Insect Biochemistry and Molecular Biology 111 (2019) 103177

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

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] NIBB 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

ARTICLE INFO

Keywords:
Chemical defence
Exocrine gland
Mevalonate pathway
Nasus
Termitidae

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.

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

* Corresponding author.
E-mail address: masaruhojo0911@gmail.com (M. Hojo).

https://doi.org/10.1016/j.ibmb.2019.103177
Received 8 March 2019; Received in revised form 18 May 2019; Accepted 16 June 2019
Available online 20 June 2019
0965-1748/

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.

2.2. RNA extraction and cDNA library construction

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

2.3. RNA sequencing

Following the procedure described by Hayashi et al. (2017), we used TruSeq RNA sample preparation kits (Version 2) (Illumina, San Diego, CA, USA) to generate a complementary DNA (cDNA) library from 1 μg of total RNA from each caste or tissue. The cDNA libraries were individually indexed for multiplex sequencing before being pooled and sequenced on a HiSeq 2000 platform (Illumina) with 101 bp paired-end reading.

2.4. De novo transcriptome assembly and expression analysis

We trimmed the low-quality ends (< QV30) and adapter sequences using the cutadapt programme (Martin, 2011), whereas the reads shorter than 50 bp were discarded. The filtered reads obtained from the five cDNA libraries derived from the same termite colony were pooled and subjected to *de novo* transcriptome assembly using the Trinity software (Grabherr et al., 2011). Open reading frames (ORFs) were predicted using the TransDecoder programme (Trinity package). The predicted ORF sequences were grouped using the cd-hit-est programme (Li and Godzik, 2006) with a minimum identity of 97%, from which we obtained a non-redundant (nr) coding sequence set. A sequence similarity search was performed against the NCBI's nr protein database (ver. June 2018) with BLASTp using predicted proteins as queries.

When we evaluated the Trinity assembly, we noticed that the contigs derived from different *GGPPS* gene paralogs were erroneously assembled together; this was probably because the *GGPPS* gene paralogs are so similar to each other that Trinity assembling with a k-mer size = 25 collapsed paralogs sharing sequences longer than 25 base pairs. To overcome this issue, we used the velvet programme (ver. 1.2.07) (Zerbino and Birney, 2008) for assembling *GGPPS* gene paralogs because velvet programme allows a flexible setting of the parameter of k-mer size as longer values. We collected the Illumina reads derived from *GGPPS* gene paralogs and attempted assembling using the velvet programme with various k-mer sizes ranging from 75 to 99. We found that the *GGPPS* ORF assembly produced by a parameter of k-mer size = 95 was the best, and we used it for the downstream analyses.

For differential expression analysis, short reads were mapped to the reference ORF sequences using Bowtie2 software (Langmead and Salzberg, 2012). Transcript abundances were estimated using eXpress software (Roberts and Pachter, 2013). To adjust the library sizes and skewed expressions of transcripts, we normalised the estimated abundance values using the trimmed mean of M values (TMM) normalisation method (Robinson and Oshlack, 2010). Differential expression analysis was conducted using the edgeR statistical package (Robinson et al., 2010).

2.5. Identification of genes involved in diterpene biosynthesis and cDNA cloning

We used the tBLASTx, with the nucleotide sequences of *Acyrthosiphon pisum* as queries, to search the reference ORF sequences of *N. takasagoensis* for the genes involved in diterpene biosynthesis. For confirming the accurate assembly, we determined cDNA sequences, including the untranslated region (UTR) of the genes of interest, as follows. Total RNA was extracted from soldier heads as described above. First-stranded cDNA was synthesised using SuperScript III First-Strand Synthesis SuperMix (Thermo Fisher Scientific). Each cDNA sequence was obtained using RT-PCR and rapid amplification of cDNA ends (RACE) with gene-specific primers (Table S1). PCR products were purified and nucleotide sequences were determined with a Big Dye Terminator v3.1 Cycle Sequencing Kit using a 3130xl Genetic Analyzer (Thermo Fisher Scientific).

2.6. Amino acid sequences alignment, phylogenetic analysis and protein structure prediction

Using amino acid sequences of *N. takasagoensis* as queries, we used tBLASTx to search for homologous genes of the GGPP synthesis pathway genes in closely related insect species, such as *Blattella germanica* Linnaeus, *Zootermopsis nevadensis* Hagen, *Cryptotermes secundus* Hill, and *Macrotermes natalensis* Haviland, whose genome sequences were published (Terrapon et al., 2014; Poulsen et al., 2014; Harrison et al., 2018). We used MEGA7 software (Kumar et al., 2016) to perform sequence alignments and construct a phylogenetic tree of the *GGPPS* genes obtained from termites and other insects and organisms. Sequences were aligned using the MUSCLE algorithm. Phylogenetic relationships were inferred using the Maximum Likelihood method based on an LG model (Le and Gascuel, 2008). We obtained initial phylogenetic trees for the heuristic search by applying the Neighbour-Joining method (Saitou and Nei, 1987). To infer the structural differences among GGPPSs, the protein structures were predicted based on the deduced amino acid sequences using Phyre2 web site (Kelley et al., 2015) with *Z. nevadensis* GGPPS amino acid sequences (XP_021933767) as a reference.

