temperature to 95°C. The data were analysed using the relative standard curve method. The cycle threshold values of the analysed genes were normalised using the values of reference genes. The data were statistically analysed by one-way ANOVA in conjunction with post-hoc Tukey's HSD tests.

3. Results

3.1. RNA sequencing and assembly

We created RNA-seq libraries to construct a comprehensive catalogue of genes expressed in the heads of male castes of *N. takasagoensis*. We sequenced the six libraries using the Illumina HiSeq platform, which yielded 61,154,873 paired-end sequence reads (Tables 1 and S2). All raw Illumina reads were deposited in the DNA Data Bank of Japan

(DDBJ) Sequence Read Archive under the accession number DRA007712. After trimming adapter and low-quality bases, the cleaned reads of the five libraries from the main colony were used to create a comprehensive catalogue using Trinity. *De novo* assembly by Trinity generated 118,787 contigs with N50 of 2,864b, ranging from 201b to 43,151b. From these contigs, we predicted 40,863 nr ORFs with N50 of 1,668b, ranging from 150b to 38,781b (Table 1). All aforementioned ORF sequences were deposited in the DDBJ Transcriptome Shotgun Assembly (TSA) division (accession nos. IAEB01000001-IAEB01040863) and used for downstream analyses as a reference coding sequence of *N. takasagoensis*.

3.2. Genes enriched in soldier heads containing the frontal gland Class 1 cell layer

The counts per million (cpm) value, normalised by TMM, is shown in Table S3. All predicted ORF sequences were compared with the NCBI nr protein database using the Basic Local Alignment Search Tool (BLAST). In our ORF list, 20,291 ORFs were retrieved from the NCBI nr protein database with e-value thresholds of 1.0E−5 (Table S4). After using BLAST, we examined the gene enrichment in soldier head tissue containing the frontal gland Class 1 cell layer. To perform this, we compared gene expression levels between the Class 1 cell layer of the frontal gland of soldier head (library 1 and 6) and the remaining tissues of soldier head and the heads of the other castes (libraries 2–5). The 50 most abundant ORFs in this soldier head tissue are shown in Table 2; significantly abundant ORFs are shown in Table S5 (FDR < 0.05). In the ORF catalogue, nine homologous genes (IAEB01006569, IAEB01006570, IAEB01013267, IAEB01015961, IAEB01004244, IAEB01037270, IAEB01012708, IAEB01012707 and IAEB01006409) involved in the diterpene synthesis pathway and 14 putative CYP genes

(IAEB01039737, IAEB01019592, IAEB01019615, IAEB01010255, IAEB01010256, IAEB01007818, IAEB01037022, IAEB01030092, IAEB01025415, IAEB01036662, IAEB01015134, IAEB01024991, IAEB01014765 and IAEB01019613) were statistically enriched in the soldier head tissue containing the frontal gland Class 1 cell layer (FDR < 0.05) (Tables 2 and S5).

3.3. Nine genes related to diterpene synthesis in N. takasagoensis

Using the presumed nine genes involved in diterpene synthesis in *A. pisum* as queries, we used BLAST to retrieve 20 ORF contigs from our ORF list, all with e-values less than 1.0E−10 (Table S6). Among the 20 ORF contigs, there were two different sequences for *hydroxymethylglutaryl-CoA synthases (HMGSs)*, three different sequences for *FPP synthases (FPPSs)* and more than six different sequences for the *GGPPS* genes. In the ORF sequences of the diterpene synthesis pathway, the sequences of IAEB01013267 and IAEB01037270 were identical to *HMGS* and *hydroxymethylglutaryl-CoA reductase (HMGR)* genes of *N. takasagoensis*, respectively (Hojo et al., 2012: *NtHMGS*, accession no. AB733008; *NtHMGR*, AB733009). Moreover, the sequences of the IAEB01015961 were identical to the *GGPPS* gene of *N. takasagoensis* (*NtGGPPS3-B*, AB266081). However, we were unable to detect ORF sequences for the other *GGPPS* genes of *N. takasagoensis* previously determined (*NtGGPPS1-B*, AB266075; *NtGGPPS2-B*, AB266079) in our list. Next, we used the velvet programme with an increased k-mer size to construct *GGPPS* contigs using the same reads used for the *GGPPS* contigs in the Trinity assembly. As a result, nine more *GGPPS* ORFs (TSA accession nos. IAEB01040864-IAEB01040872) were constructed. The sequences of IAEB01040867,

IAEB01040866 and IAEB01040864 were identical to the sequences of *NtGGPPS1-B*, *NtGGPPS2-B* and *NtGGPPS3-B*, respectively.

