Genomic and transcriptomic analyses of the subterranean termite *Reticulitermes speratus*: Gene duplication facilitates social evolution

Shuji Shigenobu[a,b,1,2], Yoshinobu Hayashi[c,1], Dai Watanabe[d,e], Gaku Tokuda[f], Masaru Y. Hojo[f,g], Kouhei Toga[e,h], Ryota Saiki[e], Hajime Yaguchi[e,i], Yudai Masuoka[e,j], Ryutaro Suzuki[e,k], Shogo Suzuki[e], Moe Kimura[l], Masatoshi Matsunami[d,m], Yasuhiro Sugime[d], Kohei Oguchi[d,k,n], Teruyuki Niimi[b,o], Hiroki Gotoh[p,q], Masaru K. Hojo[i], Satoshi Miyazaki[r], Atsushi Toyoda[s], Toru Miura[d,n,2], and Kiyoto Maekawa[t,2]

[a]NIBB Research Core Facilities, National Institute for Basic Biology, Okazaki 444-8585 Japan; [b]Department of Basic Biology, School of Life Science, The Graduate University for Advanced Studies, SOKENDAI, Okazaki, Aichi 444-8585, Japan; [c]Department of Biology, Keio University, Hiyoshi, Yokohama 223-8521, Japan; [d]Faculty of Environmental Earth Science, Hokkaido University, Sapporo, Hokkaido 060-0810, Japan; [e]Graduate School of Science and Engineering, University of Toyama, Toyama 930-8555, Japan; [f]Tropical Biosphere Research Center, Center of Molecular Biosciences, University of the Ryukyus, Nishihara, Okinawa 903-0213, Japan; [g]Global Education Institute, University of the Ryukyus, Nishihara, Okinawa 903-0213, Japan; [h]Department of Biosciences, College of Humanities and Sciences, Nihon University, Setagaya-ku, Tokyo 156-8550, Japan; [i]Department of Bioscience, School of Science and Technology, Kwansei Gakuin University, Sanda, Hyogo 669-1337, Japan; [j]Institute of Agrobiological Sciences, National Agriculture and Food Research Organization, Tsukuba 305-8634, Japan; [k]Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology, Tsukuba 305-8566, Japan; [l]School of Science, University of Toyama, Toyama 930-8555, Japan; [m]Graduate School of Medicine, University of the Ryukyus, Nishihara, Okinawa 903-0125, Japan; [n]Misaki Marine Biological Station, School of Science, The University of Tokyo, Miura, Kanagawa 238-0225, Japan; [o]Division of Evolutionary Developmental Biology, National Institute for Basic Biology, Okazaki 444-8585 Japan; [p]Ecological Genetics Laboratory, Department of Genomics and Evolutionary Biology, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan; [q]Department of Biological Sciences, Faculty of Science, Shizuoka University, Shizuoka, Shizuoka 422-8529, Japan; [r]Graduate School of Agriculture, Tamagawa University, Machida, Tokyo 194-8610, Japan; [s]Advanced Genomics Center, National Institute of Genetics, Mishima, Shizuoka 411-8540 Japan; and [t]Faculty of Science, Academic Assembly, University of Toyama, Gofuku, Toyama 930-8555, Japan

Edited by Raghavendra Gadagkar, Centre for Ecological Sciences, Indian Institute of Science, Bangalore, India; received June 26, 2021; accepted December 1, 2021

Termites are model social organisms characterized by a polyphenic caste system. Subterranean termites (Rhinotermitidae) are ecologically and economically important species, including acting as destructive pests. Rhinotermitidae occupies an important evolutionary position within the clade representing a transitional taxon between the higher (Termitidae) and lower (other families) termites. Here, we report the genome, transcriptome, and methylome of the Japanese subterranean termite *Reticulitermes speratus*. Our analyses highlight the significance of gene duplication in social evolution in this termite. Gene duplication associated with caste-biased gene expression was prevalent in the *R. speratus* genome. The duplicated genes comprised diverse categories related to social functions, including lipocalins (chemical communication), cellulases (wood digestion and social interaction), lysozymes (social immunity), geranylgeranyl diphosphate synthase (social defense), and a novel class of termite lineage–specific genes with unknown functions. Paralogous genes were often observed in tandem in the genome, but their expression patterns were highly variable, exhibiting caste biases. Some of the assayed duplicated genes were expressed in caste-specific organs, such as the accessory glands of the queen ovary and the frontal glands of soldier heads. We propose that gene duplication facilitates social evolution through regulatory diversification, leading to caste-biased expression and subfunctionalization and/or neofunctionalization conferring caste-specialized functions.