2.7. Relative expression analysis by qRT-PCR

RNA samples were separately extracted from head and body tissues of the four terminal castes (i.e. male soldier, female major worker, male alate and female alate). Each biological sample was replicated from four different colonies. We determined the quality and quantity of the extracted RNA through spectroscopic measurements at 230, 260 and 280 nm using the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). The total RNA was transcribed using the PrimeScript RT reagent Kit with gDNA Eraser (Takara Bio, Shiga, Japan), according to the manufacturer's instructions. The qRT-PCR 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 (Table 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,

Table 1
Summary of the sequencing and assembly.

	No. of component	Total base	Mean length (bp)
Sequencing for six libraries			
Raw reads	12,23,09,746	12,35,32,84,346	101
High quality reads	12,23,09,746	11,02,15,45,062	90.1
Assembly without another colony			
Reads used in assembly	10,29,69,192	9,27,46,00,844	90.1
Number of contigs	1,18,787	16,56,68,514	1395
N50			2864
Remove redundancy			
Number of contigs	92,028	9,25,24,553	1005
N50			1983
ORF prediction and remove redundancy			
Number of contigs	40,863	3,09,53,460	758
N50			1668

Table 2
The 50 most abundant ORF of the soldier head containing frontal gland class 1 cell layer of Nasutitermes takasagoensis.

TSA accession no.	Log FC	Log CPM	P-value	FDR	BLAST nr top hit
IAEB01019068	−8.133452139	6.05744782	5.10E-24	2.08E-19	reverse transcriptase [Lasius niger]
IAEB01019069	−9.001441264	5.651927734	1.54E-23	3.15E-19	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like [Linepithema humile]
IAEB01012008	−10.50050201	3.930967442	2.97E-22	4.05E-18	No significant homology
IAEB01031403	−8.030272479	3.586219116	1.63E-20	1.33E-16	No significant homology
IAEB01039737	−7.447042948	10.73861483	1.34E-20	1.33E-16	cytochrome P450 6j1-like [Zootermopsis nevadensis]
IAEB01031404	−6.485176261	4.962612213	3.02E-19	2.06E-15	No significant homology
IAEB01019592	−8.242869986	13.88454172	4.70E-19	2.75E-15	cytochrome P450 6j1-like [Zootermopsis nevadensis]
IAEB01019022	−8.286817796	11.63668811	8.79E-19	4.49E-15	soldier-caste specific protein [Coptotermes formosanus]
IAEB01019071	−7.62147701	3.621106324	1.27E-18	5.76E-15	No significant homology
IAEB01005576	−5.796239678	7.093763326	4.86E-17	1.99E-13	arylsulfatase B-like isoform X1 [Zootermopsis nevadensis]
IAEB01035994	−7.181479713	3.243637079	1.09E-16	4.06E-13	No significant homology
IAEB01018454	−5.823074409	5.821271446	1.59E-16	5.40E-13	putative RNA-directed DNA polymerase from transposon X-element, partial [Stegodyphus mimosarum]
IAEB01011597	−5.429993116	11.03807601	2.62E-16	7.66E-13	ABC transporter G family member 20 [Zootermopsis nevadensis]
IAEB01019020	−9.046535761	10.21553261	2.60E-16	7.66E-13	soldier-caste specific protein [Coptotermes formosanus]
IAEB01030057	−9.030191716	2.449737497	4.86E-16	1.32E-12	soldier-caste specific protein [Coptotermes formosanus]
IAEB01019593	−7.794204008	8.262705779	5.88E-16	1.50E-12	hypothetical protein L798_09889 [Zootermopsis nevadensis]
IAEB01037587	−8.536542279	8.711372922	1.37E-15	3.12E-12	hypothetical protein [Coptotermes formosanus]
IAEB01037603	−9.288596071	2.696022813	1.30E-15	3.12E-12	No significant homology
IAEB01012495	−6.507137242	7.523137034	1.81E-15	3.89E-12	hypothetical protein L798_09317 [Zootermopsis nevadensis]
IAEB01019021	−8.927227896	10.43910459	4.37E-15	8.93E-12	soldier-caste specific protein [Coptotermes formosanus]
IAEB01006569	−6.378697763	8.147069418	4.99E-15	9.71E-12	farnesyl pyrophosphate synthase [Zootermopsis nevadensis]
IAEB01002528	−6.163938655	5.261737637	5.33E-15	9.90E-12	No significant homology
IAEB01013757	−5.113705923	8.947850898	6.04E-15	1.07E-11	uncharacterized protein LOC110831736 isoform X4 [Zootermopsis nevadensis]
IAEB01019070	−5.161811143	3.824520797	1.09E-14	1.85E-11	No significant homology
IAEB01003974	−8.071734319	7.898731857	2.51E-14	4.10E-11	No significant homology
IAEB01019137	−6.17247084	10.5845805	3.83E-14	6.01E-11	juvenile hormone acid O-methyltransferase [Zootermopsis nevadensis]
IAEB01018457	−4.990097341	3.986628006	5.42E-14	8.21E-11	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like [Anoplophora glabripennis]
IAEB01018459	−5.800935929	2.963762287	5.93E-14	8.65E-11	No significant homology
IAEB01019956	−8.934883612	11.984693	6.47E-14	9.11E-11	hypothetical protein L798_10332 [Zootermopsis nevadensis]
IAEB01018455	−5.054023301	5.228460194	8.57E-14	1.17E-10	RNA-directed DNA polymerase from mobile element jockey, partial [Stegodyphus mimosarum]
IAEB01005579	−6.681373329	4.823487868	1.00E-13	1.32E-10	arylsulfatase B-like isoform X1 [Zootermopsis nevadensis]
IAEB01037264	−5.983362886	2.42221365	1.20E-13	1.53E-10	No significant homology
IAEB01007617	−6.278143698	4.062032303	1.33E-13	1.65E-10	coiled-coil domain-containing protein AGAP005037 [Zootermopsis nevadensis]
IAEB01018456	−4.774811005	4.988063538	1.51E-13	1.77E-10	putative endonuclease and reverse transcriptase-like protein [Operophtera brumata]
IAEB01029322	−7.089609488	1.976437394	1.51E-13	1.77E-10	No significant homology
IAEB01000071	−8.635385992	2.075787935	1.66E-13	1.84E-10	No significant homology
IAEB01003973	−8.538590332	10.41808026	1.66E-13	1.84E-10	Alpha-tocopherol transfer protein-like [Zootermopsis nevadensis]
IAEB01013823	−4.932846459	7.230948601	2.28E-13	2.46E-10	ethylmalonyl-CoA decarboxylase-like isoform X1 [Zootermopsis nevadensis]
IAEB01003975	−7.479742213	7.803457348	2.81E-13	2.95E-10	No significant homology
IAEB01029812	−6.000067384	2.455398265	3.62E-13	3.70E-10	E3 ubiquitin-protein ligase siah-1 [Coptotermes formosanus]
IAEB01035995	−4.876041778	4.986251134	4.12E-13	4.11E-10	polyprotein [*Blattella germanica*]
IAEB01019139	−5.908840483	5.970937636	1.39E-12	1.35E-09	No significant homology
IAEB01005686	−4.887448263	9.146703839	1.82E-12	1.73E-09	PREDICTED: titin isoform X7 [*Tribolium castaneum*]
IAEB01019401	−5.554938253	2.885939974	2.56E-12	2.38E-09	RNA-directed DNA polymerase from mobile element jockey, partial [Trachymyrmex septentrionalis]
IAEB01003312	−6.952105459	1.870559575	2.75E-12	2.40E-09	No significant homology
IAEB01015847	−4.546589688	7.018562457	2.82E-12	2.40E-09	No significant homology
IAEB01018458	−4.443019166	3.773802655	2.82E-12	2.40E-09	No significant homology
IAEB01035996	−4.44954187	3.743290123	2.68E-12	2.40E-09	non-LTR retrotransposon CATS, partial [*Bombyx mori*]
IAEB01034499	−4.535365348	10.46989967	3.92E-12	3.27E-09	scavenger receptor class B member 1 isoform X1 [Zootermopsis nevadensis]
IAEB01019018	−7.403693811	12.39906782	4.83E-12	3.94E-09	soldier-caste specific protein [Coptotermes formosanus]

