

Construction and Characterization of Normalized cDNA Libraries by 454 Pyrosequencing and Estimation of DNA Methylation Levels in Three Distantly Related Termite Species

Yoshinobu Hayashi^{1,2}, Shuji Shigenobu^{3,4}, Dai Watanabe^{1,5}, Kouhei Toga^{5,6}, Ryota Saiki⁵, Keisuke Shimada^{5,7}, Thomas Bourguignon^{1,8}, Nathan Lo², Masaru Hojo⁹, Kiyoto Maekawa⁵, Toru Miura^{1*}

1 Graduate School of Environmental Science, Hokkaido University, Sapporo, Japan, **2** School of Biological Sciences, The University of Sydney, Sydney, New South Wales, Australia, **3** NIBB Core Research Facilities, National Institute for Basic Biology, National Institutes of Natural Sciences, Okazaki, Japan, **4** Department of Basic Biology, School of Life Science, Graduate University for Advanced Studies, Okazaki, Japan, **5** Graduate School of Science and Engineering, University of Toyama, Toyama, Japan, **6** Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan, **7** Ishikawa Museum of Natural History, Kanazawa, Japan, **8** Department of Biological Sciences, National University of Singapore, Singapore, Singapore, **9** Tropical Biosphere Research Center, University of the Ryukyus, Okinawa, Japan

Abstract

In termites, division of labor among castes, categories of individuals that perform specialized tasks, increases colony-level productivity and is the key to their ecological success. Although molecular studies on caste polymorphism have been performed in termites, we are far from a comprehensive understanding of the molecular basis of this phenomenon. To facilitate future molecular studies, we aimed to construct expressed sequence tag (EST) libraries covering wide ranges of gene repertoires in three representative termite species, *Hodotermopsis sjostedti*, *Reticulitermes speratus* and *Nasutitermes takasagoensis*. We generated normalized cDNA libraries from whole bodies, except for guts containing microbes, of almost all castes, sexes and developmental stages and sequenced them with the 454 GS FLX titanium system. We obtained >1.2 million quality-filtered reads yielding >400 million bases for each of the three species. Isotigs, which are analogous to individual transcripts, and singletons were produced by assembling the reads and annotated using public databases. Genes related to juvenile hormone, which plays crucial roles in caste differentiation of termites, were identified from the EST libraries by BLAST search. To explore the potential for DNA methylation, which plays an important role in caste differentiation of honeybees, tBLASTn searches for DNA methyltransferases (*dnmt1*, *dnmt2* and *dnmt3*) and methyl-CpG binding domain (*mbd*) were performed against the EST libraries. All four of these genes were found in the *H. sjostedti* library, while all except *dnmt3* were found in *R. speratus* and *N. takasagoensis*. The ratio of the observed to the expected CpG content (CpG O/E), which is a proxy for DNA methylation level, was calculated for the coding sequences predicted from the isotigs and singletons. In all of the three species, the majority of coding sequences showed depletion of CpG O/E (less than 1), and the distributions of CpG O/E were bimodal, suggesting the presence of DNA methylation.

Citation: Hayashi Y, Shigenobu S, Watanabe D, Toga K, Saiki R, et al. (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(9): e76678. doi:10.1371/journal.pone.0076678

Editor: Ken Mills, Queen's University Belfast, United Kingdom

Received: July 10, 2013; **Accepted:** September 1, 2013; **Published:** September 30, 2013

Copyright: © 2013 Hayashi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by a Grant-in-Aid for Young Scientists (No. 21677001 to TM) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (<http://www.mext.go.jp/english/>), and by Grant-in-Aid for Young Scientists (No. 23770274 to MH), for JSPS Fellows (Nos. 2469, 12J03468 and 249520 to YH, DW and KT respectively) and for Scientific Research (C) (No. 24570022 to KM) from the Japan Society for the Promotion of Science (<http://www.jsps.go.jp/english/>). YH was also supported by Grant for Basic Science Research Projects from Sumitomo Foundation (<http://www.sumitomo.or.jp/e/>). NL was supported by the Australian Research Council (<http://www.arc.gov.au/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors declare that they have no competing interests.

* E-mail: miu@ees.hokudai.ac.jp

Introduction

Termites, one of the major social insect groups, live in colonies and construct complex societies with highly

sophisticated division of labor among castes, which show distinctive behavior and morphology for their specialized tasks [1,2]. Division of labor is the key to efficient colony

performance, leading to the ecological success of termites particularly in tropical and subtropical terrestrial regions [2].

Distinct castes express different sets of genes and are differentiated from each other through differential gene expression during postembryonic development in response to environmental factors [3,4]. To date, a significant number of studies that focused on differential gene expression among castes and among individuals in the course of caste differentiation have been reported in termites (reviewed in Miura & Scharf [4]). These studies successfully identified genes with caste-specific or -biased expressions, and up- or down-regulated genes during caste differentiation. However, only a few castes in a few species have been investigated. Termites are represented by more than 2600 extant species [5] and exhibit considerable diversity in morphology and behavior of castes and caste developmental pathways [6]. At present, a comprehensive understanding of caste polymorphism of termites from the molecular viewpoint is still far from being reached. To facilitate gene expression analyses in termites, it is important to construct expressed sequence tag (EST) libraries, which are used as an information source for gene discovery, gene structure identification, and so on [7-10].

To date, large-scale EST libraries have been constructed in only a handful of termite species. In *Coptotermes formosanus* (Rhinotermitidae), Husseneder et al. [11] obtained 4,726 ESTs generated from an assembly of 7,663 sequence reads by Sanger sequencing of a cDNA library originating from various castes with normalization of cDNA (which equalizes transcript concentrations in a cDNA library) [12,13]. Zhang et al. [14] generated normalized cDNA library from workers, soldiers, nymphs, and male and female alates of *C. formosanus*, and obtained 16,691 contigs and 9,248 singletons that resulted from assembly of 131,636 Sanger sequence reads. In *Reticulitermes flavipes*, a total of 15,259 Sanger reads were obtained from non-normalized cDNA generated from alates, workers, soldiers, and larvae by Steller et al. [15]. Recently Hojo et al. [16] performed 454 pyrosequencing for transcriptomes of frontal gland tissues to identify genes involved in the synthesis of terpenoid defensive secretion in *Nasutitermes takasagoensis* (Termitidae) and generated 1,189 contigs that were assembled from 50,290 clean reads. Huang et al. [17] used heads of workers of *Odontotermes formosanus* for Illumina sequencing and generated 57 million reads that were assembled into 221,728 contigs. In addition to those large-scale EST sequencing, a number of smaller scale EST projects were carried out in termites. As of June 2013, 164,150 ESTs from termite species were deposited in the dbESTs of NCBI GenBank (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html) and 3 projects that produced 526,778 sequences in total in the SRA of GenBank, including a metagenomic analysis that primarily focused on gut symbionts of termites. However, the EST libraries generated to date are likely to be missing significant parts of the termite transcriptome gene repertoire. cDNA library construction using many castes, as well as massive sequencing with next generation technologies, are both required to obtain a wide range of the gene repertoire. To date, no study has combined both of these aspects.

In this study we generated normalized cDNA libraries from RNA pools of whole bodies (excluding the digestive tracts) of almost all castes, sexes and developmental stages in three species of termites, *Hodotermopsis sjostedti* (Termopsidae), *R. speratus* (Rhinotermitidae) and *N. takasagoensis* (Termitidae). We sequenced each of the libraries using 454 pyrosequencing technology (GS FLX Titanium System). The three termite species belong to different families that are phylogenetically distant from one another; *H. sjostedti* is at relatively basal position, *R. speratus* is in an intermediate position, and *N. takasagoensis* is at the most apical part in the termite phylogenetic tree [18]. Moreover, their caste developmental pathways markedly differ (Figure S1): *H. sjostedti* has a linear developmental pathway, lacking a true worker caste, while *R. speratus* and *N. takasagoensis* have a bifurcated developmental pathway with a true worker caste. Morphology of the castes also differs, especially among soldiers; soldiers of *H. sjostedti* and *R. speratus* have elongated mandibles for biting enemies, while those of *N. takasagoensis* have reduced mandibles and enlarged frontal glands which they use to spray chemicals against enemies. Importantly, manipulation experiments for studying caste differentiation, such as hormone application and RNA interference techniques, have been established in these species (for example, Ogino et al. [19] and Hattori et al. [20] for *H. sjostedti*, Tsuchiya et al. [21] and Nambu et al. [22] for *R. speratus*, and Toga et al. [23] for *N. takasagoensis*). With these techniques, molecular and physiological studies on caste differentiation have been performed and have provided a significant amount of knowledge on the molecular basis of caste polymorphism (reviewed in Miura & Scharf [4]). Construction of large-scale EST databases for these species will allow us to carry out gene expression analyses of caste polymorphism more efficiently.