social insect | termite | gene duplication | caste system

The evolution of eusociality (i.e., animal societies defined by the reproductive division of labor, cooperative brood care, and multiple overlapping generations) represents an important step in animal evolution that increased the level of biological complexity (1, 2). Eusocial insects such as bees, wasps, ants, and termites show sophisticated systems based on the division of labor among castes, which is one of the pinnacles of eusocial evolution (3). Recent advances in molecular biological technologies and omics studies have revealed many molecular mechanisms underlying eusociality and have led to the establishment of a new field of study known as "sociogenomics" (4). The genomes of major eusocial hymenopteran lineages (i.e., ants, bees, and wasps) have been sequenced, and the differences in gene expression and DNA methylation among castes have been explored (5–14). These sociogenomic studies in hymenopterans have revealed some of the genetic basis of social evolution, including the co-option of genetic toolkits of conserved genes,

Significance

Gene duplication is a major source of evolutionary innovation and is associated with the increases in biological complexity and adaptive radiation. Termites are model social organisms characterized by a sophisticated caste system. We analyzed the genome of the Japanese subterranean termite, an ecologically and economically important insect acting as a destructive pest. The analyses revealed the significance of gene duplication in social evolution. Gene duplication associated with caste-biased gene expression was prevalent in the termite genome. Many of the duplicated genes were related to social functions, such as chemical communication, social immunity, and defense, and they were often expressed in caste-specific organs. We propose that gene duplication facilitates social evolution through regulatory diversification leading to caste-biased expression and functional specialization.

Author contributions: S. Shigenobu, Y.H., T.M., and K.M. designed research; S. Shigenobu, Y.H., D.W., G.T., M.Y.H., K.T., R. Saiki, H.Y., Y.M., R. Suzuki, S. Suzuki, M.K., M.M., Y.S., K.O., T.N., H.G., M.K.H., S.M., A.T., T.M., and K.M. performed research; S. Shigenobu, Y.H., and K.M. contributed new reagents/analytic tools; S. Shigenobu, Y.H., D.W., G.T., M.Y.H., K.T., R. Saiki, H.Y., Y.M., R. Suzuki, S. Suzuki, M.K., M.M., Y.S., K.O., T.N., H.G., M.K.H., S.M., A.T., T.M., and K.M. analyzed data; and S. Shigenobu, Y.H., D.W., G.T., M.Y.H., K.T., R. Saiki, H.Y., R. Suzuki, S. Suzuki, M.K., M.M., Y.S., K.O., T.N., H.G., M.K.H., S.M., A.T., T.M., and K.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]S.S. and Y.H. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: shige@nibb.ac.jp, miu@mmbs.s.u-tokyo.ac.jp, or kmaekawa@sci.u-toyama.ac.jp.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2110361119/-/DCSupplemental.

Published January 18, 2022.

changes in protein-coding genes, cis-regulatory evolution leading to genetic network reconstruction, epigenetic modifications, and taxonomically restricted genes (TRGs) (15, 16).

Termites (Isoptera) are another representative insect lineage exhibiting highly sophisticated eusociality and a wide range of social complexities (17). Termites are hemimetabolous, diploid insects that are phylogenetically distant from hymenopterans with holometaboly and haplodiploidy. Termite societies are characterized by reproductives of both sexes, workers, and soldiers. In the termite lineage, eusociality is thought to have evolved once, although the levels of social complexity features, such as colony size, feeding habitat, symbiosis with microorganisms, and caste developmental pathways, have diverged among termite species (18). These characteristics are especially different between the two major termite sublineages: the early-branching families (referred to as "lower" termites) and the most apical family Termitidae ("higher" termites) (Fig. 1*A*). Based on the whole-genome sequences of a few termite species reported to date, the commonality and diversity of the genetic repertoires between Isoptera and Hymenoptera or between termite lineages and their solitary outgroup (i.e., cockroach) have been investigated (20–23). Additionally, in *Zootermopsis nevadensis*, clear differences in gene expression levels among castes (20) and in DNA methylation between alates and workers (24) have been detected.