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) (Table 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

Table 3
Determined sequences of nine genes involved with diterpene synthesis pathway and their corresponding ORF IDs by RNA-seq.

Enzyme name	Gene name	Accession no.	CDS length	TSA accession no.	Registration
acetyl-Coenzyme A acetyltransferase, cytosolic	*NtACAT2*	LC440320	1185	IAEB01004244	present study
3-hydroxy-3-methylglutaryl-Coenzyme A synthase	*NtHMGS1*	LC440321	1389	IAEB01021509	present study
	*NtHMGS2**	AB733008	1419	IAEB01013267	Hojo et al. (2012)
3-hydroxy-3-methylglutaryl-Coenzyme A reductase	*NtHMGR**	AB733009	2610	IAEB01037270	Hojo et al. (2012)
mevalonate kinase	*NtMevK*	LC440322	1305	IAEB01035397	present study
phosphomevalonate kinase	*NtMevPK*	LC440323	621	IAEB01011905	present study
diphosphomevalonate decarboxylase	*NtMevPPD*	LC440324	1176	IAEB01039000	present study
isopentenyl diphosphate isomerase	*NtIPPI*	LC440325	702	IAEB01039163	present study
farnesyl diphosphate synthase	*NtFPPS1*	LC440326	1251	IAEB01016062	present study
	NtFPPS2	LC440327	1299	IAEB01006569	present study
	NtFPPS3	LC440328	1299	IAEB01006570	present study
geranylgeranyl diphosphate synthase	*NtGGPPS1-B*	AB266075	951	IAEB01040867	Hojo et al. (2007)
	NtGGPPS2-B	AB266079	951	IAEB01040866	Hojo et al. (2007)
	NtGGPPS3-B	AB266081	951	IAEB01015961, IAEB01040864	Hojo et al. (2007)
	NtGGPPS4	LC440329, AB266082	951	IAEB01040865, IAEB01040868, IAEB01040869, IAEB01040871	present study
	NtGGPPS5	LC440330	945	IAEB01012707, IAEB01012708	present study
	NtGGPPS6	LC440331	954	IAEB01039951	present study
	NtGGPPS7	LC440332	954	IAEB01006415, IAEB01006417, IAEB01006409	present study

*Gene names were renamed in this study.