In addition to EST library construction, we explored the possibility of DNA methylation in the three termite species by examining whether gene sequences of DNA (cytosine-5) methyltransferases, namely, *dnmt1*, *dnmt2* and *dnmt3*, and methyl-CpG binding domain, *mbd*, were present in the EST libraries. Furthermore, we estimated DNA methylation levels of coding sequences by computational analysis on the EST data. DNA methylation is known to be one of the most important mechanisms for generating differential gene expression triggered by environmental cues [24]. In animals, DNA methylation mostly occurs in cytosine nucleotides that are located next to guanine nucleotides, which are known as CpG dinucleotides [25,26]. In invertebrates, exon regions are the main targets of DNA methylation [27,28]. Furthermore, it is known that DNA methylation can be predicted from normalized CpG nucleotide content, CpG O/E, which is the ratio of observed and expected frequencies of CpG sequences [26]. Due to hyper mutability of methylated cytosines to thymines [29,30], CpG O/E of heavily methylated genes in germlines decreases over evolutionary time [26,30,31]. Here, we calculated CpG O/E of coding sequences predicted from the EST data and discussed DNA methylation levels of the three termite species.

Results and Discussion

Normalized cDNA library construction

The primary purpose of this study is to construct EST libraries that contain wide ranges of expressed-gene repertoires and thus we collected as many castes, sexes, and developmental stages of termite individuals as possible, because the different types of individuals may express different sets of genes [4]. We extracted total RNA from 18, 26, and 25 categories of individuals that were classified based on caste/sex/developmental stage in *Hodotermopsis sjostedti*, *Reticulitermes speratus*, and *Nasutitermes takasagoensis*, respectively (Tables S1-S3). These RNA transcripts from different categories were pooled for each of the species. To facilitate the sequencing of rare transcripts, we applied a normalization technique to the cDNA generated by reverse transcription of the RNA.

Sequencing and assembly of normalized cDNA libraries

The normalized cDNA libraries were sequenced with a 454 GS FLX Titanium system (Roche, Indianapolis, IN, USA). All of the reads obtained have been deposited in the DDBJ Sequence Read Archive (DRA) database under accession numbers DRA000538 and DRA001044 for *H. sjostedti*, DRA001045 for *R. speratus*, and DRA001046 for *N. takasagoensis*. After removing adaptor sequences and low quality nucleotides, 1.22 M, 1.32 M, and 1.39 M reads in *H. sjostedti*, *R. speratus*, and *N. takasagoensis*, respectively, were retained for further analyses, yielding a total of 401.8 M, 443.0 M, and 517.6 M bases (Table 1).

To assemble the cleaned reads, we used a GS de novo Assembler (Newbler) version 2.5.3 (Roche) with cDNA mode. The assembling produced 49,919 isotigs, 90 contigs and 83,549 singletons in *H. sjostedti*, 55,476 isotigs, 160 contigs and 87,191 singletons in *R. speratus*, and 27,408 isotigs, 151 contigs and 128,438 singletons in *N. takasagoensis*. For simplicity, isotigs and contigs are hereafter collectively referred to as isotigs ($n = 50,009$ in *H. sjostedti*; 55,636 in *R. speratus*; and 27,559 in *N. takasagoensis*; Table 1). Singletons less than 100 bp ($n = 54,007$ in *H. sjostedti*; 47,065 in *R. speratus*; 25,315 in *N. takasagoensis*) were discarded for subsequent analyses.

Transcriptome completeness

To assess transcriptome completeness of the EST libraries, we surveyed how many genes out of gene sets conserved among taxa are present in the EST libraries. We performed BLASTX searches for the isotigs and singletons with various E-value thresholds ($\leq 1e-5$, $\leq 1e-10$, $\leq 1e-15$ and $\leq 1e-20$) against Core Eukaryotic Genes (CEGs), which consist of 458 conserved genes [32]. The BLASTX results indicated that most of the CEGs (98.3%, 98.3%, and 99.1% in *H. sjostedti*, *R. speratus* and *N. takasagoensis* respectively; Table 2) showed significant similarity to at least one of the isotigs and singletons even with the most stringent E-value threshold, $1e-20$.

Furthermore, transcriptome completeness was assessed also by using gene sets conserved among insects. OrthoDB6 database defined orthologous gene groups among taxa [33]

Table 1. Sequencing and assembly statistics in three termite species.

	<i>Hodotermopsis sjostedti</i>	<i>Reticulitermes speratus</i>	<i>Nasutitermes takasagoensis</i>
No. of reads	1,221,634	1,317,986	1,387,437
No. of bases	402,779,218	444,211,498	518,789,889
No. of clean reads	1,221,416	1,317,777	1,387,263
No. of clean bases	401,761,320	443,013,802	517,610,329
No. of isogroups	41,306	43,201	16,635
No. of isotigs	50,009	55,636	27,559
mean length of isotigs	621.5	654.6	820.0
median length of isotigs	494	506	681
N50 length of isotigs	711	773	974
No. of singletons	83,549	87,191	128,438
mean length of singletons	155.1	184.6	304.3
median length of singletons	71	80	346
N50 length of singletons	339	359	411
No. of singletons ≥ 100 bp	29,542	40,126	103,123
mean length of singletons ≥ 100 bp	318.1	321.7	358.6
median length of singletons ≥ 100 bp	337	338	381
N50 length of singletons ≥ 100 bp	399	388	411

doi: 10.1371/journal.pone.0076678.t001

(<http://cegg.unige.ch/orthodb6>). We retrieved protein sequences of the orthologs that were conserved in all, all but one, and $\geq 90\%$ of the insect species registered in OrthoDB6 (278, 1,332 and 4,969 gene groups respectively). Then BLASTX searches were performed for the isotigs and singletons against the OrthoDB6 gene sets. The BLASTX searches revealed that more than 94%, 93% and 88% of genes conserved among all, among all but one and among $\geq 90\%$ of the insect species, respectively, were detected from the EST libraries with a $1e-20$ E-value threshold (Table 2).

These results suggest that the EST libraries cover most of the genes conserved among broad ranges of taxa.

Functional annotation

For functional annotation, the isotigs and singletons were subjected to BLASTX searches against the non-redundant (nr) database of GenBank with threshold values of $\leq 1e-4$ for E-value and ≥ 40 for bit score. We found that 29,249 (36.8%) [24,003 (48.0%) and 5,246 (17.8%)] of the isotigs and the singletons, respectively] in *H. sjostedti*, 32,191 (33.6%) [25,982 (46.7%) and 6209 (15.5%)] in *R. speratus*, and 42,815 (32.8%) [17,585 (63.8%) and 25,230 (24.5%)] in *N. takasagoensis* showed similarity to at least one gene from the nr database based on the threshold. The lack of BLASTX hits for the

Table 2. Gene coverage of EST libraries of three termite species.

Gene set	Hodotermopsis sjostedti			Reticulitermes speratus			Nasutitermes takasagoensis					
	1e-5	1e-10	1e-15	1e-20	1e-5	1e-10	1e-15	1e-20	1e-5	1e-10	1e-15	1e-20
Core Eukaryotic Genes [458] [*]	458 (100.0)	457 (99.8)	455 (99.8)	450 (98.3)	457 (99.8)	455 (99.3)	453 (98.9)	450 (98.3)	458 (100.0)	457 (99.8)	456 (99.6)	454 (99.1)
orthologs present in all of the insect species [278] [*]	272 (97.8)	270 (97.1)	267 (96.0)	262 (94.2)	275 (98.9)	274 (98.6)	273 (98.2)	268 (96.4)	276 (99.3)	272 (97.8)	269 (96.8)	265 (95.3)
orthologs present in all but one of the insect species [1332] [*]	1312 (98.5)	1287 (96.6)	1259 (94.5)	1245 (93.5)	1319 (99.0)	1299 (97.5)	1278 (95.9)	1258 (94.4)	1315 (98.7)	1298 (97.4)	1273 (95.6)	1259 (94.5)
orthologs present in > 90% of the insect species [4969] [*]	4665 (93.9)	4569 (92.0)	4475 (90.1)	4392 (88.4)	4737 (95.3)	4652 (93.6)	4576 (92.1)	4473 (90.0)	4714 (94.9)	4621 (93.0)	4530 (91.2)	4438 (89.3)

Numbers (percentages) of the Core Eukaryotic Genes and OrthoDB6 orthologous gene groups across insects to which ESTs of termites showed BLASTX hit with E-values $\leq 1e-5$, $\leq 1e-10$, $\leq 1e-15$, and $\leq 1e-20$ are shown.
^{*} The number in brackets indicates the total number of the Core Eukaryotic Genes or the orthologous gene groups obtained from OrthoDB6 database.
doi: 10.1371/journal.pone.0076678.t002

majority of the isotigs and singletons may be due to a small number of deposited protein sequences of termites and closely related insects, i.e., cockroaches and mantises.