Evolutionary novelties are often brought about by gene duplications (25) (reviewed in ref. 26). Gene duplication allows the subsequent divergent evolution of the resultant gene copies, enabling evolutionary innovations in protein functions and/or expression patterns (27–29). The transition to eusociality in Hymenoptera has been associated with gene family expansion, including the expansion of odorant receptor (30–34), insulin signaling (35), and vitellogenin genes (36, 37). In two termites with available genome information, *Z. nevadensis* and *Cryptotermes secundus*, copy numbers were found to be expanded for genes such as ionotropic receptor (20, 21), vitellogenin (20, 38), insulin receptor (39), and juvenile hormone biosynthesis genes (40). Despite the accumulating examples of duplications of some categories of genes in social insects, their genome-wide impact and the evolutionary significance remain unclear.

Among the more than 2,900 extant species of termites (Isoptera) (41), subterranean termites (Rhinotermitidae), especially those belonging to two genera, *Reticulitermes* and *Coptotermes*, occupy an important evolutionary position (Fig. 1*A*). Recent phylogenetic studies showed that Rhinotermitidae is paraphyletic, and a clade including *Reticulitermes*, *Coptotermes*, and *Heterotermes* was shown to be sister to Termitidae (42, 43). In particular, *Reticulitermes* exhibits intermediate social complexity between higher (Termitidae) and lower (all the other families) termites (44), although it belongs to the latter group [e.g., this genus displays primitive feeding ecology and gut symbiont features, a relatively complex colony structure and a caste development mode known as the bifurcated pathway (Fig. 1*B*)]. Moreover, *Reticulitermes* is the most common termite group in palearctic (45) and nearctic (46) regions and includes major pests causing serious damage to human-made wooden structures (47). For these reasons, members of this genus are probably among the most studied termites (17). Nevertheless, despite their evolutionary, ecological, and economic relevance, subterranean termites remain an understudied group in terms of both genetics and genomics.

In this study, we targeted the Japanese subterranean termite *Reticulitermes speratus*. We conducted whole-genome sequencing, caste-specific RNA sequencing (RNA-seq) analysis and whole-genome bisulfite sequencing (BS-seq) of *R. speratus* to understand the genomic, transcriptomic, and epigenetic bases of the social life of this termite species. We also compared the omics data of *R. speratus* with those of sequenced higher and lower termites (Fig. 1*A*). Our integrative analyses revealed that gene duplications are often associated with caste-biased gene expression and caste-specific functions, which highlights the significant role of gene duplication in eusocial evolution in the termite lineage.

Results and Discussion

Genomic Features of *R. speratus*. Genome sequencing of *R. speratus* was performed with genomic DNA isolated from female secondary reproductives (nymphoids) (Fig. 1*B*). *R. speratus* nymphoids are almost exclusively produced parthenogenetically by automixis with terminal fusion in primary queens, such that the genome should be homozygous at most loci (48), which provides an advantage in genome assembly. We generated a total of 86 Gb of Illumina HiSeq sequence data and performed de novo assembly into 5,817 scaffolds with an N50 of 1.97 Mb and total size of 881 Mb (*SI Appendix*, Table S1), covering 88% of the genome based on the genome size (1.0 Gb) estimated by flow cytometry (49). The assembled *R. speratus* genome shows high coverage of coding regions, capturing 99.2% (98.5% complete; 0.7% fragmented) of 1,367 Insecta benchmarking universal single-copy orthologs (50) (*SI Appendix*, Table S1). The *R. speratus* genome is rich in repetitive elements, which make up 40.4% of the genome. The high accumulation of repetitive elements may be relevant to the evolution of eusociality, as proposed for termites and snapping shrimp (22, 51). A total of 15,591 protein-coding genes were predicted by combining the reference-guided assembly of RNA-seq reads (36 libraries derived from different castes, sexes, and body parts; see *Transcriptome Differentiation among Castes* for details) and homology-based gene prediction followed by manual curation of gene families of interest (Fig. 1*C*). Whole-genome BS-seq revealed extensive gene body methylation of the *R. speratus* genome, accounting for 8.8% of methylated cytosines in the CG context (*SI Appendix*, Fig. S1). The genome-wide DNA methylation landscape was similar to that of the dampwood termite *Z. nevadensis* (12%) (24). These omics data and a genome browser are available at http://www.termite.nibb.info/retsp/.