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 (Fig. S1) were identical in all species examined (Table 4). However, the number of genes in the following GGPP synthesis pathway differed among species: there were one or two *FPPS* genes in all species except for *N. takasagoensis*, which had three. In particular, the number of *GGPPS* genes was larger in *N. takasagoensis* (at least eight) compared with the cockroach and other termite species examined (only one).

3.5. Phylogenetic relationship of the GGPPS genes in termites

Phylogenetic analysis indicated the termite *GGPPS* genes are divided into two clades (Fig. 1); one clade containing homologous genes identified from all species examined (ancestral clade) and the other clade containing genes identified only from termitid species (expanded clade). In the latter clade, genes could be further divided into two clades; one included the *GGPPS* genes of *N. takasagoensis* (*NtGGPPS6*) and *M. natalensis*, and the other included all the other genes only obtained from *N. takasagoensis* (*NtGGPPS1 to NtGGPPS5*).

3.6. Structures of GGPPS proteins of N. takasagoensis

The amino acid sequence alignment of seven ORF determined by cloning indicated that two aspartate-rich regions, involved in the binding of the diphosphate moieties of the allylic substrate through Mg^{2+} bridges (Liang et al., 2002), were conserved in all NtGGPPS proteins with some substitutions (Fig. 2). The amino acid residues at the fifth position upstream of the first aspartate-rich region (FARM) in all NtGGPPS proteins were neutral residues with relatively smaller molecular masses, such as serine, alanine and 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.

Table 4
Number of genes or contigs involved in GGPP synthetic pathway.

Symbol	*B. germanica*	*Z. nevadensis*	*C. secundus*	*M. natalensis*	*N. takasagoensis*
ACAT2	1	1	1	1	1
HMGS	2	2	2	2	2
HMGR	1	1	1	1	1
MevK	1	1	1	1	1
MevPK	1	1	1	1	1
MevPPD	1	1	1	1	1
IPPI	1	1	1	1	1
FPPS	2	1	2	1	3
GGPPS	1	1	1	2	>8

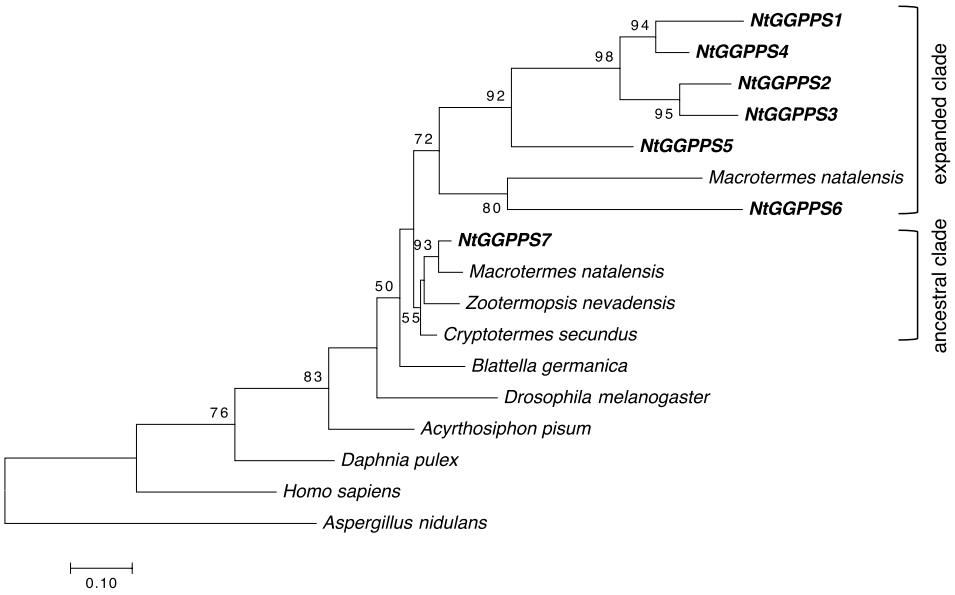

Fig. 1. Molecular phylogenetic tree based on the GGPP synthase amino acid sequences in *N. takasagoensis* (NtGGPPS1–7) and other organisms. Inferred relationships were supported using the Maximum Likelihood method, based on the Le and Gascuel (2008) model. Bootstrap probability (1000 replicates, only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

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

4.1. Predicted ORFs obtained from male caste head tissues

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

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

4.2. Gene duplication and functionalisation of GGPPS

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

4.3. Evolution of the termite GGPPS genes

Whole genome sequences in some termite and cockroach species revealed that the gene duplication of *GGPPS* has occurred only in