To confirm the accuracy of the assembled sequences, we performed cDNA cloning to determine the nucleotide sequences with their UTRs. We determined 13 cDNA sequences related to diterpene synthesis (accession nos. LC440320-LC440332), which were named (Table 3) and deposited in the DDBJ database. Among the *GGPPS* genes, *NtGGPPS4* (AB266082) was identical to IAEB01040865, IAEB01040868, IAEB01040869 and IAEB01040871 sequences. Furthermore, *NtGGPPS5* was identical to the IAEB01012707 and very similar to IAEB01012708 sequences. *NtGGPPS6* was identical to IAEB01039951 sequences, whereas *NtGGPPS7* was identical to IAEB01006415, IAEB01006417 and IAEB01006409 sequences. Contig IAEB01006416 was partially identical to *NtGGPPS7*. The IAEB01021611 sequences differed from the other seven *GGPPS* sequences determined (Fig. S3). All these transcripts possessed the distinct UTR sequences (Fig. S3), showing they were unlikely splice variants. To summarise, our results predicted at least eight genes encoding GGPPS in the *N. takasagoensis* genome.

3.4. Number of genes related to diterpene synthesis in cockroaches and termites

The genome sequences of three termite species (*Z. nevadensis*, *M. natalensis* and *C. secundus*) and one cockroach species (*B. germanica*) have recently been reported (Terrapon et al., 2014; Poulsen et al., 2014; Harrison et al., 2018). We subjected the deduced amino acid sequences of our identified genes responsible for diterpene synthesis in *N. takasagoensis* to BLASTp against the original gene set of the other termites and one cockroach species. Our results showed that the numbers of mevalonate pathway genes

304 (Fig S1) were identical in all species examined (Table 4). However, the number of genes
305 in the following GGPP synthesis pathway differed among species: there were one or two
306 *FPPS* genes in all species except for *N. takasagoensis*, which had three. In particular, the
307 number of *GGPPS* genes was larger in *N. takasagoensis* (at least eight) compared with
308 the cockroach and other termite species examined (only one).
309
310 *3.5. Phylogenetic relationship of the* GGPPS *genes in termites*
311 Phylogenetic analysis indicated the termite *GGPPS* genes are divided into two
312 clades (Fig. 1); one clade containing homologous genes identified from all species
313 examined (ancestral clade) and the other clade containing genes identified only from
314 termitid species (expanded clade). In the latter clade, genes could be further divided into
315 two clades; one included the *GGPPS* genes of *N. takasagoensis* (*NtGGPPS6*) and *M.*
316 *natalensis*, and the other included all the other genes only obtained from *N. takasagoensis*
317 (*NtGGPPS1 to NtGGPPS5*).
318
319 *3.6. Structures of GGPPS proteins of* N. takasagoensis
320 The amino acid sequence alignment of seven ORF determined by cloning indicated
321 that two aspartate-rich regions, involved in the binding of the diphosphate moieties of the
322 allylic substrate through Mg^{2+} bridges (Liang et al., 2002), were conserved in all
323 NtGGPPS proteins with some substitutions (Fig. 2). The amino acid residues at the fifth
324 position upstream of the first aspartate-rich region (FARM) in all NtGGPPS proteins were
325 neutral residues with relatively smaller molecular masses, such as serine, alanine and
326 glycine, typically found in GGPPSs (Ohnuma et al., 1996) (Fig. 2). One of the inserted

residues in FARM is normally glutamic acid in eukaryotic GGPPSs (Bouvier et al., 2005), whereas those of NtGGPPSs (except for NtGGPPS1-B) were found to be different.

The 3D structure of GGPPSs of *N. takasagoensis* and *Z. nevadensis* predicted from the deduced amino acid sequence was shown in Fig. S4. Although the structure of the examined GGPPSs was well conserved between these termite species, slight conformational variations were recognised in NtGGPPS4 (Fig. S4).

3.7. Relative expression of nine genes related to diterpene synthesis

We performed qRT-PCR analysis using the four terminal castes (i.e. male soldier, female major worker, male alate and female alate) to compare the relative expression of genes related to diterpene synthesis. In these comparisons, the expression level of *elongation factor 1 alpha* (*EF1a*) was the most stable among the three candidate reference genes [i.e. *beta-actin* (AB501107), *EF1a* (AB501108) and *glyceraldehyde 3-phosphate dehydrogenase* (*GAPDH*); Table S7].

Gene expression analysis indicated that only a gene copy (*NtHMGS1*, *NtFPPS1* and *NtGGPPS7*) was universally expressed in each of the gene groups that possessed multiple gene copies (i.e., *HMGS*, *FPPS* and *GGPPS*), whereas the other genes (*NtHMGS2*, *NtFPPS2*, *NtFPPS3* and *NtGGPPS1* to *NtGGPPS5*) were expressed only in soldier heads [Fig. 3, one-way analysis of variance (ANOVA) in conjunction with post-hoc Tukey's HSD tests, $p < 0.05$]. However, the higher expression levels of *NtGGPPS6* in the soldier head was not supported by statistical tests. The other single copy genes were essentially expressed at the highest levels in soldier heads.

4. Discussion

351 *4.1. Predicted ORFs obtained from male caste head tissues*

352 We successfully constructed a gene catalogue expressed in frontal gland Class 1
353 cells, because the contig of *soldier specific protein 1* (IAEB01010718) was listed among
354 the abundant ORFs from soldier-head tissues containing frontal gland cells (Table S5),
355 which is highly expressed in Class 1 cells of soldier frontal gland in a previous study
356 (Hojo et al., 2005). Since the genomic and/or large-scale transcriptomic analyses of
357 termitid species, especially nasute termites, have not yet been performed in detail, the
358 transcriptome information provided here would be helpful to identify genes of interests,
359 including those related to the biosynthetic pathway of diterpenoid secretions.