We compared the *R. speratus* gene repertoire with those of 88 other arthropods, including the two termites *Z. nevadensis* and *Macrotermes natalensis* (20, 52). Ortholog analysis showed that 12,032 (82.9%) of the 15,591 genes in *R. speratus* genes were shared with other arthropods, while 1,773 were taxonomically restricted (TRGs) to Isoptera, among which 430 were shared with the other two termites and 1,343 were unique to *R. speratus* (Fig. 1*C*). Whole-genome comparisons with two sequenced termites, *M. natalensis* and *Z. nevadensis,* showed a high degree of synteny conservation (Fig. 1*D*). We identified 2,799 syntenic blocks (N50: 858.4 kb) shared with *M. natalensis*, covering 95.1% of the *R. speratus* genome, in which 560.4 Mb of nucleotides were aligned, while 3,650 syntenic blocks (N50: 591.1 kb) were shared with *Z. nevadensis*, covering 72.1% of the *R. speratus* genome, 116.7 Mb of which were aligned. Only a few cases of large genomic rearrangements were found between the termite genomes, at least at the contiguity level of the current assemblies, suggesting overall conservation of genome architecture in the lineage of termites over 135 million y (Fig. 1*A*). Interestingly, despite such high conservation of macrosynteny, interruptions or breaks in local synteny were observed and were often associated with tandem gene duplications. For example, when we examined regions containing large tandem gene duplications (>5-gene tandem duplications), the synteny between the *R. speratus* and *M. natalensis* genomes was found to be interrupted in 10 of 21 regions (examples shown in *SI Appendix*, Fig. S2).

Transcriptome Differentiation among Castes. Distinct castes arise from the same genome via a phenomenon called caste polyphenism, which is a distinctive hallmark of social insects (53, 54). To