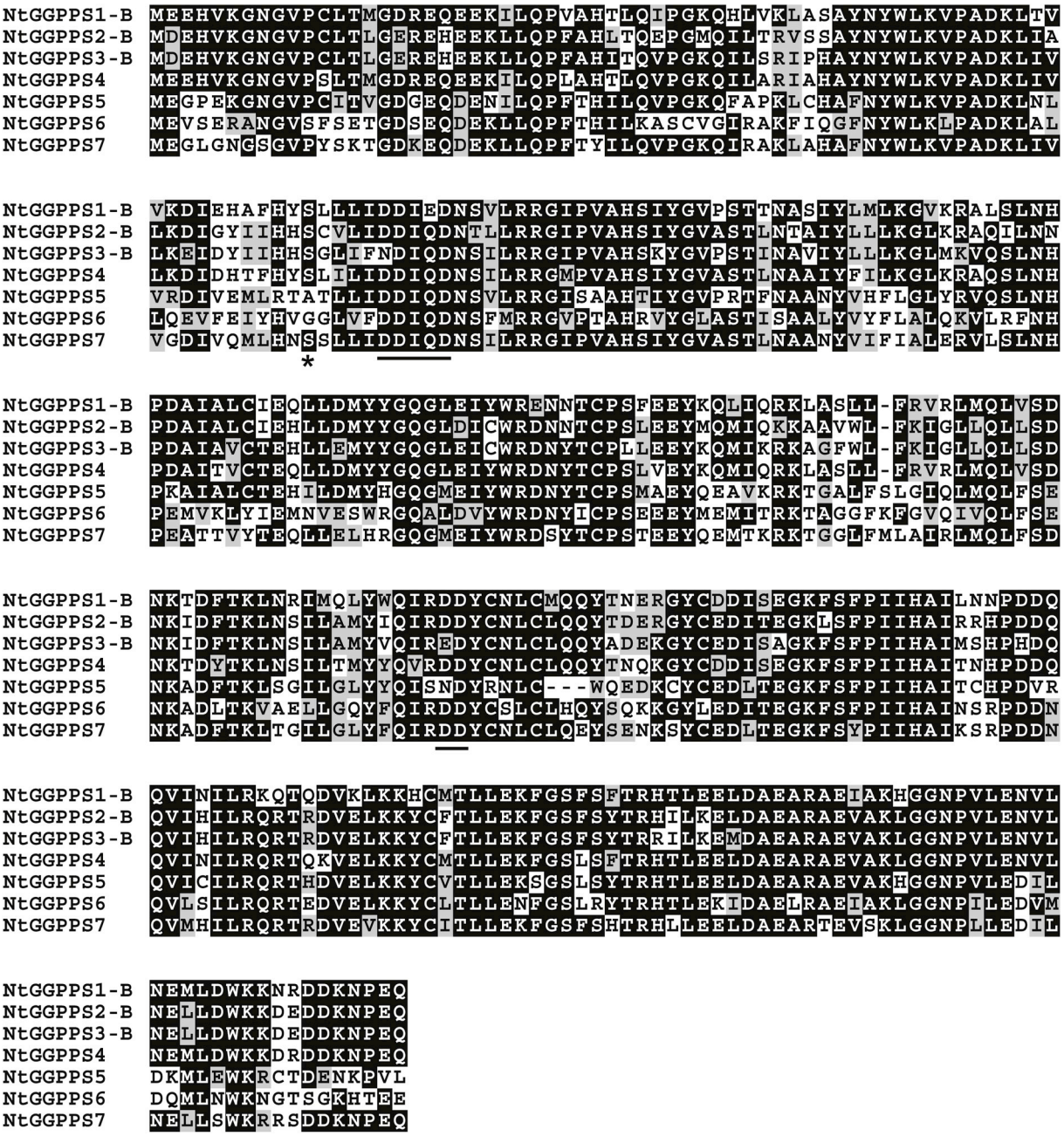

Fig. 2. Alignment of NtGGPPS amino acid sequences produced by MEGA7. Two conserved aspartate-rich domains for isoprenyl diphosphate synthase are underlined. Asterisk indicates the amino acid residue at the fifth position upstream of the first aspartate-rich consensus region. Filled and shaded boxes indicate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

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

The fact that only *NtGGPPS7* was expressed at any tissues in any castes (Fig. 3) suggests that *NtGGPPS7* likely has essential house-