Based on the top hits of BLASTX searches, we assigned Gene Ontology (GO) annotation by using ID mapping data of the UniProt database. Of the total isotigs and singletons, 13,141 (16.5%) [11,774 (23.5%) and 1367 (4.6%) of the isotigs and the singletons, respectively], 14,435 (15.1%) [11664 (21.0%) and 2771 (6.9%)], and 19,051 (14.6%) [7942 (28.8%), 11109 (10.8%)] were annotated with GO terms in *H. sjostedti*, *R. speratus* and *N. takasagoensis*, respectively. Frequency of genes categorized into each of GO terms is shown in Figure 1.

We used the InterProScan software [34] with the Pfam database to search for functional protein domains in the isotigs and the singletons. In all of the termite species, the three most frequent domains were 'Zinc-finger double domain' (PF13465), 'WD domain, G-beta repeat' (PF00400), and 'Zinc finger, C2H2 type' (PF00096), with frequencies of 955, 471, and 234 respectively in *H. sjostedti*, 955, 471, and 234 in *R. speratus*, and 1303, 440 and 318 in *N. takasagoensis* (Table S4). This analysis also revealed that among the isotigs and singletons with no hits in the BLASTX searches, 136, 168, and 200 were predicted to contain functional protein domains.

Identification of juvenile hormone-related genes

Juvenile hormone (JH) plays crucial roles in caste differentiation in termites [35,36]. The EST libraries were searched for JH-related genes that are listed in Table 3 using the TBLASTN algorithm with E-value threshold of $1e-20$. As query sequences of the TBLASTN searches, we used insect orthologs of JH-related genes that were listed in Table 3 of The International Aphid Genomics Consortium [37]. We detected all of the JH-related genes from the EST libraries, except for 'allatostatin receptor' in the three termite species and 'methoprene-tolerant' in *H. sjostedti*. This information will be useful for future molecular studies on caste differentiation.

Species distribution of BLASTX top hit genes

We examined species distributions of the top-hit genes in the BLASTX analysis with the nr database. In all of the termite species, *Tribolium castaneum* was the most frequent species in the distributions [2788 (11.7%), 2894 (11.8%), and 4256 (11.8%) of total isotigs and singletons with BLASTX hits in *H. sjostedti*, *R. speratus*, and *N. takasagoensis*, respectively; Table S5], followed by *Pediculus humanus* [2562 (10.7%), 2645 (10.8%), and 3935 (10.9%)]. The protist *Trichomonas vaginalis* was also present in the distributions but with much fewer hits [652 (2.7%), 251 (1.0%), 23 (0.06%)]. This is likely to result from contamination of RNA of symbiotic intestinal protists into the termite RNA samples, despite the fact that we removed guts of termites before RNA extraction. It is well known that lower termites such as *H. sjostedti* and *R. speratus* harbor symbiotic protists belonging to the order Trichomonadida [38,39], and the presence of symbiotic protists has also been suggested in *Nasutitermes* [40]. By summarizing counts of the species distribution, we found that more than 80% of blast top hits (81.3%, 81.9% and 84.1%) were to insects and other arthropods, and 5% or less were to protists and bacteria

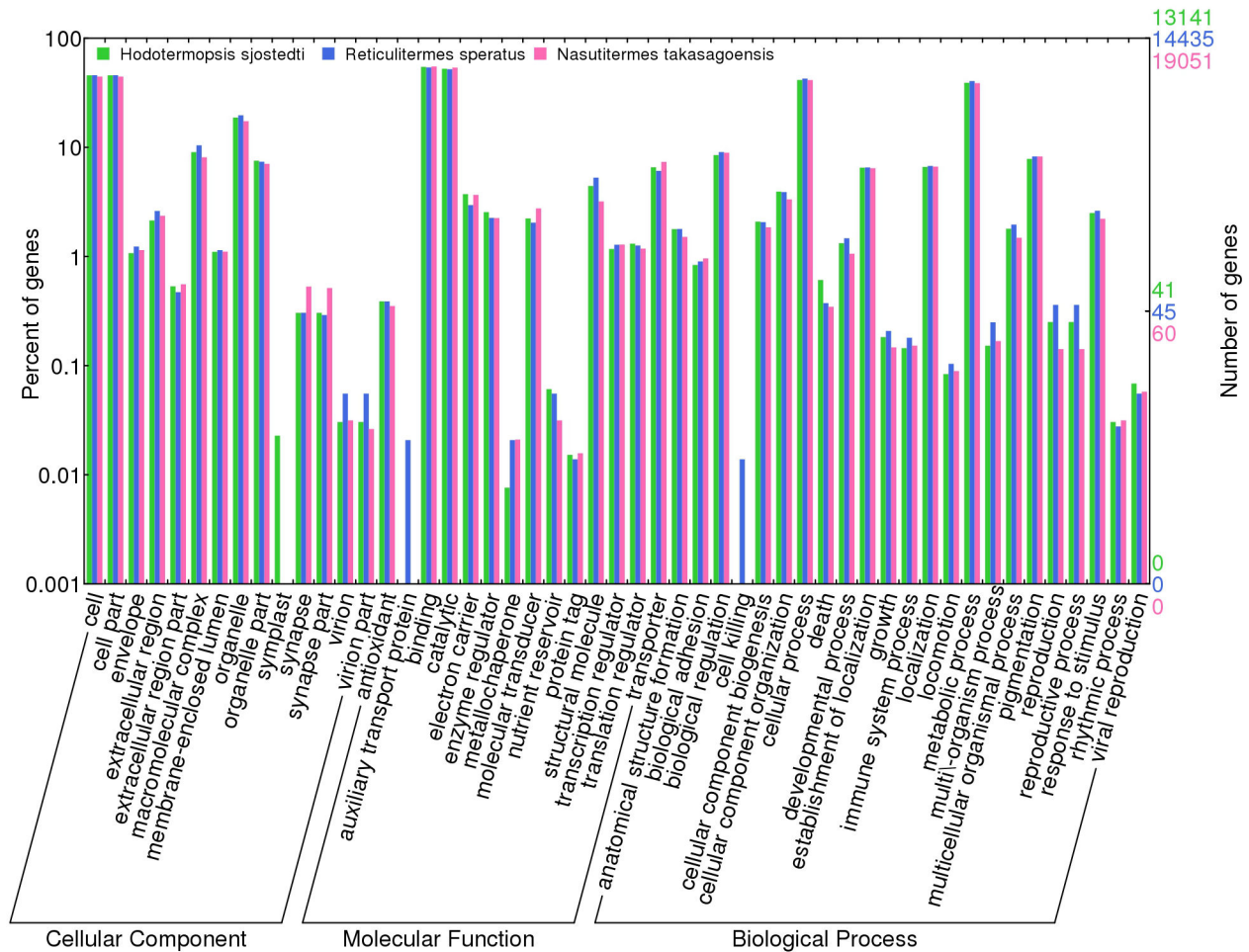


Figure 1. Gene Ontology annotation of EST libraries of three termite species. Frequency and percentage of the isotigs and singletons annotated by Gene Ontology terms are shown.

doi: 10.1371/journal.pone.0076678.g001

(protists: 3.7%, 2.3% and 1.0%; bacteria: 1.3%, 1.3% and 0.84%). Compared to an EST library of *Coptotermes formosanus* in which 42% of the top BLAST hits originated from protists and bacteria [11], the EST libraries constructed in this study contained much fewer ESTs derived from microorganisms.

Orthologous genes among three termite species

Protein coding regions were predicted and extracted from the isotigs and singletons by OrfPredictor [41], and orthology of the predicted protein sequences among the three termite species was determined by InParanoid [42] and Multiparanoid [43]. As a result, we found 7337 orthologous gene groups that were shared among the three termite species (Figure 2). Among these orthologous gene groups, 377 did not show sequence similarity in the aforementioned BLASTX searches against the nr database. These gene groups without similarity to known genes potentially include novel genes that have evolved in the termite lineage and are widespread among extant termite

species. However, to determine whether these genes are unambiguously termite-specific novel genes, it is necessary to examine transcriptome or genome of close relatives of termites such as the wood roach *Cryptocercus* [44], neither of which have been generated so far.