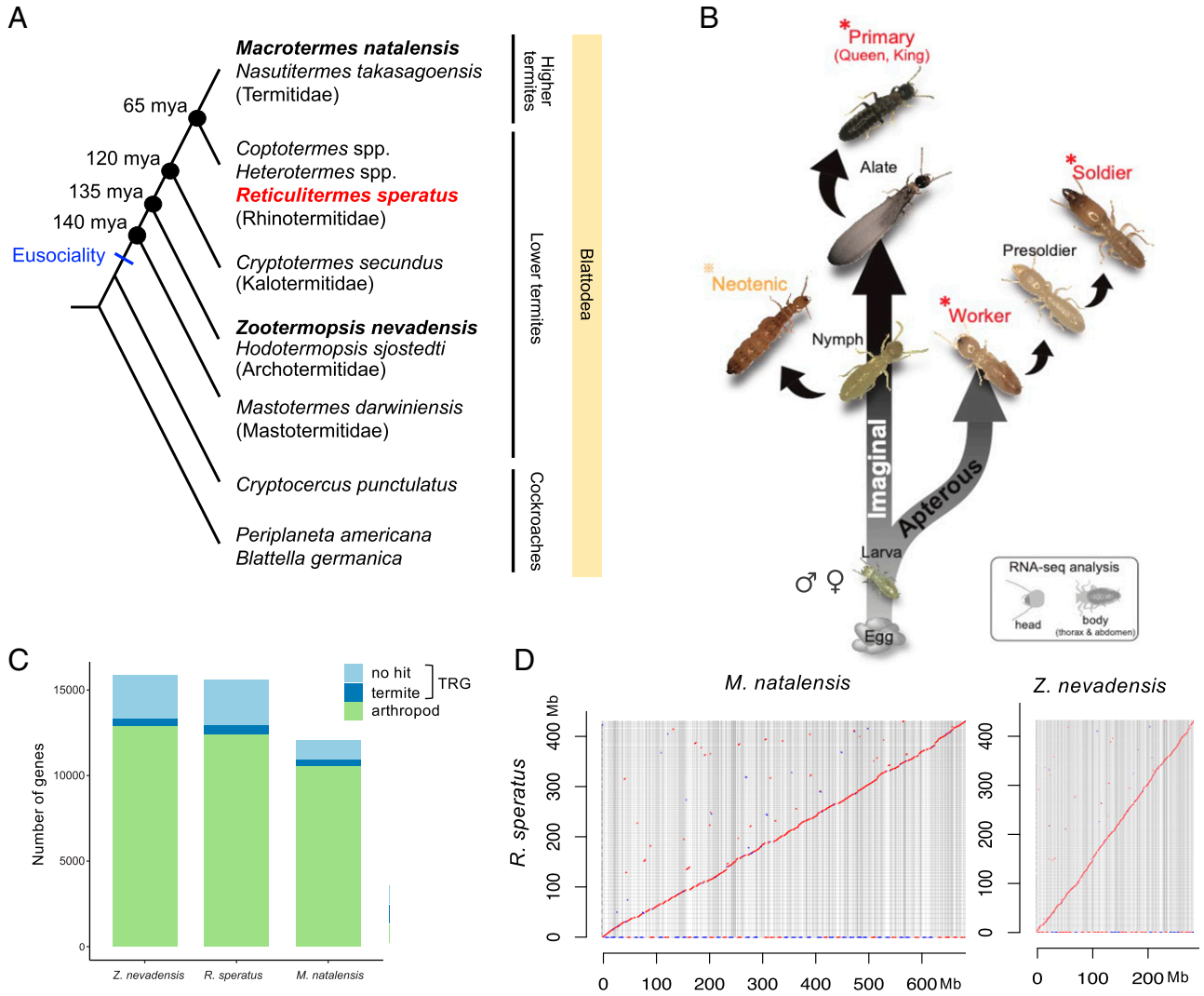

Fig. 1. Phylogenetic position of *R. speratus* in Blattodea, its developmental pathway, and the evolution of its gene repertoire and genome structure. (*A*) Phylogenetic tree of termites and cockroaches. Estimated divergence dates (mya: million years ago) are based on Bucek et al. (19). *R. speratus* is marked in red, and the two termites mainly used for comparison in this study are marked with bold characters. (*B*) Developmental pathway of *R. speratus*. Both sexes have the same developmental pathway. There are two larval stages before molting into a nymph (with wing buds) or worker (no wing buds). There are six imaginal stages, and the sixth-stage nymphs molt into alates, which are primary reproductives (queen and king). Secondary reproductives (neotenics or nymphoids) differentiate from third- to sixth-stage nymphs. In the apterous line, there are at least five stages of workers. Some workers in the colony molt into presoldiers and soldiers. The female neotenics used for genome sequencing and the three castes used for RNA-seq are marked with asterisks. (*C*) Gene repertoire of *R. speratus*, *Z. nevadensis*, and *M. natalensis* categorized by orthology. For each species, genes were compared to those of other 88 arthropods and grouped into three classes: orthologs shared with other arthropods (labeled "arthropod"), orthologs shared with either of the other termites but with no orthologs in other arthropods (labeled "termite"), and orphan genes unique to each species (labeled "no hit"). Genes classified as "termite" and "no hit" are designated as TRGs. (*D*) High conservation of synteny between termite genomes revealed by dot plots generated by comparing *R. speratus* with *Z. nevadensis* and *M. natalensis*. Scaffolds longer than 2.0 Mb in the *R. speratus* assembly are used for plotting. Forward alignments are plotted in red, and reverse alignments are plotted in blue.

elucidate caste-biased gene expression to understand the mechanism underlying caste-specific phenotypes, we compared the transcriptomes of three castes (primary reproductives, workers, and soldiers) of *R. speratus* (Fig. 1*B* and *SI Appendix*, Table S2). We sequenced 36 RNA-seq libraries, representing three biological replicates of both sexes and two body parts ("head" and "thorax + abdomen") for each of the three castes.

The results clearly showed that the termite castes were distinctively differentiated at the gene expression level. The multidimensional scaling plot depicted the three castes as clearly distinct transcriptomic clusters for both the head and thorax + abdomen transcriptomes (Fig. 2*A*). However, little sexual difference was detected within each caste, although reproductives showed substantial transcriptomic differences in thorax + abdomen samples between queens and kings, probably due to the differences in the reproductive organs (Fig. 2*A*). Using a generalized linear model (GLM) with caste and sex as explanatory variables, we identified 1,579 and 2,076 genes that were differentially expressed among castes in the head and thorax + abdomen samples, respectively, according to the criterion of a false discovery rate (FDR)-corrected $P < 0.01$, while we identified only 6 and 79 genes that were differentially expressed between the sexes in the head and thorax + abdomen samples, respectively, with the same criteria. We focused on the genes that were differentially expressed among castes (caste-DEGs) and further classified them into three categories of caste-biased genes (i.e., reproductive-, worker-, and soldier-biased genes), according to the criterion of >twofold higher expression than that in the other two castes (Fig. 2*B* and