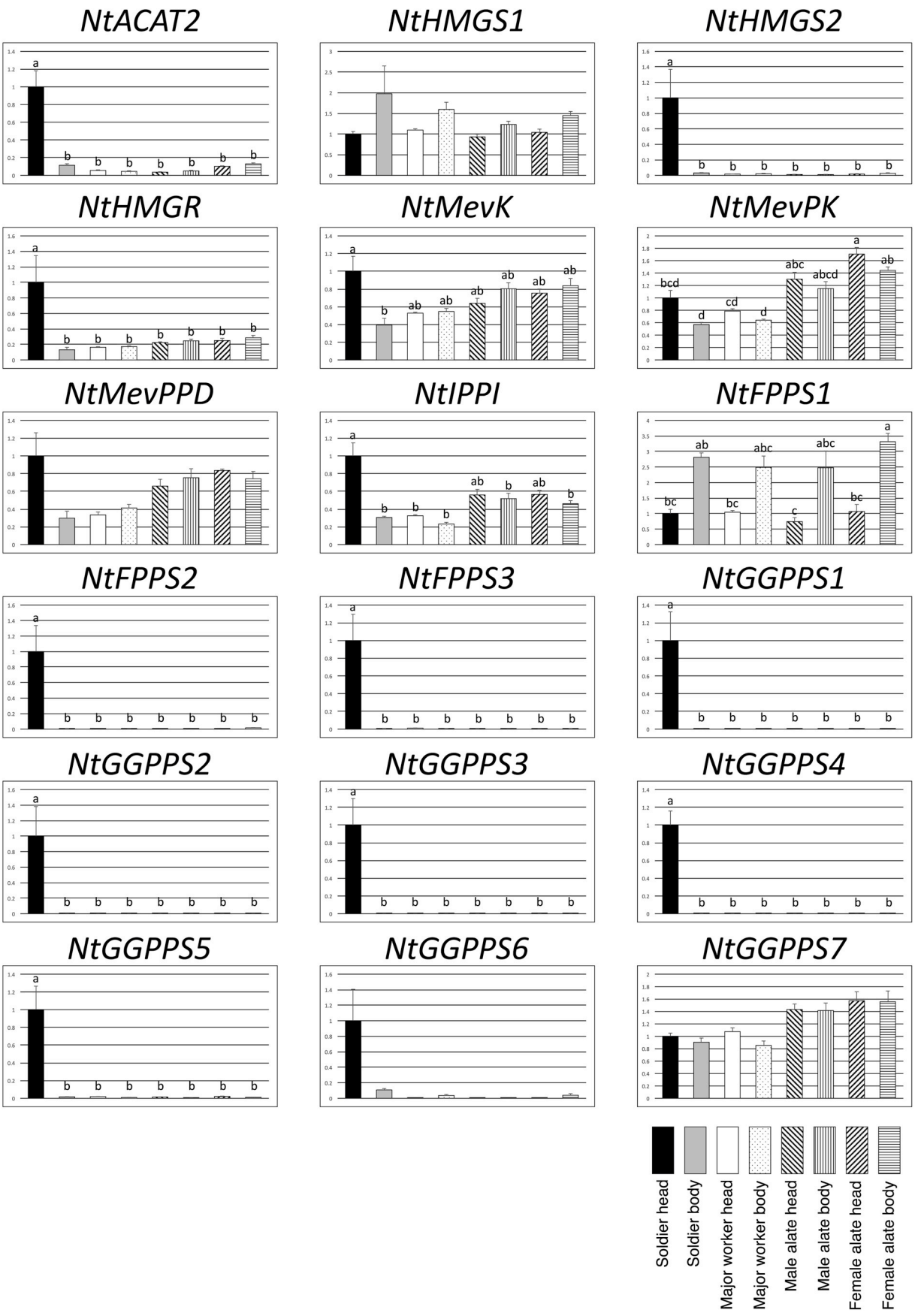

Fig. 3. Relative expression of genes responsible for the GGPP synthesis pathway quantified in heads and bodies of soldiers, major workers, male alates and female alates of *N. takasagoensis*. The expression levels were normalised to those expressed in the heads of the soldier caste (indexed as 1.0).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

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

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

4.5. Conclusions

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

Acknowledgements

We thank K. Yamaguchi and Y. Hayashi for technical support in Illumina sequencing. This work was financially supported by JSPS KAKENHI (Nos.15K07798 and 17H01510) provided by the Ministry of Education, Culture, Sports, Science and Technology, Japan. This work was also supported by NIBB Collaborative Research Program (Nos. 18-332 and 19-337) to M. H.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ibmb.2019.103177.

References

Andersen, C.L., Jensen, J.L., Ørntoft, T.F., 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 64, 5245–5250.

Aratake, S., Tomura, T., Saitoh, S., Yokokura, R., Kawanishi, Y., Shinjo, R., Reimer, J.D., Tanaka, J., Maekawa, H., 2012. Soft coral *Sarcophyton* (Cnidaria: Anthozoa: Octocorallia) species diversity and chemotypes. PLoS One 7, e30410.

Beck, G., Coman, D., Herren, E., Ruiz-Sola, M.Á., Rodríguez-Concepción, M., Gruissem, W., Vranová, E., 2013. Characterization of the *GGPP* synthase gene family in *Arabidopsis thaliana*. Plant Mol. Biol. 82, 393–416.

Bouvier, F., Rahier, A., Camara, B., 2005. Biogenesis, molecular regulation and function of plant isoprenoids. Prog. Lipid Res. 44, 357–429.

Buděšínský, M., Valterová, I., Sémon, E., Cancello, E., Bordereau, C., 2005. NMR structure determination of (11*E*)-trinervita-1(14),2,11-triene, a new diterpene from sexual glands of termites. Tetrahedron 61, 10699–10704.

Cane, D.E., 1999. Isoprenoid biosynthesis: overview. In: In: Cane, D.E. (Ed.), Comprehensive Natural Products Chemistry, vol. 2. Pergamon Press, Oxford, pp. 1–13.

Costa-Leonardo, A.M., De Salvo, C.R., 1987. A comparative study of the frontal glands in three species of Brazilian termite soldiers (Isoptera, Termitidae). Rev. Bras. Entomol. 31, 465–471.

Croteau, R., Kutchan, T.M., Lewis, N.G., 2000. Natural products (secondary metabolites). In: Buchanan, B., Gruissem, W., Jones, R. (Eds.), Biochemistry & Molecular Biology of Plants. American Society of Plant Physiologists, Rockville, pp. 1250–1318.

Deligne, J., Quennedey, A., Blum, M.S., 1981. The enemies and defense mechanisms of termites. In: In: Hermann, H.R. (Ed.), Social Insects, vol. II. Academic Press, New York, pp. 1–76.