We further examined the putative termite-specific genes by using the results of the InterProScan searches with Pfam database. We found that of the 377 genes seven contained Pfam motifs: two of them had 'PBP/GOBP family' (Pfam ID: PF01395) domain and the rest contained 'EB module' (PF01683), 'Regulator of G protein signalling domain' (PF00615), 'THAP domain' (PF05485), 'Cystatin domain' (PF00031), and 'Zinc-finger of C2H2 type' (PF12874). It is interesting that genes categorized in the PBP/GOBP (Pheromone/general odorant binding protein) family were candidates of the termite-specific novel genes, because, in social insects, pheromonal communication is a very important component of colony organization [45-47]. These PBP/GOBP

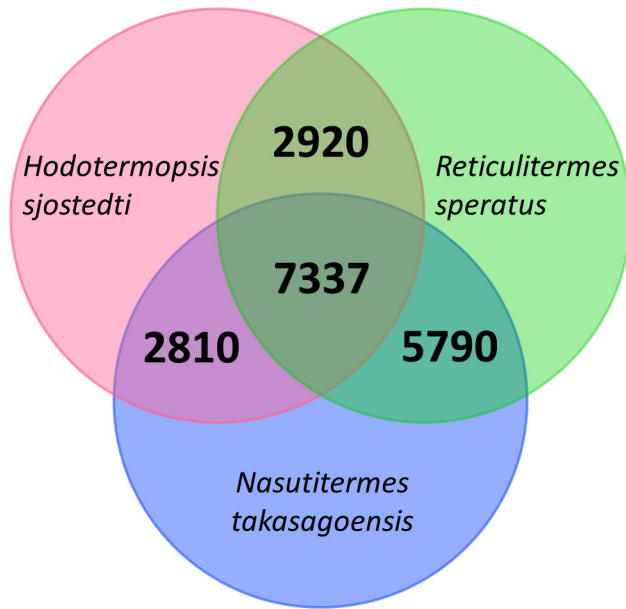


Figure 2. Gene orthology among three termite species. Venn diagram showing overlap of orthologous gene groups among *Hodotermopsis sjostedti*, *Reticulitermes speratus* and *Nasutitermes takasagoensis*.

doi: 10.1371/journal.pone.0076678.g002

family proteins might have evolved in association with the new communication functions for social life.

DNA methylation

We explored the possibility of DNA methylation in the three termite species by examining the presence of DNA methyltransferases (*dnmt1*, *dnmt2* and *dnmt3*) and methyl-CpG binding domain (*mbd*) in the EST libraries. DNMT1 is required to maintain pre-existing methylation patterns in the newly synthesized strand during DNA replication [48]. Although DNMT2 was considered to be a DNA methyltransferase, it is now recognized as a tRNA methyltransferase [49,50]. DNMT3, known as a *de novo* methyltransferase, establishes methylation patterns on unmethylated DNA [51]. MBD proteins have motifs that specifically recognize and are responsible for binding to methyl-CpG [52]. We carried out TBLASTN searches against the EST libraries by using protein sequences of DNMTs and MBDs derived from *Acyrtosiphon pisum*, *Apis mellifera*, *Nasonia vitripennis*, *Pediculus humanus* and *Tribolium castaneum* as query sequences (E-value threshold of $1e-20$). The result suggests that all *dnmts* and *mbd* were present in *H. sjostedti* (Table 3). In *R. speratus* and *N. takasagoensis*, *dnmt1*, *dnmt2*, and *mbd* genes were suggested to be present, while *dnmt3* sequences were not detected at the threshold. However, DNMT3 sequences of *A. mellifera* and *N. vitripennis* showed similarity with E values of $2e-19$ and $1e-7$ to the ESTs of *R. speratus* and *N. takasagoensis*, respectively. In these two species, a further study is needed to clarify the presence of *dnmt3*. These results suggest occurrence of DNA methylation at least in *H. sjostedti*, as suggested in previous studies in

rhinotermitid termites, *C. lacteus* [53], *C. formosanus* and *R. flavipes* [54].

To estimate levels of DNA methylation in coding regions, we calculated CpG O/E of the coding sequences predicted from the isotigs and singletons. We also examined distribution of CpG O/E for bimodality by using NOCOM software [55]. In all of the termite species, the distributions were bimodal, and regarded as mixtures of two normal distributions (log-likelihood ratio test for mono- vs bi-modal distribution model, $\chi^2 = 1333.4$, *d.f.* = 2, $p < 0.001$ in *H. sjostedti*; $\chi^2 = 1056.4$, *d.f.* = 2, $p < 0.001$ in *R. speratus*; $\chi^2 = 721.6$, *d.f.* = 2, $p < 0.001$ in *N. takasagoensis*; Figure 3). The estimated mean \pm SD of the two mixed normal distributions were 0.39 ± 0.14 and 0.81 ± 0.14 , 0.41 ± 0.15 and 0.82 ± 0.15 , and 0.44 ± 0.17 and 0.83 ± 0.17 , with the proportion of the former normal distribution of 0.84, 0.81, and 0.81 in *H. sjostedti*, *R. speratus* and *N. takasagoensis* respectively. This result indicates that the majority of the genes showed low CpG O/E values (less than 1). Depletion of CpG O/E and bimodality of its distribution are representative features of methylated genomes [54]. Because methylated cytosines have a tendency to mutate to thymines through deamination, CpG O/E of heavily methylated regions decreases, and instead, TpG O/E and CpA O/E increase over evolutionary time [30]. In the three termite species, the distributions of TpG O/E and CpA O/E were shifted to the right compared to the distributions of the other dinucleotide combinations (Figures S2-S4). This result also supports DNA methylation in the majority of the genes of the three termites.

To characterize coding sequences with relatively low- and high-CpG O/E, we first classified them into either low- or high-CpG O/E class by using the intersection of the two normal curves of CpG O/E obtained by NOCOM analysis as threshold of the classification. Then, we determined enriched GO terms in low- and high-CpG O/E genes. In all of the three species, 'cellular process' (GO:0009987), 'localization' (GO:0051179) and 'establishment of localization' (GO:0051234), all of which are belonging to the BP category, were significantly enriched in the low-CpG O/E class (Figure S5). On the other hand, 'extracellular region' (GO:0005576) of the CC category, 'structural molecule activity' (GO:0005198), 'transporter activity' (GO:0005215), 'electron carrier activity' (GO:0009055) and 'molecular transducer activity' (GO:0060089) of the MF category were significantly enriched in the high-CpG O/E class. A previous study of DNA methylation [54] showed that 'extracellular region' and 'structural molecule activity' were enriched in the high-CpG O/E class also in *R. flavipes* and in *C. formosanus*, respectively. On the other hand, in the low-CpG O/E class, there were no consistent GO categories with the previous study. As suggested in the previous study [54], the inconsistency in GO categories enriched in the high- and low-CpG O/E genes might be due to differences in transcriptome completeness between EST libraries, rather than due to species-specific methylation change over evolutionary time.

Conclusions

We generated normalized cDNA libraries using almost all castes, sexes and developmental stages, and sequenced them

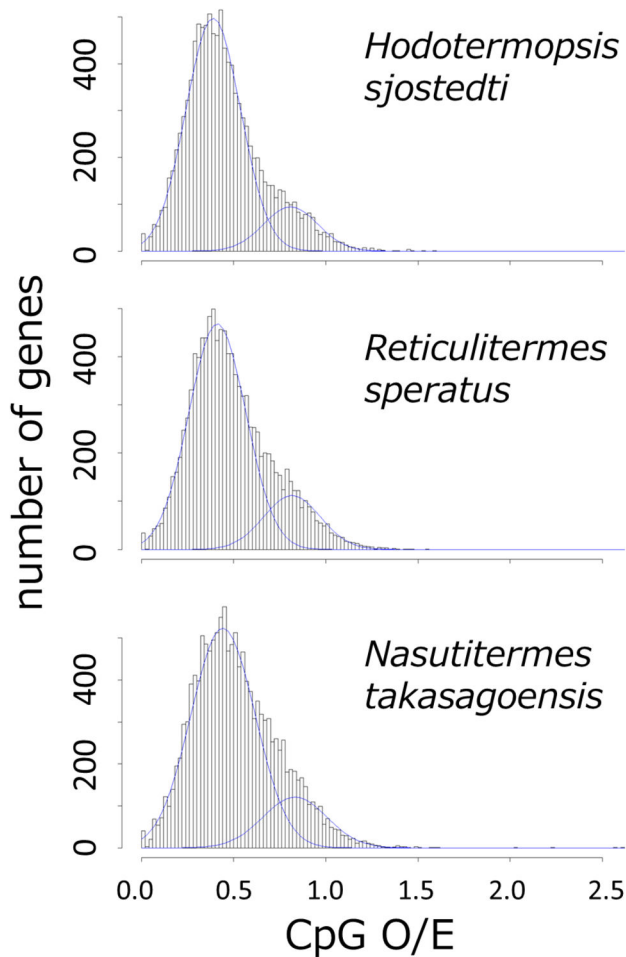


Figure 3. Distribution of normalized CpG content (CpG O/E) in three termite species. Predicted normal curves were fitted to the distributions.