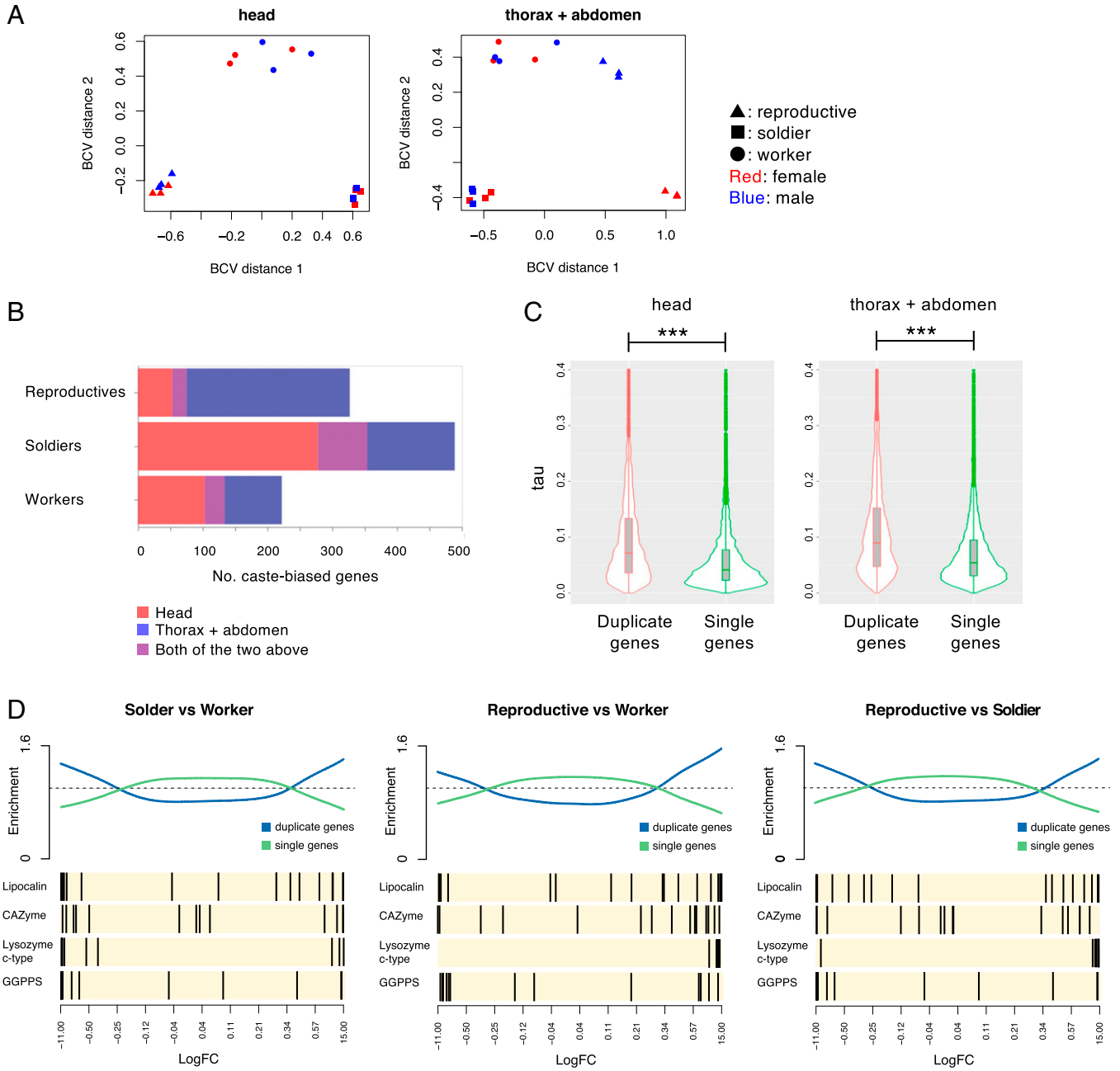

Fig. 2. Caste-specific transcriptome analysis and the enrichment of duplicate genes for caste-biased genes. (*A*) Multidimensional scaling plot of RNA-seq data showing the relatedness between the expression profiles of different castes (reproductive, soldier, and worker) and sexes (male and female). *Left* plots RNA-seq data from head samples, and *Right* plots data from thorax + abdomen samples. Three biological replicates were analyzed for each condition and were plotted individually. (*B*) Numbers of caste-biased genes with >twofold higher expression levels than in the other two castes. The colors in each bar indicate the differences identified from RNA-seq data. (*C*) Violin plots showing the distribution of the tau indices of duplicate genes and single genes. Tau values range between 0 and 1, with low values indicating invariable and constitutive expression between castes and higher values supporting caste specificity. In both analyzed body parts, the tau values of duplicate genes were significantly greater than those of single genes ($P < 2.2 \times 10^{-16}$, Wilcoxon rank sum test). (*D*) Enrichment of caste-DEGs (DEGs among castes) for duplicate genes. In each comparison between castes (soldier versus worker, reproductive versus worker, and reproductive versus soldier), all genes are ranked and ordered by log-fold-change value along the horizontal axis. The black bars mark the positions of genes. Genes of sociality-related functions highlighted in the text are selected and plotted in *Lower*. Curved lines in *Upper* show the relative enrichment of duplicated genes (blue line) or single genes (green line) relative to uniform ordering.