360 Among the abundant ORFs in the soldier head containing frontal gland tissues, many
361 CYP genes were identified (Tables 2 and S5), that encode the terminal oxidases
362 contributing to the chemical diversity in all kingdoms. In higher plants, the copy numbers
363 of CYP genes are extremely high (e.g. 245 in *Arabidopsis thaliana*, 343 in *Oryza sativa*;
364 Nelson and Werck-Reichhart, 2011) in contrast to other organisms. In these plants, the
365 CYP gene copies are suggested to contribute to diversified structures and functions of
366 terpenoids (Hamberger and Bak, 2013). In a rhinotermitid termite *P. simplex*, in which
367 soldiers produce NPD as a defensive component, 10 gene candidates related to CYP were
368 highly expressed in soldier frontal gland (Jirošová et al., 2017, 2018). In *Nasutitermes*
369 species, more than 30 diterpene compounds have been reported in frontal gland secretions
370 (Prestwich, 1979b; Šobotník et al., 2010), so that in *N. takasagoensis*, frontal gland
371 secretions is thought to comprise diterpene compounds with various structures. These
372 suggest that the oxygenation by CYPs may contribute to the diterpene diversity. To clarify
373 this possibility, functional analyses of the CYP genes identified in this study will be
374 required in future studies.

375

376 *4.2. Gene duplication and functionalisation of* GGPPS

377 GGPP, a common precursor of diterpenes, is used for C20-prenylation, which is
378 essential for the localisation and activation of proteins in animals (Zhang and Casey,
379 1996). The mammalian and *Drosophila melanogaster GGPPS* genes are ubiquitously
380 expressed and coded by a single gene (Ericsson et al., 1998; Lai et al., 1998). However,
381 multiple gene copies encoding GGPPS were identified in *N. takasagoensis*. Furthermore,
382 all the *GGPPS* genes previously described in *N. takasagoensis* were expressed only in
383 the soldier caste (Hojo et al., 2007, 2012). Among animals, diterpene biosynthesis has
384 been assumed only in a few marine invertebrates besides nasute termites (Gavagnin and
385 Fontana, 2000; Aratake et al., 2012; Farag et al., 2017), although the evolution of the
386 diterpene biosynthesis pathway in animals has yet to be elucidated. Since theoretical
387 simulation showed that gene duplication leads to subfunctionalisation, followed by
388 neofunctionalisation (Rastogi and Liberles, 2005), multiple copies of *GGPPS* gene in *N.*
389 *takasagoensis* are suggested to drive subfunctionalisation and/or neofunctionalisation of
390 *GGPPS* genes in the termite lineage. As an example of neofunctionalisation, genes coding
391 luciferases in fireflies are suggested to have evolved from a peroxisomal fatty acetyl-CoA
392 synthetase gene after tandem duplications (Fallon et al., 2018). In future studies, however,
393 functional assays on the enzymatic activities of GGPPSs that are coded by different gene
394 copies should be required.

395

396 *4.3. Evolution of the termite* GGPPS *genes*

397 Whole genome sequences in some termite and cockroach species revealed that the gene
398 duplication of *GGPPS* has occurred only in derived termite taxa (Nasutitermitinae and

399 Macrotermitinae; Table 4). Although the genome of a flowering plant *A. thaliana*
400 possesses 10 functional *GGPPS* genes that were differentially expressed in various
401 organelles, tissues and/or developmental stages (Beck et al., 2013), in *N. takasagoensis,*
402 all the identified GGPPS were cytosolic (estimated from TargetP,
403 http://www.cbs.dtu.dk/services/TargetP/, Emanuelsson et al., 2000). Additionally, all the
404 seven NtGGPPS had aspartate-rich regions (Fig. 2), suggesting that these enzymes bind
405 the allylic substrate and pyrophosphate of IPP. The amino acid residues at the fifth
406 position upstream of the first aspartate-rich region of these sequences were smaller
407 residues, which are probably involved in the determination of product length (Tarshis et
408 al., 1996; Ohnuma et al., 1996). These proteins must have a function similar to that of the
409 plant GGPPSs, which could convert FPP with IPP to 20-carbon isoprenoid products
410 (Bouvier et al., 2005; Wang et al., 2016). All the expanded *GGPPS* genes (*NtGGPPS1* to
411 *NtGGPPS6*) were expressed specifically in the soldier heads (Fig. 3), suggesting that
412 these genes were acquired by gene duplication in the lineage of nasute termites, and play
413 roles in the synthesis of defensive diterpenes. However, the *NtGGPPS6* was slightly
414 expressed in the bodies of soldiers, major workers and female alates (Fig. 3). In nasute
415 termite, nonoxygenated diterpene hydrocarbons neocembrene and trinerviatriene are used
416 for trail pheromones in workers and soldiers and for sex pheromones in female alates
417 (Buděšínský et al., 2005; Sillam-Dussès et al., 2010), so that *NtGGPPS6* may have
418 distinct functions in the synthesis of these pheromones in other body parts and in other
419 castes.