Eisner, T., Kriston, I., Aneshansley, D.J., 1976. Defensive behavior of a termite (*Nasutitermes exitiosus*). Behav. Ecol. Sociobiol. 1, 83–125.

Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300, 1005–1016.

Engel, M.S., Grimaldi, D.A., Krishna, K., 2009. Termites (Isoptera): their phylogeny, classification, and rise to ecological dominance. Am. Mus. Novit. 3650, 1–27.

Ericsson, J., Greene, J.M., Carter, K.C., Shell, B.K., Duan, D.R., Florence, C., Edwards, P.A., 1998. Human geranylgeranyl diphosphate synthase: isolation of the cDNA, chromosomal mapping and tissue expression. J. Lipid Res. 39, 1731–1739.

Fallon, T.R., Lower, S.E., Chang, C.-H., Bessho-Uehara, M., Martin, G.J., Bewick, A.J., Behringer, M., Debat, H.J., Wong, I., Day, J.C., Suvorov, A., Silva, C.J., Stanger-Hall, K.F., Hall, D.W., Schmitz, R.J., Nelson, D.R., Lewis, S.M., Shigenobu, S., Bybee, S.M., Larracuente, A.M., Oba, Y., Weng, J.K., 2018. Firefly genomes illuminate parallel origins of bioluminescence in beetles. eLife 7, e36495.

Farag, M.A., Al-Mahdy, D.A., Meyer, A., Westphal, H., Wessjohann, L.A., 2017. Metabolomics reveals biotic and abiotic elicitor effects on the soft coral *Sarcophyton ehrenbergi* terpenoid content. Sci. Rep. 7, 648.

Gavagnin, M., Fontana, A., 2000. Diterpenes from marine opisthobranch molluscs. Curr. Org. Chem. 4, 1201–1248.

Goldstein, J.L., Brown, M.S., 1990. Regulation of the mevalonate pathway. Nature 343, 425–430.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.

Hamberger, B., Bak, S., 2013. Plant P450s as versatile drivers for evolution of species-specific chemical diversity. Phil. Trans. R. Soc. B 368, 20120426.