doi: 10.1371/journal.pone.0076678.g003

using the 454 GS FLX Titanium system in three species of termites, *Hodotermopsis sjostedti*, *Reticulitermes speratus* and *Nasutitermes takasagoensis*. The EST libraries of the three species contained most of the genes conserved among a wide range of taxa, that is, CEGs and genes conserved among the insects registered in OrthoDB6. Genes that are not conserved among the taxa, that is, lineage-specific genes, are also likely to be present in the EST libraries. The result of BLASTX searches with the nr database showed that more than a half of ESTs did not exhibit similarity to known genes. Such ESTs without BLASTX hits may contain lineage-specific genes. Therefore, the EST libraries of the three species of termites are expected to cover most of their transcriptomes and be useful for future molecular biological studies on termites.

Our computational analyses suggested that DNA methylation occurs in the three species of termites. However, it is still unclear whether caste differentiation is influenced by DNA methylation in termites. DNA methylation is a candidate proximate mechanism for generating polyphenism in insects

[56,57]; for example, in the honey bee, *Apis mellifera*, DNA methylation was suggested to influence caste development [58]. In two rhinotermitid termites, *C. formosanus* and *R. flavipes*, CpG O/E was associated with gene expression profile among castes [54]. However, on the other hand, in the termite *C. lacteus*, caste-specific DNA methylation was not detected by methylation-sensitive amplified fragment length polymorphism analysis [53]. Thus, further extensive analyses on association between DNA methylation and caste polymorphism are needed to understand social organization from a molecular perspective in termites.

Methods

Termite collection

For EST library construction, we used termite individuals derived from 13 colonies of *Hodotermopsis sjostedti* that were collected during 2002-2011 from Yakushima Island, Kagoshima Prefecture, Japan. Seven colonies of *Reticulitermes speratus* were collected in 2010 from Toyama and Ishikawa Prefectures, Japan. Three colonies of *Nasutitermes takasagoensis* were collected in 2011 from Iriomote Island, Okinawa Prefecture, Japan. Until RNA extraction or rearing experiments, the termites were maintained with nest materials at room temperature ($25\pm 1^\circ\text{C}$) in the laboratory, except for nymphs of *H. sjostedti*, which had been stored at -80°C , and for a king, a queen and N3 nymphs of *N. takasagoensis*, which had been stored in RNAlater (Ambion, Austin, TX, USA) at -80°C (see Tables S1-S3 for detailed description of castes).

No specific permits were required for the described field studies and the locations are not privately-owned or protected in any way. All the three termite species are not endangered or protected species.

Alate pairing and establishment of incipient colonies

To obtain queens, kings, small presoldiers and small soldiers of *H. sjostedti* (see Table S1 for descriptions of these castes), we established incipient colonies by pairing alates, which become kings and queens after dealation and initiation of colonies. Alates derived from two of the collected colonies were paired and reared in plastic boxes (50 mm \times 60 mm \times 25 mm) filled with decaying wood flakes at room temperature for four months. After four months, kings, queens, small presoldiers and small soldiers were collected and immediately used for RNA extraction.

Similarly, in *R. speratus*, to obtain queens, kings, eggs, and larvae, we paired alates derived from two of the field colonies and reared them in glass vials ($\phi 21$ mm \times 45 mm) filled with decaying wood flakes at room temperature. After seven months from the colony establishment, kings, queens, eggs, and larvae were collected for RNA extraction.

Induction of presoldier differentiation by juvenile hormone analog treatment

To obtain RNA of pseudergates/workers that are in the process of developing into presoldiers, we carried out presoldier induction experiments with juvenile hormone

analogs (JHAs) in *H. sjostedti*, *R. speratus*, and *N. takasagoensis*. For the induction experiments in *H. sjostedti* we established experimental colonies, each of which consisted of 10 pseudergates, in a petri dish of 70 mm in diameter with filter paper contained 10 µg of pyriproxyfen; other conditions of the induction were based on Ogino et al. [19]. In *R. speratus*, we followed Tsuchiya et al. [21] for presoldier induction and used 80 µg of juvenile hormone III per experimental colony. In *N. takasagoensis*, we induced presoldiers with 100 µg of hydrophore according to Toga et al. [59]. Presoldiers induced by JHA were collected for total RNA extraction.

Induction of neotenic differentiation

To obtain RNA of neotenic, namely nymphoids and ergatoids, of *R. speratus*, we induced neotenic differentiation in artificially-established colonies that lacked reproductive individuals. Under some conditions (e.g. in the absence of reproductive individuals), nymphs and workers of *R. speratus* differentiate into nymphoids and ergatoids, respectively, through special molt [60,61]. To collect nymphoids, we established artificial colonies, each of which was composed of 20 nymphs, and maintained them as described in Saiki & Maekawa [62]. We collected some nymphoids that appeared in the artificial colonies within 24 hrs from molting, and referred them as 'newly emerged nymphoid' (Table S2). The other nymphoids, referred as 'nymphoid' (Table S2), were continued to be reared for a while and then collected. To collect ergatoids, we established 10 artificial colonies, each of which was composed of 50 workers and maintained in a plastic box (50 mm × 60 mm × 25 mm) lined with moistened filter paper at 25°C under constant darkness. We collected ergatoids that were differentiated from the workers in the artificial colonies. These nymphoids and ergatoids were used for RNA extraction.

Total RNA extraction

For RNA extraction, we used termite individuals that were classified into 18, 26, and 25 categories on the basis of their castes/sexes/developmental stages for *H. sjostedti*, *R. speratus*, and *N. takasagoensis*, respectively (Tables S1-S3). Caste and developmental stage were determined based on external morphology and body size, and sex was determined based on morphology of sternites, as described in previous studies (caste and developmental stage: Miura et al. [63,64] for *H. sjostedti*, Kawamura [65] and Takematsu [66] for *R. speratus*, and Hojo et al. [67] for *N. takasagoensis*; sex determination: Miura et al. [63] for *H. sjostedti*, Hayashi et al. [68] for *R. speratus*, and Hojo et al. [67] for *N. takasagoensis*).

Before RNA extraction, to reduce contamination of RNA of intestinal symbiotic protozoans, we dissected out and removed the digestive tracts of the termites, except for young-instar larvae, which lack the symbionts, using forceps. For the first purification of total RNA we used combination of ISOGEN and High-Salt Precipitation Solution (Nippon Gene, Toyama, Japan), following the protocol provided by the manufacturer after termites were homogenized in the ISOGEN solution, and obtained total RNA dissolved in 50-100 µl of nuclease-free water. The total RNA was then subjected to DNase treatment and second purification with SV Total RNA Isolation System

(Promega, Madison, WI). Finally the total RNA was dissolved in appropriate volumes (80-160 µl) of nuclease-free water.