420 The fact that only *NtGGPPS7* was expressed at any tissues in any castes (Fig. 3)
421 suggests that *NtGGPPS7* likely has essential house-keeping functions like protein
422 prenylation, as reported by Zhang and Casey (1996). Phylogenetic analyses also

supported this idea, in which *NtGGPPS7* was located at the basal position to a clade of expanded *GGPPS* genes (Fig. 1).

Also, the phylogenetic results suggest that extensive gene duplication have occurred in the derived termite lineage, i.e., Termitidae. Although soldiers in some Macrotermitinae species have a frontal gland, diterpene defensive compounds have not been detected from the frontal glands (Prestwich, 1984). Moreover, it was also shown that the *GGPPS* genes have duplicated multiple times in *Nasutitermes* (Fig. 1). Considering that *NtGGPPS1* to *NtGGPPS5* genes were expressed only in the heads of soldier termites (Fig. 3), the subfunctionalisation of *GGPPS* genes is suggested to be involved in the acquisition of diterpene synthesis in the lineage of *Nasutitermes* species. However, 3D structures of all GGPPSs of *N. takasagoensis* predicted were very similar (Fig. S4), suggesting that the enzymatic functions are not largely different between ubiquitous and soldier-specific expressed GGPPSs. Nevertheless, the possibility that the slight conformational variations observed in NtGGPPS4 or other trivial structural differences contributing to the functional diversification cannot be ruled out. Further biochemical analyses should be performed to elucidate the key difference for each function among GGPPSs obtained in this study.

4.4. Novel role of GGPP synthesis pathway genes in termites

In our ORF catalogue, 8 genes that encode all the enzymes responsible for the early steps the GGPP synthesis (*ACAT2*, *HMGS*, *HMGR*, *MevK*, *MevPK*, *MevPPD*, *IPPI* and *FPPS*) were identified. Among these, *HMGS* and *FPPS* were encoded by multiple gene copies. For *HMGS,* there were two gene copies in any insect species, while the copy number of *FPPS* genes varied among insect species (Table 4). Our gene expression results

447 indicated that one *HMGS* gene (*NtHMGS2*) and two of the *FPPS* genes (*NtFPPS2* and
448 *NtFPPS3*) were highly expressed in soldier heads (Fig. 3), suggesting that the soldier-
449 specific expressions are also involved in the diterpene biosynthesis.

450 Isoprenoids (including GGPP) play various roles in living organisms (Zhang and
451 Casey, 1996). In *N. takasagoensis,* the *ACAT2* gene which encodes the enzyme at the
452 initial step of the isoprenoid synthesis, was highly expressed in soldier heads (Fig. 3).
453 This suggests that the diterpenoid synthesis for defensive secretions takes place from the
454 initial step of the isoprenoid synthesis in soldier heads. All the single-copy genes were
455 expressed in all termite castes, suggesting that they possess the roles of both the primary
456 metabolism and the diterpenoid synthesis for defensive secretion.

457

458 *4.5. Conclusions*

459 The findings of this study suggest that the gene duplication and the following
460 subfunctionalisation of *GGPPS* gene may contribute to the acquisition of defensive
461 substances in a termite lineage. Future works by further identifications and
462 characterisations of the *GGPPS* gene in a variety of termite lineages, particularly in
463 species possessing soldiers with frontal glands, may enable us to determine the timings
464 of *GGPPS* gene duplication events and the evolution of chemical defense in termite
465 lineages.

466

467 **Acknowledgements**

468 We thank K. Yamaguchi and Y. Hayashi for technical support in Illumina
469 sequencing. This work was financially supported by a JSPS KAKENHI (No.15K07798)
470 to M.H. and S.S. provided by the Ministry of Education, Culture, Sports, Science and

471 Technology, Japan. This work was also supported by NIBB Collaborative Research
472 Program (No. 18-332) to M. H.
473

474 **References**

475 Andersen, C.L., Jensen, J.L., Ørntoft, T.F., 2004. Normalization of real-time quantitative
476 reverse transcription-PCR data: a model-based variance estimation approach to
477 identify genes suited for normalization, applied to bladder and colon cancer data sets.
478 Cancer Res. 64, 5245–5250.
479 Aratake, S., Tomura, T., Saitoh, S., Yokokura, R., Kawanishi, Y., Shinjo, R., Reimer,
480 J.D., Tanaka, J., Maekawa, H., 2012. Soft coral *Sarcophyton* (Cnidaria: Anthozoa:
481 Octocorallia) species diversity and chemotypes. PLOS ONE 7, e30410.
482 Beck, G., Coman, D., Herren, E., Ruiz-Sola, M.Á., Rodríguez-Concepción, M., Gruissem,
483 W., Vranová, E., 2013. Characterization of the *GGPP* synthase gene family in
484 *Arabidopsis thaliana*. Plant Mol. Biol. 82, 393–416.
485 Bouvier, F., Rahier, A., Camara, B. 2005. Biogenesis, molecular regulation and function
486 of plant isoprenoids. Prog. Lipid Res. 44, 357–429.
487 Buděšínský, M., Valterová, I., Sémon, E., Cancello, E., Bordereau, C., 2005. NMR
488 structure determination of (11*E*)-trinervita-1(14),2,11-triene, a new diterpene from
489 sexual glands of termites. Tetrahedron 61, 10699–10704.
490 Cane, D.E., 1999. Isoprenoid biosynthesis: overview. In: Cane, D.E. (Ed.),
491 Comprehensive Natural Products Chemistry, Vol. 2. Pergamon Press, Oxford, pp.
492 1–13.