Harrison, M.C., Jongepier, E., Robertson, H.M., Arning, N., Bitard-Feildel, T., Chao, H., Childers, C.P., Dinh, H., Doddapaneni, H., Dugan, S., Gowin, J., Greiner, C., Han, Y., Hu, H., Hughes, D.S.T., Huylmans, A.-K., Kemena, C., Kremer, L.P.M., Lee, S.L., Lopez-Ezquerra, A., Mallet, L., Monroy-Kuhn, J.M., Moser, A., Murali, S.C., Muzny, D.M., Otani, S., Piulachs, M.-D., Poelchau, M., Qu, J., Schaub, F., Wada-Katsumata, A., Worley, K.C., Xie, Q., Ylla, G., Poulsen, M., Gibbs, R.A., Schal, C., Richards, S., Belles, X., Korb, J., Bornberg-Bauer, E., 2018. Hemimetabolous 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., 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. Bull. Am. Mus. Nat. Hist. 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.
Lai, C., McMahon, R., Young, C., Mackay, T.F.C., Langley, C.H., 1998. *quemao*, a Drosophila bristle locus, encodes geranylgeranyl pyrophosphate synthase. Genetics 149, 1051–1061.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2 nat. Methods 9, 357–359.
Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25, 1307–1320.
Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659.
Liang, P.-H., Ko, T.-P., Wang, A.H.-J., 2002. Structure, mechanism and function of prenyltransferases. Eur. J. Biochem. 269, 3339–3354.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMB Net J. 17, 10–12.
Nelson, D., Werck-Reichhart, D., 2011. A P450-centric view of plant evolution. Plant J. 66, 194–211.
Noirot, C., 1989. Social structure in termite societies. Ethol. Ecol. Evol. 1, 1–17.
Noirot, C., Darlington, J.P.E., 2000. Termite nests: architecture, regulation and defence. In: Abe, T., Bignell, D.E., Higashi, M. (Eds.), Termites: Evolution, Sociality, Symbioses, Ecology. Kluwer Academic Publishers, Dordrecht, pp. 121–139.
Noirot, C., Quennedey, A., 1974. Fine structure of insect epidermal glands. Annu. Rev. Entomol. 19, 61–80.
Ogura, K., Koyama, T., 1998. Enzymatic aspects of isoprenoid chain elongation. Chem. Rev. 98, 1263–1276.
Ohnuma, S., Narita, K., Nakazawa, T., Ishida, C., Takeuchi, Y., Ohto, C., Nishino, T., 1996. A role of the amino acid residue located on the fifth position before the first aspartate-rich motif of farnesyl diphosphate synthase on determination of the final product. J. Biol. Chem. 48, 30748–30754.
Pfaffl, M.W., Tichopad, A., Prgomet, C., Neuvians, T.P., 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–excel-based tool using pair-wise correlations. Biotechnol. Lett. 26, 509–515.
Poulsen, M., Hu, H., Li, C., Chen, Z., Xu, L., Otani, S., Nygaard, S., Nobre, T., Klaubauf, S., Schindler, P.M., Hauser, F., Pan, H., Yang, Z., Sonnenberg, A.S.M., de Beer, Z.W., Zhang, Y., Wingfield, M.J., Grimmelikhuijzen, C.J.P., de Vries, R.P., Korb, J., Aanen, D.K., Wang, J., Boomsma, J.J., Zhang, G., 2014. Complementary symbiont contributions to plant decomposition in a fungus-farming termite. Proc. Natl. Acad. Sci. U.S.A. 111, 14500–14505.
Prestwich, G.D., 1979a. Chemical defense by termite soldiers. J. Chem. Ecol. 5, 459–480.
Prestwich, G.D., 1979b. Interspecific variation in the defence secretions of *Nasutitermes* soldiers. Biochem. Syst. Ecol. 7, 211–221.
Prestwich, G.D., 1983a. The chemical defenses of termites. Sci. Am. 249, 68–75.
Prestwich, G.D., 1983b. Chemical systematics of termite exocrine secretions. Annu. Rev. Ecol. Systemat. 14, 287–311.
Prestwich, G.D., 1984. Defense mechanisms of termites. Annu. Rev. Entomol. 29, 201–232.
Prestwich, G.D., 1988. The chemicals of termite societies (Isoptera). Sociobiology 14, 175–191.
Prestwich, G.D., Jones, R.W., Collins, M.S., 1981. Terpene biosynthesis by nasute termite soldiers (Isoptera: Nasutitermitinae). Insect Biochem. 11, 331–336.
Quennedey, A., 1984. Morphology and ultrastructure of termite defense glands. In: Hermann, H. (Ed.), Defensive Mechanisms in Social Insects. Praeger Publishers, New York, pp. 151–200.
Rastogi, S., Liberles, D.A., 2005. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. BMC Evol. Biol. 5, 28.
Roberts, A., Pachter, L., 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. Nat. Methods 10, 71–73.
Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11, R25.
Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.
Roisin, Y., 2000. Diversity and evolution of caste patterns. In: Abe, T., Bignell, D.E., Higashi, M. (Eds.), Termites: Evolution, Sociality, Symbioses, Ecology. Kluwer Academic Publishers, Dordrecht, pp. 95–119.
Roux, E.A., Korb, J., 2004. Evolution of eusociality and the soldier caste in termites: a validation of the intrinsic benefit hypothesis. J. Evol. Biol. 17, 869–875.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Sillam-Dussès, D., Sémon, E., Robert, A., Cancello, E., Lenz, M., Valterovà, I., Bordereau, C., 2010. Identification of multi-component trail pheromones in the most evolutionarily derived termites, the Nasutitermitinae (Termitidae). Biol. J. Linn. Soc. 99, 20–27.
Singh, A.K., Prestwich, G.D., 1986. Synthesis of tritiumlabelled cembrene-A. J. Label. Comp. Radiopharm. 23, 377–382.
Šobotník, J., Jirošová, A., Hanus, R., 2010. Chemical warfare in termites. J. Insect Physiol. 56, 1012–1021.
Tarshis, L.C., Proteau, P.J., Kellogg, B.A., Sacchettini, J.C., Poulter, C.D., 1996. Regulation of product chain length by isoprenyl diphosphate synthases. Proc. Natl. Acad. Sci. U.S.A. 93, 15018–15023.
Terrapon, N., Li, C., Robertson, H.M., Ji, L., Meng, X., Booth, W., Chen, Z., Childers, C.P., Glastad, K.M., Gokhale, K., Gowin, J., Gronenberg, W., Hermansen, R.A., Hu, H., Hunt, B.G., Huylmans, A.K., Khalil, S.M., Mitchell, R.D., Munoz-Torres, M.C., Mustard, J.A., Pan, H., Reese, J.T., Scharf, M.E., Sun, F., Vogel, H., Xiao, J., Yang, W., Yang, Z., Yang, Z., Zhou, J., Zhu, J., Brent, C.S., Elsik, C.G., Goodisman, M.A., Liberles, D.A., Roe, R.M., Vargo, E.L., Vilcinskas, A., Wang, J., Bornberg-Bauer, E., Korb, J., Zhang, G., Liebig, J., 2014. Molecular traces of alternative social organization in a termite genome. Nat. Commun. 5, 3636.
Tian, L., Zhou, X., 2014. The soldiers in societies: defense, regulation, and evolution. Int. J. Biol. Sci. 10, 296–308.
Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3 RESEARCH0034.
Wang, K.C., Ohnuma, S., 2000. Isoprenyl diphosphate synthases. Biochim. Biophys. Acta 1529, 33–48.
Wang, C., Chen, Q., Fan, D., Li, J., Wang, G., Zhang, P., 2016. Structural analyses of short-chain prenyltransferases identify an evolutionarily conserved GFPPS Clade in brassicaceae plants. Mol. Plant 9, 195–204.
Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.
Zhang, F.L., Casey, P.J., 1996. Protein prenylation: molecular mechanisms and functional consequences. Annu. Rev. Biochem. 65, 241–269.
Zhao, Y., Yu, L., Gao, J., Fu, Q., Hua, Y., Zhang, H., Zhao, S., 2000. cDNA cloning, chromosome mapping and expression characterization of human geranylgeranyl pyrophosphate synthase. Sci. China Ser. C Life Sci. 43, 613–622.