Normalized cDNA library construction for 454 pyrosequencing and sequence assembling

Total RNAs from the different castes/sexes/developmental stages of each species were pooled together. The pooled total RNA was then reverse transcribed with M-MLV Reverse Transcriptase (Promega, Madison, WI, USA) and the primer 5'-CAAGCAGAAGACGGC ATACGACTGGAG(T)₁₆VN-3' (where V is A, C, or G, and N is any nucleotide) containing a *GsuI* recognition site. Subsequently, after oxidation of the diole group, biotinylation of the 5' Cap structure of mRNA, and RNase I treatment, full-length cDNA/RNA hybrids were selectively captured by magnetic beads attaching streptavidin (Streptavidin Sepharose High Performance; GE healthcare, Piscataway, NJ, USA). The RNA was decomposed by incubation in 50 mM NaOH at 37°C and full-length single stranded cDNA was purified. Then, two different, partially double-stranded adaptors (adaptor GN5 and N6), each of which had a *GsuI* recognition site, were mixed in a ratio of 4:1 and attached to the single strand cDNA by using DNA Ligation Kit, Mighty Mix (Takara, Ohtsu, Japan). The adaptor GN5 was composed of 5'-AATGATACGGCGCTGGAGGACAGGTTTCAGAGTTC(N)₅-3' and its combining complement 3'-TTACTATGCCGCGACCTCCTGTCCAAGTCTCAAG-5'. The adaptor N6 was composed of 5'-AATGATACGGCGCTGGAGGACAGGTTTCAGAGTTC(N)₆-3' and its combining complement 3'-TTACTATGCCGCGACCTCCTGTCCAAGTCTCAAG-5'. Following the adaptor ligation, the second strand cDNA was synthesized with Takara Bio, LA Taq (Takara) and the primer 5'-AATGATACGGCGCTGGAGGACAGGTTTCAGAGTTC-3'. To minimize difference in abundance among different transcripts, the double stranded cDNA was normalized. For normalization, the cDNA was first denatured at 98°C for 2 min, hybridized at 68°C for 5 hrs, and digested by a double-strand specific DNA nuclease (Duplex-specific Nuclease, Crab, recombinant, Solution; Wako Pure Chemical Industries, Ltd, Osaka, Japan). The normalized cDNA was subsequently amplified by PCR with the following thermal condition: preheat at 98°C at 1 min, 10 cycles of consecutive 98°C for 10 sec, 55°C at 5 sec and 72°C for 5 min, and final extension at 72°C for 5 min. Finally the PCR products were digested with *GsuI* for removing the adaptors and poly (A) sequences.

The normalized cDNA library of each termite species was sequenced in a full plate of a Roche 454 FLX Titanium instrument. Sequence assembling was performed using a GS de novo Assembler (Newbler) v2.5.3 (Roche Applied Science, Indianapolis, IN, USA) with cDNA mode (see 454 Sequencing System Software Manual, v2.5.3, Part C, for a detailed description). Newbler explicitly accounts for splice variants in its cDNA mode operation and constructs isotigs. Briefly, Newbler can identify branching structures in multiple alignments of overlapping reads and divides the alignments into multiple contigs within which have no branching conflicts. Then the contigs are grouped into the same isogroups and

used for constructing alternative connections among them with a network or graph-based approach. Any contigs or isotigs that share any read overlaps are grouped into the same isogroup. Broadly contigs can be considered as exons, although they may contain untranslated regions (UTRs) and introns, isotigs as splice variants, and isogroups as groups of splice variants that are generated from single genes. All of the isotigs and singletons obtained by the assembling, except for short singletons (<100 bp), were used for the further analyses.

The preparation of normalized cDNA libraries, 454 pyrosequencing and sequence assembling were performed as a custom service by Genaris, Inc (Yokohama, Kanagawa, Japan).

Functional annotation with non-redundant database, Gene Ontology, and Pfam database

To annotate isotigs and singletons, we carried out sequence similarity searches using BLASTX algorithm (version 2.2.26) [69] against the non-redundant (nr) database (downloaded on 30 Oct 2012) of GenBank. We set E-value threshold of $\leq 1e-4$ and bit score threshold of ≥ 40 for BLAST hit. Based on the BLASTX top hit genes, Gene Ontology (GO) term IDs were assigned to the isotigs and singletons by using ID mapping data of UniProt (28 Nov 2012 released version, downloaded from ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/idmapping/idmapping_selected.tab.gz) and a histogram of GO terms was generated with WEGO [70].

Protein motifs and domains were identified for all of the predicted gene models (see "Coding sequence prediction" of "Methods") by running InterProScan (version 5) [34] with default parameters using known domains from Pfam (Release 26.0).

Species distribution of BLASTX top hit genes

Based on the results of BLASTX with the nr database, we surveyed species distributions of the top hit genes. For this analysis, we used only one isotig selected randomly from a single isogroup when there were multiple isotigs in a single isogroup, because isotigs from the same isogroups have similar sequences and almost always had the same top-hit genes in BLASTX searches, and thus are redundant for the analysis of species distribution.

Coding sequence prediction

The coding sequences and their corresponding amino acid sequences were extracted from the isotigs and singletons by using OrfPredictor ([41]; <http://proteomics.yzu.edu/tools/OrfPredictor.html>).

Orthologous gene determination

InParanoid [42] was used for identification of orthologous gene pairs of *H. sjostedti* and *N. takasagoensis*, *H. sjostedti* and *R. speratus*, and *N. takasagoensis* and *R. speratus*. Then MultiParanoid [43] was used to merge them into multiple species orthologous groups. InParanoid and MultiParanoid were executed with default parameters.

Normalized CpG content (CpG O/E)

For calculation of normalized CpG content (CpG O/E), we used the coding sequences obtained by OrfPredictor. Because our EST libraries contained genes derived from organisms other than the termites, such as intestinal protists, we used only the coding sequences originated from the isotigs and singletons whose top hits of the BLASTX search with the nr database were genes derived from insect species. In addition, when there were multiple coding sequences originating from the same isogroups, we randomly selected and used only one coding sequence from an isogroup to avoid redundancy of the sequences. Finally, short coding sequences (<300 bp) were discarded for this analysis.

The normalized CpG content is defined and calculated by the following formula:

$$\text{CpG O/E} = P_{\text{CpG}}/P_{\text{C}}P_{\text{G}}$$

where P_{CpG} , P_{C} , and P_{G} are the frequencies of CpG dinucleotides, C nucleotides, and G nucleotides, respectively. We statistically tested whether the frequency distributions of CpG O/E were bimodal rather than unimodal. For the test of bimodality we estimated means, standard deviations and mixture proportions of two normal distributions, and calculated log-likelihoods of frequency distributions under models of uni- and bi-modal distributions with the NOCOM software (Ott 1979). Chi-square tests were then performed for the log-likelihoods with the statistics $G^2 = 2[\ln(L_1) - \ln(L_2)]$ which approximately follows a chi-square distribution with 2 degrees of freedom.

The coding sequences were classified into low- and high-CpG O/E genes by using the intersection of the two normal curves in the CpG O/E frequency distributions as a threshold of the classification. We then made histograms of GO terms for the two classes of the coding sequences. To find enriched GO terms in the two classes, we performed Fisher's exact tests for the rate of annotated genes to the rest of all genes in each GO term in the two classes. False discovery rate was controlled at $q < 0.05$ by the Benjamini & Hochberg method [71] for each of BP, CC and MF categories.

Database search for *dnmt* and *mbd* gene sequences

TBLASTN searches were performed for examining presence of DNA (cytosine-5) methyltransferase genes, *dnmt1*, *dnmt2* and *dnmt3*, and methyl-CpG-binding domain, *mbd*, in the EST libraries. We used DNMT and MBD protein sequences of the following insect species downloaded from GenBank as query sequences for the TBLASTN searches against the EST libraries: *Acyrtosiphon pisum* (DNMT1, XP_003243626; DNMT2, XP_001949338; DNMT3a, XP_003241627; DNMT3b, XP_003240668; MBD, NP_001156167), *Apis mellifera* (DNMT1a, NP_001164522; DNMT1b, XP_001122269; DNMT2, XP_393991; DNMT3, NP_001177350; MBD isoform 1, XP_003250633; MBD isoform 2, XP_003250634; MBD isoform 3, XP_392422), *Nasonia vitripennis* (DNMT1a, NP_001164521; DNMT1b, XP_001600175; DNMT1c, XP_001607336; DNMT2, XP_001602026; DNMT3, XP_001599223; MBD isoform A, NP_001164526; MBD isoform B, NP_001164527), *Pediculus humanus* (DNMT1a, XP_002432160; DNMT1b, XP_002431878; DNMT2,

XP_002432555; MBD, XP_002428735) and *Tribolium castaneum* (DNMT1, XP_001814230; DNMT2, EFA09160; MBD, XP_969537). E value of $\leq 1e-20$ was set as a threshold for significant hit of the TBLASTN searches.

Supporting Information

Figure S1. Caste developmental pathways. (a) *Hodotermopsis sjostedti*, (b) *Reticulitermes speratus*, and (c) *Nasutitermes takasagoensis*. Each arrow indicates a molt. Dotted lines indicate potential molts, which are suggested to occur under natural conditions. It is known that ergatoids are differentiated from workers in *R. speratus*, and from workers or larvae in *N. takasagoensis*, while instars that have the potential to develop into ergatoids have not been identified.

(PDF)

Figure S2. Histograms of normalized contents of dinucleotides in *Hodotermopsis sjostedti*.