493 Costa-Leonardo, A.M., De Salvo, C.R., 1987. A comparative study of the frontal glands
494 in three species of Brazilian termite soldiers (Isoptera, Termitidae). Rev. Bras. Ent.
495 31, 465–471.
496 Croteau, R., Kutchan, T.M., Lewis, N.G., 2000. Natural products (secondary metabolites).
497 In: Buchanan, B., Gruissem, W., Jones, R. (Eds.), Biochemistry & Molecular
498 Biology of Plants. American Society of Plant Physiologists, Rockville, pp. 1250–
499 1318.
500 Deligne, J., Quennedey, A., Blum, M.S., 1981. The enemies and defense mechanisms of
501 termites. In: Hermann, H.R. (Ed.), Social Insects, Vol. II. Academic Press, New
502 York, pp. 1–76.
503 Eisner, T., Kriston, I., Aneshansley, D.J., 1976. Defensive behavior of a termite
504 (*Nasutitermes exitiosus*). Behav. Ecol. Sociobiol. 1, 83–125.
505 Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000. Predicting subcellular
506 localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol.
507 300, 1005–1016.
508 Engel, M.S., Grimaldi, D.A., Krishna, K., 2009. Termites (Isoptera): their phylogeny,
509 classification, and rise to ecological dominance. Am. Mus. Novit. 3650, 1–27.
510 Ericsson, J., Greene, J.M., Carter, K.C., Shell, B.K., Duan, D.R., Florence, C., Edwards,
511 P.A., 1998. Human geranylgeranyl diphosphate synthase: isolation of the cDNA,
512 chromosomal mapping and tissue expression. J. Lipid Res. 39, 1731–1739.
513 Fallon, T.R., Lower, S.E., Chang, C.-H., Bessho-Uehara, M., Martin, G.J., Bewick, A.J.,
514 Behringer, M., Debat, H.J., Wong, I., Day, J.C., Suvorov, A., Silva, C.J., Stanger-
515 Hall, K.F., Hall, D.W., Schmitz, R.J., Nelson, D.R., Lewis, S.M., Shigenobu, S.,

516 Bybee, S.M., Larracuente, A.M., Oba, Y., Weng, J.K., 2018. Firefly genomes
517 illuminate parallel origins of bioluminescence in beetles. elife 7, e36495.
518 Farag, M.A., Al-Mahdy, D.A., Meyer, A., Westphal, H., Wessjohann, L.A., 2017.
519 Metabolomics reveals biotic and abiotic elicitor effects on the soft coral *Sarcophyton*
520 *ehrenbergi* terpenoid content. Scientific Reports 7, 648.
521 Gavagnin, M., Fontana, A., 2000. Diterpenes from marine opisthobranch molluscs.
522 Current Organic Chemistry 4, 1201–1248.
523 Goldstein, J.L., Brown, M.S., 1990. Regulation of the mevalonate pathway. Nature 343,
524 425–430.
525 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis,
526 X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N.,
527 Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K.,
528 Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq
529 data without a reference genome. Nat. biotech. 29, 644–652.
530 Hamberger, B., Bak, S., 2013. Plant P450s as versatile drivers for evolution of species-
531 specific chemical diversity. Phil. Trans. R. Soc. B 368, 20120426.
532 Harrison, M.C., Jongepier, E., Robertson, H.M., Arning, N., Bitard-Feildel, T., Chao, H.,
533 Childers, C.P., Dinh, H., Doddapaneni, H., Dugan, S., Gowin, J., Greiner, C., Han,
534 Y., Hu, H., Hughes, D.S.T., Huylmans, A.-K., Kemena, C., Kremer, L.P.M., Lee,
535 S.L., Lopez-Ezquerra, A., Mallet, L., Monroy-Kuhn, J.M., Moser, A., Murali, S.C.,
536 Muzny, D.M., Otani, S., Piulachs, M.-D., Poelchau, M., Qu, J., Schaub, F., Wada-
537 Katsumata, A., Worley, K.C., Xie, Q., Ylla, G., Poulsen, M., Gibbs, R.A., Schal, C.,
538 Richards, S., Belles, X., Korb, J., Bornberg-Bauer, E., 2018. Hemimetabolous
539 genomes reveal molecular basis of termite eusociality. Nat. Ecol. Evol. 2, 557–566.

Hayashi, Y., Shigenobu, S., Watanabe, D., Toga, K., Saiki, R., Shimada, K., Bourguignon, T., Lo, N., Hojo, M., Maekawa, K., Miura, T., 2013. Construction and characterization of normalized cDNA libraries by 454 pyrosequencing and estimation of DNA methylation levels in three distantly related termite species. PLoS One 8, e76678.