Dataset S1). These caste-biased genes should account for the specialized functions of each caste. Soldier samples exhibited the highest number of caste-specific genes, suggesting the highly specialized functions of the soldier caste (Fig. 2*B*). This was consistent with the finding of a previous RNA-seq analysis of the Eastern subterranean termite *Reticulitermes flavipes* indicating that a majority of the identified DEGs were soldier specific (55); 73 of the 93 identified DEGs were up- or down-regulated specifically in the soldier caste. In addition to these soldier-specific *R. flavipes* genes (e.g., *troponin C* and fatty *acyl-CoA reductase*), the caste-biased genes identified in our transcriptome analysis of *R. speratus* included genes previously reported to be caste-biased genes in other termites (56–59), such as *vitellogenin* (reproductives), *geranylgeranyl pyrophosphate synthase* (soldiers), and *beta-*

glucosidase genes (probably associated with cellulase; workers). This consistency between the transcriptome analyses of different termite species indicates that the RNA-seq analysis conducted in this study is reliable and that the regulation and perhaps the functions of these caste-biased genes are conserved across the termite lineage.

The caste-DEGs were enriched for gene ontology terms related to a wide array of functions (*SI Appendix*, Table S3), such as terpenoid metabolism, hormone metabolism, chitin metabolism, hydrolase activity, oxidoreductase activity, lipid metabolism, signaling, eye pigmentation, serine-type peptidase, thermotaxis, heme binding, and lysozyme activity. Protein motifs enriched in the caste-DEGs were also identified, including cytochrome P450, lipocalin, lysozyme, glycosyl hydrolase family, TGF-beta, and trypsin motifs (*SI Appendix*, Table S4). The genes of these categories were further investigated and discussed in the context of termite biology related to 14 categories (lipocalins, GGPP synthases, cellulases, lysozyems, sex determination, epigenetics, chemosensory system, biogenic amines and neuropeptides, juvenile hormone, ecdysone, insulin signaling, wing formation, immunity, insecticide target, and detoxification) in *Gene Duplication and Caste-Biased Gene Expression* and *SI Appendix*.

Among the 1,773 TRGs that were restricted to Isoptera, the termite-shared TRGs showed strong enrichment for caste-DEG (Fisher's exact test, $P < 1.0 \times 10^{-7}$ for head samples and $P < 1.0 \times 10^{-10}$ for thorax + abdomen samples), while the TRGs found only in *R. speratus* (orphan genes) did not ($P = 0.99$ and $P = 0.97$, respectively).

To investigate the relationship between caste-biased gene expression and DNA methylation, we analyzed differential methylation levels among three castes (reproductives, workers, and soldiers). The BS-seq data showed that the global CpG methylation patterns were very similar among the castes (*SI Appendix*, Fig. S3 *A* and *B*), in contrast to the methylation pattern of *Z. nevadensis*, in which DNA methylation differs strongly between castes (winged adults versus final-instar larvae) and is strongly linked to caste-specific splicing (24). Instead, gene body DNA methylation in *R. speratus* seems to be important for the expression of housekeeping genes, as reported in the drywood termite *C. secundus* (21). Housekeeping genes exhibited a high degree of gene body methylation in all castes of *R. speratus*, while caste-biased genes showed a significantly lower level of DNA methylation (*SI Appendix*, Fig. S3 *C* and *D*).

Gene Duplication and Caste-Biased Gene Expression. Evolutionary novelties are often brought about by gene duplications (25, 26). Our ortholog analysis comparing the *R. speratus* gene repertoire with those of 88 other arthropods identified 1,396 multigene families that are duplicated in the *R. speratus* genome (*SI Appendix*, Table S5). Interestingly, compared to the genome as a whole, the set of caste-DEGs described in *Transcriptome Differentiation among Castes* was significantly enriched for genes in multigene families (X-squared = 218.62, df = 1, $P < 2.2 \times 10^{-16}$). We also calculated the tau score as a proxy of the caste specificity of gene expression for all genes and found that duplicated genes were significantly more caste specific than single-copy genes ($P < 2.2 \times 10^{-16}$, Wilcoxon rank sum test) in both transcriptome datasets (head and thorax + abdomen) (Fig. 2*C*). Additionally, gene set tests showed that sets of duplicated genes were differentially expressed in all pairwise comparisons between castes (Fig. 2*D*). These data highlight the important roles of duplicated genes for the genetic regulation of caste differentiation in *R. speratus*.