(PDF)

Figure S3. Histograms of normalized contents of dinucleotides in *Reticulitermes speratus*.

(PDF)

Figure S4. Histograms of normalized contents of dinucleotides in *Nasutitermes takasagoensis*.

(PDF)

Figure S5. Frequency and percentage of high- and low-CpG genes annotated by Gene Ontology terms in three termite species. The terms in which significant differences in frequencies between high- and low-CpG genes were found are indicated by asterisks.

(PDF)

Table S1. Summary of samples used for cDNA library construction in *Hodotermopsis sjostedti*. Caste, sex, and description of samples, number of individuals, and field colonies from which termite samples originated are shown.

(PDF)

Table S2. Summary of samples used for cDNA library construction in *Reticulitermes speratus*. Caste, sex, and

description of samples, number of individuals, and field colonies from which termite samples originated are shown.

(PDF)

Table S3. Summary of samples used for cDNA library construction in *Nasutitermes takasagoensis*. Caste, sex, and description of samples, number of individuals, and field colonies from which termite samples originated are shown.

(PDF)

Table S4. Summary of protein domain search result in EST libraries of three termite species. The 30 most frequently occurring Pfam domains/families in the isotigs and singletons of the three termite species are shown.

(PDF)

Table S5. Distribution of BLASTX top hits in EST libraries of three termite species.

(PDF)

Acknowledgements

We are grateful to M. Suzuki [National Institute for Basic Biology (NIBB)] for her useful comments on DNA methylation. We thank S. Nakamura, M. Yoshimura, H. Yaguchi, Y. Hashimoto, F. Nakayama (University of Toyama), M.K. Hojo, Y. Ishikawa, S. Ohno, T. Yodoi, and K. Tanabe (Hokkaido University) for their help in collecting termites and extracting RNA. Data Integration and Analysis Facility of NIBB is acknowledged for providing computer resources. This study was carried out as a part of Collaborative Experiments Using Next Generation DNA Sequencers of NIBB (No. 11-723, 12-701).

Author Contributions

Conceived and designed the experiments: YH SS KM TM. Performed the experiments: YH SS DW KT RS KS MH. Analyzed the data: YH SS DW KT RS KS TB NL MH KM. Contributed reagents/materials/analysis tools: YH SS DW KT RS KS MH. Wrote the manuscript: YH SS DW KT RS KS TB NL MH KM TM.

References

- Wilson EO (1971) The insect societies. Cambridge, MA: Belknap Press. 548pp.
- Oster GF, Wilson EO (1978) Caste and ecology in the social insects. Princeton, NJ: Princeton University Press. 352pp.
- Noirot C (1991) Caste differentiation in Isoptera - basic features, role of pheromones. Ethol Ecol Evol: 3-7.
- Miura T, Scharf ME (2011) Molecular basis underlying caste differentiation in termites. In: D Bignell Y Roisin N Lo. Biology of Termites: a Modern Synthesis. Dordrecht, Netherlands: Springer Verlag, pp. 211-253.
- Kambhampati S, Eggleton P (2000) Taxonomy and phylogeny of termites. In: T Abe DE Bignell M Higashi. Termites: evolution, sociality, symbioses, ecology. Dordrecht, Netherlands: Kluwer Publishing House Academic Publishers. pp. 1-23.
- Roisin Y (2000) Diversity and evolution of caste patterns. In: T Abe DE Bignell M Higashi. Termites: evolution, sociality, symbioses, ecology. Dordrecht, Netherlands: Kluwer Publishing House Academic Publishers. pp. 95-119.
- Jongeneel CV (2000) Searching the expressed sequence tag (EST) databases: panning for genes. Brief Bioinform 1: 76-92. doi: 10.1093/bib/1.1.76. PubMed: 11466975.
- Rudd S (2003) Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci 8: 321-329. doi:10.1016/S1360-1385(03)00131-6. PubMed: 12878016.
- Dong Q, Kroiss L, Oakley FD, Wang BB, Brendel V (2005) Comparative EST analyses in plant systems. Methods Enzymol 395: 400-418. doi:10.1016/S0076-6879(05)95022-2. PubMed: 15984049.