Hayashi, Y., Maekawa, K., Nalepa, C.A., Miura, T., Shigenobu, S., 2017. Transcriptome sequencing and estimation of DNA methylation level in the subsocial wood‐feeding cockroach *Cryptocercus punctulatus* (Blattodea: Cryptocercidae). Appl. Entomol. Zool. 52, 643–651.

Hojo, M., Koshikawa, S., Matsumoto, T., Miura, T., 2004. Developmental pathways and plasticity of neuter castes in *Nasutitermes takasagoensis* (Isoptera: Termitidae). Sociobiol. 44, 433–441.

Hojo, M., Morioka, M., Matsumoto, T., Miura, T., 2005. Identification of soldier caste-specific protein in the frontal gland of nasute termite *Nasutitermes takasagoensis* (Isoptera: Termitidae). Insect Biochem. Mol. Biol. 35, 347–354.

Hojo, M., Matsumoto, T., Miura, T., 2007. Cloning and expression of a geranylgeranyl diphosphate synthase gene: insights into the synthesis of termite defence secretion. Insect Mol. Biol. 16, 121–131.

Hojo, M., Maekawa, K., Saitoh, S., Shigenobu, S., Miura, T., Hayashi, Y., Tokuda, G., Maekawa, H. 2012. Exploration and characterization of genes involved in the synthesis of diterpene defence secretion in nasute termite soldiers. Insect Mol. Biol. 21, 545–557.

Jirošová, A., Jančařík, A., Menezes, R.C., Bazalová, O., Dolejšová, K., Vogel, H., Jedlička, P., Buček, A., Brabcová, J., Majer, P., Hanus, R., Svatoš, A. 2017. Co-

option of the sphingolipid metabolism for the production of nitroalkene defensive chemicals in termite soldiers. Insect Biochem. Mol. Biol. 82, 52–61.

Jirošová, A., Jančařík, A., Menezes, R.C., Bazalová, O., Dolejšová, K., Vogel, H., Jedlička, P., Buček, A., Brabcová, J., Majer, P., Hanus, R., Svatoš, A. 2018. Metabolomic and transcriptomic data on major metabolic/biosynthetic pathways in workers and soldiers of the termite Prorhinotermes simplex (Isoptera: Rhinotermitidae) and chemical synthesis of intermediates of defensive (E)-nitropentadec-1-ene biosynthesis. Data Brief. 18, 1614–1627.

Kainou, T., Kawamura, K., Tanaka, K., Matsuda, H., Kawamukai, M., 1999. Identification of the *GGPS1* genes encoding geranylgeranyl diphosphate synthases from mouse and human. Biochim. Biophys. Acta 1437, 333–340.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858.

Krishna, K., Grimaldi, D.A., Krishna, V., Engel, M.S., 2013. Treatise on the isoptera of the world. Bulletin of the American Museum of Natural History B377, 1–200.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Kuzuguchi, T., Morita, Y., Sagami, I., Sagami, H., Ogura, K., 1999. Human geranylgeranyl diphosphate synthase: cDNA cloning and expression. J. Biol. Chem. 274, 5888–5894.

Kuzuyama, T., 2002. Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Biosci Biotechnol Biochem* **66**: 1619–1627.

586 Lai, C., McMahon, R., Young, C., Mackay, T.F.C., Langley, C.H., 1998. *quemao*, a
587 Drosophila bristle locus, encodes geranylgeranyl pyrophosphate synthase. Genetics
588 149, 1051–1061.
589 Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2 Nat.
590 Methods 9, 357–359.
591 Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Mol.
592 Biol. Evol. 25, 1307–1320.
593 Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets
594 of protein or nucleotide sequences. Bioinformatics 22, 1658–1659.
595 Liang, P.-H., Ko, T.-P., Wang, A. H.-J., 2002 Structure, mechanism and function of
596 prenyltransferases. Eur. J. Biochem. 269, 3339–3354.
597 Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing
598 reads. EMBnet journal 17, 10–12.
599 Nelson, D, Werck-Reichhart, D., 2011. A P450-centric view of plant evolution. Plant J.
600 66, 194–211.
601 Noirot, C., 1989. Social structure in termite societies. Ethol. Ecol. Evol. 1, 1–17.
602 Noirot, C., Darlington, J.P.E., 2000. Termite nests: architecture, regulation and defence.
603 In: Abe, T., Bignell, D.E., Higashi, M. (Eds), Termites: Evolution, Sociality,
604 Symbioses, Ecology. Kluwer Academic Publishers, Dordrecht, pp. 121–139.
605 Noirot, C., Quennedey, A., 1974. Fine structure of insect epidermal glands. Annu. Rev.
606 Entomol. 19, 61–80.
607 Ogura, K., Koyama, T., 1998. Enzymatic aspects of isoprenoid chain elongation. Chem.
608 Rev. 98, 1263–1276.

609 Ohnuma, S., Narita, K., Nakazawa, T., Ishida, C., Takeuchi, Y., Ohto, C., Nishino, T.,
610 1996 A role of the amino acid residue located on the fifth position before the first
611 aspartate-rich motif of farnesyl diphosphate synthase on determination of the final
612 product. J. Biol. Chem. 48, 30748–30754.
613 Pfaffl, M.W., Tichopad, A., Prgomet, C., Neuvians, T.P., 2004. Determination of stable
614 housekeeping genes, differentially regulated target genes and sample integrity:
615 BestKeeper–excel-based tool using pair-wise correlations. Biotechnol. Lett. 26,
616 509–515.
617 Poulsen, M., Hu, H., Li, C., Chen,Z., Xu, L., Otani, S., Nygaard, S., Nobre, T., Klaubauf,
618 S., Schindler, P.M., Hauser, F., Pan, H., Yang, Z., Sonnenberg, A.S.M., de Beer,
619 Z.W., Zhang, Y., Wingfield, M.J., Grimmelikhuijzen, C.J.P, de Vries, R.P., Korb, J.,
620 Aanen, D.K., Wang, J., Boomsma, J.J., Zhang, G., 2014. Complementary symbiont
621 contributions to plant decomposition in a fungus-farming termite. Proc. Natl. Acad.
622 Sci. USA 111, 14500–14505.
623 Prestwich, G.D., 1979a. Chemical defense by termite soldiers. J. Chem. Ecol. 5, 459–480.
624 Prestwich, G.D., 1979b. Interspecific variation in the defence secretions of *Nasutitermes*
625 soldiers. Biochem. System. Ecol. 7, 211–221.
626 Prestwich, G.D., 1983a. The chemical defenses of termites. Scientific American 249, 68–
627 75.
628 Prestwich, G.D., 1983b. Chemical systematics of termite exocrine secretions. Annu. Rev.
629 Ecol. Syst. 14, 287–311.
630 Prestwich, G.D., 1984. Defense mechanisms of termites. Annu. Rev. Entomol. 29, 201–
631 232.

632 Prestwich, G.D., 1988. The chemicals of termite societies (Isoptera). Sociobiology 14,
633 175–191.
634 Prestwich, G.D., Jones, R.W., Collins, M.S., 1981. Terpene biosynthesis by nasute
635 termite soldiers (Isoptera: Nasutitermitinae). Insect Biochem. 11, 331–336.
636 Quennedey, A., 1984. Morphology and ultrastructure of termite defense glands. In:
637 Hermann, H. (Ed.), Defensive mechanisms in social insects. Praeger Publishers,
638 New York, pp. 151–200.
639 Rastogi, S., Liberles, D.A., 2005. Subfunctionalization of duplicated genes as a transition
640 state to neofunctionalization. BMC Evol. Biol. 5, 28.
641 Roberts, A., Pachter, L., 2013. Streaming fragment assignment for real-time analysis of
642 sequencing experiments. Nat. Methods 10, 71–73.
643 Robinson, M.D., Oshlack, A. 2010. A scaling normalization method for differential
644 expression analysis of RNA-seq data. Genome Biol. 11, R25.
645 Robinson, M.D., McCarthy, D.J., Smyth, G.K. 2010. edgeR: a Bioconductor package for
646 differential expression analysis of digital gene expression data. Bioinformatics 26,
647 139–140.
648 Roisin, Y., 2000. Diversity and evolution of caste patterns. In: Abe, T., Bignell, D.E.,
649 Higashi, M. (Eds.), Termites: Evolution, Sociality, Symbioses, Ecology. Kluwer
650 Academic Publishers, Dordrecht, pp. 95–119.
651 Roux, E. A., Korb, J., 2004. Evolution of eusociality and the soldier caste in termites: a
652 validation of the intrinsic benefit hypothesis. J. Evol. Biol. 17, 869–875.
653 Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing
654 phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

655 Sillam-Dussès, D., Sémon, E., Robert, A., Cancello, E., Lenz, M., Valterovà, I.,
656 Bordereau, C., 2010. Identification of multi-component trail pheromones in the most
657 evolutionarily derived termites, the Nasutitermitinae (Termitidae). Biol. J. Linn. Soc.
658 99, 20–27.
659 Singh, A.K., Prestwich, G.D., 1986. Synthesis of tritiumlabelled cembrene-A. J. Labelled
660 Compd. Radiopharm. 23, 377–382.
661 Šobotník, J., Jirošová, A., Hanus, R., 2010. Chemical warfare in termites. J. Insect Physiol.
662 56, 1012–1021.
663 Tarshis, L.C., Proteau, P.J., Kellogg, B.A., Sacchettini, J.C., Poulter, C.D., 1996.
664 Regulation of product chain length by isoprenyl diphosphate synthases. Proc. Natl.
665 Acad. Sci. USA 93, 15018–15023.
666 Terrapon, N., Li, C., Robertson, H.M., Ji, L., Meng, X., Booth,W., Chen, Z., Childers,
667 C.P., Glastad, K.M., Gokhale, K., Gowin, J., Gronenberg, W., Hermansen, R.A., Hu,
668 H., Hunt, B.G., Huylmans, A.K., Khalil, S.M., Mitchell, R.D., Munoz-Torres, M.C.,
669 Mustard, J.A., Pan, H., Reese, J.T., Scharf, M.E., Sun, F., Vogel, H., Xiao, J., Yang,
670 W., Yang, Z., Yang, Z., Zhou, J., Zhu, J., Brent, C.S., Elsik, C.G., Goodisman, M.A.,
671 Liberles, D.A., Roe, R.M., Vargo, E.L., Vilcinskas, A., Wang, J., Bornberg-Bauer,
672 E., Korb, J., Zhang, G., Liebig, J., 2014. Molecular traces of alternative social
673 organization in a termite genome. Nat. Commun. 5, 3636.
674 Tian, L., Zhou, X., 2014. The soldiers in societies: defense, regulation, and evolution. Int.
675 J. Biol. Sci. 10, 296–308.
676 Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A.,
677 Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data