10. Nagaraj SH, Gasser RB, Ranganathan S (2007) A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief Bioinform* 8: 6-21. PubMed: 16772268.
11. Husseneder C, McGregor C, Lang RP, Collier R, Delatte J (2012) Transcriptome profiling of female alates and egg-laying queens of the Formosan subterranean termite. *Comp Biochem Physiol D Genomics Proteomics* 7: 14-27. doi:10.1016/j.cbd.2011.10.002. PubMed: 22079412.
12. Shcheglov AS, Zhulidov PA, Bogdanova EA, Shagin DA (2007) Normalization of cDNA Libraries. In: AA Buzdin SA Lukyanov. *Nucleic Acids Hybridization: Modern Applications*. Dordrecht, Netherlands: Springer Verlag, pp. 97-124.
13. Soares MB, de Fatima Bonaldo M, Hackett JD, Bhattacharya D (2009) Expressed sequence tags: normalization and subtraction of cDNA libraries expressed sequence tags. *Methods Mol Biol* 533: 109-122. doi:10.1007/978-1-60327-136-3_6. PubMed: 19277560.
14. Zhang D, Lax AR, Henriessat B, Coutinho P, Katiya N et al. (2012) Carbohydrate-active enzymes revealed in *Coptotermes formosanus* (Isoptera: Rhinotermitidae) transcriptome. *Insect Mol Biol* 21: 235-245. doi:10.1111/j.1365-2583.2011.01130.x. PubMed: 22243654.
15. Steller MM, Kambhampati S, Caragea D (2010) Comparative analysis of expressed sequence tags from three castes and two life stages of the termite *Reticulitermes flavipes*. *BMC Genomics* 11: 463. doi:10.1186/1471-2164-11-463. PubMed: 20691076.
16. Hojo M, Maekawa K, Saitoh S, Shigenobu S, Miura T et al. (2012) Exploration and characterization of genes involved in the synthesis of diterpene defence secretion in nasute termite soldiers. *Insect Mol Biol* 21: 545-557. doi:10.1111/j.1365-2583.2012.01162.x. PubMed: 22984844.
17. Huang Q, Sun P, Zhou X, Lei C (2012) Characterization of head transcriptome and analysis of gene expression involved in caste differentiation and aggression in *Odontotermes formosanus* (Shiraki). *PLOS ONE* 7: e50383. doi:10.1371/journal.pone.0050383. PubMed: 23209730.
18. Inward DJ, Vogler AP, Eggleton P (2007) A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology. *Mol Phylogenet Evol* 44: 953-967. doi:10.1016/j.ympev.2007.05.014. PubMed: 17625919.
19. Ogino K, Hirono Y, Matsumoto T, Ishikawa H (1993) Juvenile-hormone analog, S-31183, causes a high-level induction of presoldier differentiation in the Japanese damp-wood termite. *Zool Sci* 10: 361-366.
20. Hattori A, Sugime Y, Sasa C, Miyakawa H, Ishikawa Y et al. (2013) Soldier morphogenesis in the damp-wood termite is regulated by the insulin signaling pathway. *J Exp Zool B* 320B: 295-306. PubMed: 23703784.
21. Tsuchiya M, Watanabe D, Maekawa K (2008) Effect on mandibular length of juvenile hormones and regulation of soldier differentiation in the termite *Reticulitermes speratus* (Isoptera: Rhinotermitidae). *Appl Entomol Zool* 43: 307-314. doi:10.1303/aez.2008.307.
22. Nambu Y, Tanaka H, Enoki A, Itakura S (2010) RNA interference in the termite *Reticulitermes speratus*: silencing of the hexamerin gene using a single 21 nucleotide small interfering RNA-promoted differentiation of nymph to nymphoid. *Sociobiology* 55: 527-546.
23. Toga K, Hojo M, Miura T, Maekawa K (2012) Expression and function of a limb-patterning gene *Distal-less* in the soldier-specific morphogenesis in the nasute termite *Nasutitermes takasagoensis*. *Evol Dev* 14: 286-295. PubMed: 23017076.
24. Roberts SB, Gavery MR (2012) Is there a relationship between DNA methylation and phenotypic plasticity in invertebrates? *Front. Physiol* (Bethesda Md.) 2: 116.
25. Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenetics. *Nat Rev Genet* 9: 465-476. doi:10.1038/nrg2341. PubMed: 18463664.
26. Yi SV, Goodisman MA (2009) Computational approaches for understanding the evolution of DNA methylation in animals. *Epigenetics* 4: 551-556. doi:10.4161/epi.4.8.10345. PubMed: 20009525.
27. Suzuki MM, Kerr ARW, De Sousa D, Bird A (2007) CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res* 17: 625-631. doi:10.1101/gr.6163007. PubMed: 17420183.
28. Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328: 916-919. doi:10.1126/science.1186366. PubMed: 20395474.
29. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W (1978) Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* 274: 775-780. doi:10.1038/274775a0. PubMed: 355893.
30. Bird AP (1980) DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* 8: 1499-1504. doi:10.1093/nar/8.7.1499. PubMed: 6253938.
31. Schorderet DF, Gartler SM (1992) Analysis of CpG suppression in methylated and nonmethylated species. *Proc Natl Acad Sci U S A* 89: 957-961. doi:10.1073/pnas.89.3.957. PubMed: 1736311.
32. Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23: 1061-1067. doi:10.1093/bioinformatics/btm071. PubMed: 17332020.
33. Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV (2013) OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Res* 41: D358-D365. doi:10.1093/nar/gks1116. PubMed: 23180791.
34. Zdobnov EM, Apweiler R (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847-848. doi:10.1093/bioinformatics/17.9.847. PubMed: 11590104.
35. Nijhout HF, Wheeler DE (1982) Juvenile hormone and the physiological basis of insect polymorphisms. *Q Rev Biol* 57: 109-133. doi:10.1086/412671.
36. Nijhout HF (1994) *Insect hormones*. Princeton, NJ: Princeton University Press. 280pp.
37. the International Aphid Genomics Consortium (2010) Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLOS Biol* 8: e1000313. PubMed: 20186266.
38. Inoue T, Kitade O, Yoshimura T, Yamaoka I (2000) Symbiotic associations with protists. In: T Abe DE Bignell M Higashi. *Termites: evolution, sociality, symbioses, ecology*. Dordrecht, Netherlands: Kluwer Publishing House Academic Publishers. pp. 275-288.
39. Ohkuma M, Brune A (2011) Diversity, structure, and evolution of the termite gut microbial community. In: DE Bignell Y Roisin N Lo. *Biology of Termites: A Modern Synthesis*. Dordrecht, Netherlands: Springer Verlag. pp. 413-438.
40. Cleveland LR (1923) Symbiosis between Termites and their intestinal protozoa. *Proc Natl Acad Sci U S A* 9: 424-428. doi:10.1073/pnas.9.12.424. PubMed: 16586922.
41. Min XJ, Butler G, Storms R, Tsang A (2005) OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res* 33: W677-W680. doi:10.1093/nar/gki394. PubMed: 15980561.
42. Remm M, Storm CE, Sonnhammer EL (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* 314: 1041-1052. doi:10.1006/jmbi.2000.5197. PubMed: 11743721.
43. Alexeyenko A, Tamas I, Liu G, Sonnhammer EL (2006) Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics* 22: e9-15. doi:10.1093/bioinformatics/btl213. PubMed: 16873526.
44. Nalepa CA, Bandi C (2000) Characterizing the ancestors: paedomorphosis and termite evolution. In: T Abe DE Bignell M Higashi. *Termites: evolution, sociality, symbioses, ecology*. Dordrecht, Netherlands: Kluwer Publishing House Academic Publishers. pp. 53-76.
45. Wilson EO (1965) Chemical communication in the social insects. *Science* 149: 1064-1071. doi:10.1126/science.149.3688.1064. PubMed: 17737837.
46. Hölldobler B, Wilson EO (1990) *The ants*. Cambridge, MA: Belknap Press. 732pp.
47. Matsuura K (2012) Multifunctional queen pheromone and maintenance of reproductive harmony in termite colonies. *J Chem Ecol* 38: 746-754. doi:10.1007/s10886-012-0137-3. PubMed: 22623152.
48. Robertson KD (2002) DNA methylation and chromatin - unraveling the tangled web. *Oncogene* 21: 5361-5379. doi:10.1038/sj.onc.1205609. PubMed: 12154399.
49. Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL et al. (2006) Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science* 311: 395-398. doi:10.1126/science.1120976. PubMed: 16424344.
50. Jurkowski TP, Meusburger M, Phalke S, Helm M, Nellen W et al. (2008) Human DNMT2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism. *RNA* 14: 1663-1670. doi:10.1261/rna.970408. PubMed: 18567810.
51. Okano M, Xie S, Li E (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* 19: 219-220. doi:10.1038/890. PubMed: 9662389.
52. Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 31: 89-97. doi:10.1016/j.tibs.2005.12.008. PubMed: 16403636.
53. Lo N, Li B, Ujvari B (2012) DNA methylation in the termite *Coptotermes lacteus*. *Insectes Soc* 59: 257-261. doi:10.1007/s00040-011-0213-7.
54. Glastad KM, Hunt BG, Goodisman MA (2013) Evidence of a conserved functional role for DNA methylation in termites. *Insect Mol Biol* 22: 143-154. doi:10.1111/imb.12010. PubMed: 23278917.
55. Ott J (1979) Detection of rare major genes in lipid levels. *Hum Genet* 51: 79-91. doi:10.1007/BF00278296. PubMed: 500096.

56. Moczek AP, Snell-Rood EC (2008) The basis of bee-ing different: the role of gene silencing in plasticity. *Evol Dev* 10: 511-513. doi:10.1111/j.1525-142X.2008.00264.x. PubMed: 18803767.
57. Glstad KM, Hunt BG, Yi SV, Goodisman MA (2011) DNA methylation in insects: on the brink of the epigenomic era. *Insect Mol Biol* 20: 553-565. doi:10.1111/j.1365-2583.2011.01092.x. PubMed: 21699596.
58. Kucharski R, Maleszka J, Foret S, Maleszka R (2008) Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319: 1827-1830. doi:10.1126/science.1153069. PubMed: 18339900.
59. Toga K, Hojo M, Miura T, Maekawa K (2009) Presoldier induction by a juvenile hormone analog in the nasute termite *Nasutitermes takasagoensis* (Isoptera: Termitidae). *Zool Sci* 26: 382-388. doi:10.2108/zsj.26.382. PubMed: 19583496.
60. Watanabe H, Noda H (1991) Small-scale rearing of a subterranean termite, *Reticulitermes speratus* (Isoptera, Rhinotermitidae). *Appl Entomol Zool* 26: 418-420.
61. Miyata H, Furuichi H, Kitade O (2004) Patterns of neotenic differentiation in a subterranean termite, *Reticulitermes speratus* (Isoptera : Rhinotermitidae). *Entomol Sci* 7: 309-314. doi:10.1111/j.1479-8298.2004.00078.x.
62. Saiki R, Maekawa K (2011) Imaginal organ development and vitellogenin gene expression changes during the differentiation of nymphoids of the termite *Reticulitermes speratus*. *Sociobiology* 58: 499-511.
63. Miura T, Hirono Y, Machida M, Kitade O, Matsumoto T (2000) Caste developmental system of the Japanese damp-wood termite *Hodotermopsis japonica* (Isoptera: Termopsidae). *Ecol Res* 15: 83-92. doi:10.1046/j.1440-1703.2000.00320.x.
64. Miura T, Koshikawa S, Machida M, Matsumoto T (2004) Comparative studies on alate wing formation in two related species of rotten-wood termites: *Hodotermopsis sjostedti* and *Zootermopsis nevadensis* (Isoptera, Termopsidae). *Insectes Soc* 51: 247-252.
65. Kawamura T (1982) Development and caste differentiation in *Reticulitermes speratus* (Kolbe). *Shiroari* 49: 44-52. (in Japanese).
66. Takematsu Y (1992) Biometrical study on the development of the castes in *Reticulitermes speratus* (Isoptera, Rhinotermitidae). *Jpn J Entomol* 60: 67-76.
67. Hojo M, Koshikawa S, Matsumoto T, Miura T (2004) Developmental pathways and plasticity of neuter castes in *Nasutitermes takasagoensis* (Isoptera: Termitidae). *Sociobiology* 44: 433-441.
68. Hayashi Y, Kitade O, Kojima J (2003) Parthenogenetic reproduction in neotenic of the subterranean termite *Reticulitermes speratus* (Isoptera: Rhinotermitidae). *Entomol Sci* 6: 253-257. doi:10.1046/j.1343-8786.2003.00030.x.
69. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402. doi:10.1093/nar/25.17.3389. PubMed: 9254694.
70. Ye J, Fang L, Zheng H, Zhang Y, Chen J et al. (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res* 34: W293-W297. doi:10.1093/nar/gkl031. PubMed: 16845012.
71. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 57: 289-300.