678 by geometric averaging of multiple internal control genes. Genome Biol. 3,
679 RESEARCH0034.
680 Wang, K.C., Ohnuma, S., 2000. Isoprenyl diphosphate synthases. Biochim. Biophys.
681 Acta 1529, 33–48.
682 Wang, C., Chen, Q., Fan, D., Li, J., Wang, G., Zhang, P., 2016. Structural analyses of
683 short-chain prenyltransferases identify an evolutionarily conserved GFPPS Clade in
684 brassicaceae plants. Mol. Plant 9, 195–204.
685 Zerbino, D.R., Birney, E., 2008. Velvet: Algorithms for de novo short read assembly
686 using de Bruijn graphs. Genome Res. 18, 821–829.
687 Zhang, F.L., Casey, P.J., 1996. Protein prenylation: molecular mechanisms and functional
688 consequences. Annu. Rev. Biochem. 65, 241–269.
689 Zhao, Y., Yu, L., Gao, J., Fu, Q., Hua, Y., Zhang, H., Zhao, S., 2000. cDNA cloning,
690 chromosome mapping and expression characterization of human geranylgeranyl
691 pyrophosphate synthase. Science in China Series C Life Sci. 43, 613–622.

692 **Figure Captions**

693 Fig. 1. Molecular phylogenetic tree based on the GGPP synthase amino acid sequences
694 in *N. takasagoensis* (NtGGPPS1–7) and other organisms. Inferred relationships were
695 supported using the Maximum Likelihood method, based on the Le and Gascuel (2008)
696 model. Bootstrap probability (1000 replicates, only >50%) is shown on each branch.
697 Scale bar indicates the substitution rate per site.

698

699 Fig. 2. Alignment of NtGGPPS amino acid sequences produced by MEGA7. Two
700 conserved aspartate-rich domains for isoprenyl diphosphate synthase are underlined.
701 Asterisk indicates the amino acid residue at the fifth position upstream of the first
702 aspartate-rich consensus region. Filled and shaded boxes indicate the conserved and the
703 same characteristic residues in all sequences, respectively.

704

705 Fig. 3. Relative expression of genes responsible for the GGPP synthesis pathway
706 quantified in heads and bodies of soldiers, major workers, male alates and female alates
707 of *N. takasagoensis*. The expression levels were normalised to those expressed in the
708 heads of the soldier caste (indexed as 1.0). Data are expressed as mean ± SD (biological
709 replication numbers = 4). Differences in letters above the index bars denote significant
710 differences (based on one-way ANOVA, followed by multiple comparison tests with
711 Bonferroni correction, Padj < 0.05).

712

713

714 **Supporting Information**

715 Figure S1.

716 Putative biosynthetic pathway of cembrane-derived polycyclic diterpene in nasute termite.

717

718 Figure S2.

719 Diagram of separation of soldier head used for RNA extraction for RNA-seq library

720 construction.

721

722 Figure S3.

723 Nucleotide sequence aliment of *GGPPS* genes in *N. takasagoensis*. Forward and reverse

724 primers are underlined and double-underlined, respectively. Capital letters and small

725 capital letters indicate the sequences of coding regions and untranslated regions,

726 respectively. Asterisks indicate the sequences of start and stop codon.

727

728 Figure S4.

729 Comparison of the 3D protein structures of GGPPSs in *N. takasagoensis* (NtGGPPSs)

730 and *Z. nevadensis* (ZnGGPPS). Arrows in NtGGPPS4 indicate some conformational

731 variations compared to the other termite GGPPSs.

732

733 Table S1.

734 Primer sequences and optimal primer concentration for qRT-PCR.

735

736 Table S2.

737 Number of raw reads in each library.

738

739 Table S3.

740 Counts per million (cpm) value normalized by TMM.

741

742 Table S4.

743 Results of top hit protein by BLAST against ncbi nr protein database.

744

745 Table S5.

746 The abundant ORF of the soldier hear containing frontal gland class 1 cell layer of

747 *Nasutitermes takasagoensis* (FDR<0.05).

748

749 Table S6.

750 Result of BLASTp using nine genes involved with diterpene synthesis pathway of

751 *Acyrthosiphon pisum* as query.

752

753 Table S7.

754 The gene expression stability values generated by three software programmes.