Multigene Families Related to Caste-Specific Traits in *R. speratus*. Caste-biased multigene families were associated with diverse functional categories, some of which were strongly related to caste-specific behaviors and tasks. Here, we highlight five families, namely, lipocalins (protein transporters for social communication and physiological signaling), cellulases (carbohydrate-active enzymes [CAZymes] or worker wood digestion), lysozymes (immune-related genes for social immunity), geranylgeranyl diphosphate (GGPP) synthases (metabolic enzymes for the production of soldier defensive chemicals), and a termite-specific gene family with unknown functions, as examples of multigene families relevant to termite sociality. Molecular evolution studies have shown that the redundancy caused by gene duplication may allow one paralog to acquire a new function (neofunctionalization) or divide the ancestral function among paralogs (subfunctionalization) (25, 26). We are particularly interested in the evolutionary impact of gene duplication on caste specialization through neo/subfunctionalization.

Lipocalins. Lipocalins belong to a family of proteins with molecular recognition properties, such as the ability to bind a range of small hydrophobic molecules (e.g., pheromones) and specific cell surface receptors and to form complexes with soluble macromolecules (60). A previous study identified a gene of the lipocalin family, SOL1, that is exclusively expressed in the mandibular glands of mature soldiers of the rotten-wood termite *Hodotermopsis sjostedti* (61). SOL1 is thought to function as a signaling molecule for defensive social interactions among termite colony members (53). Moreover, RNA-seq analysis showed that a lipocalin gene, *Neural Lazarillo homolog 1* (*ZnNlaz1*), was specifically expressed in soldier-destined larvae in an incipient colony of *Z. nevadensis* (62). Gene function and protein localization analyses suggested that ZnNLaz1 was a crucial regulator of soldier differentiation through the regulation of trophallactic interactions with a queen. Thus, it was of interest that the lipocalin-related motif (Pfam PF00061; lipocalin/cytosolic fatty-acid binding protein family) was significantly enriched in the list of caste-DEGs (FDR < 0.05; (*SI Appendix*, Table S4).

We identified 18 lipocalin family genes in the *R. speratus* genome (Fig. 3 *A*–*C* and *SI Appendix*, Table S6). The number of lipocalin genes was larger than those in other insects (Fig. 3*C*). Phylogenetic analysis of lipocalin family genes identified in arthropods, including three termite species, revealed a highly dynamic evolutionary history of this protein family (Fig. 3*A*). Two subfamilies, namely, clades A and B, were observed to have experienced extensive expansion in the termite lineage (Fig. 3 *A* and *C*). The most drastic expansion was found in clade A, which included *H. sjostedti* SOL1. In this clade, 5, 8, 3, and 4 genes were identified in *R. speratus*, *M. natalensis*, *Z. nevadensis*, and *C. secundus*, respectively, and extensive and independent gene expansions occurred in each species. Clade B was also composed of genes with termite lineage–specific duplication. In many cases, these lipocalin genes were found in tandem arrays in the *R. speratus* genome (Fig. 3*B*). The inferred phylogenetic tree indicated that duplications in each clade occurred after the divergence of termites from a common ancestor. We detected signs of positive selection in the four *R. speratus* genes of Clades A and B (RS008823, RS008824, RS13913, and RS008884) (*SI Appendix*, Fig. S4).

A comparison of the transcriptome among castes revealed that most lipocalin genes (15 of 18) showed caste-biased gene expression (Figs. 2*D* and 3*B*). The caste specificity, however, varied among genes, regardless of sequence similarity and positional proximity in the genome. In particular, the expression levels of genes in clades A and B differed greatly among castes. For example, *RS008823* and *RS008824* displayed solder-specific expression, the expression of *RS013912* was biased toward workers, and *RS013913* was down-regulated in soldiers. *RS008881* and *RS008884* were exclusively expressed in the queen body (thorax + abdomen), while *RS008882*, a gene located next to the two aforementioned genes, showed quite different expression patterns, with high expression levels in heads, especially those of workers. These results indicate that termite lipocalin genes have undergone dynamic expansion in terms of the gene repertoire and regulatory

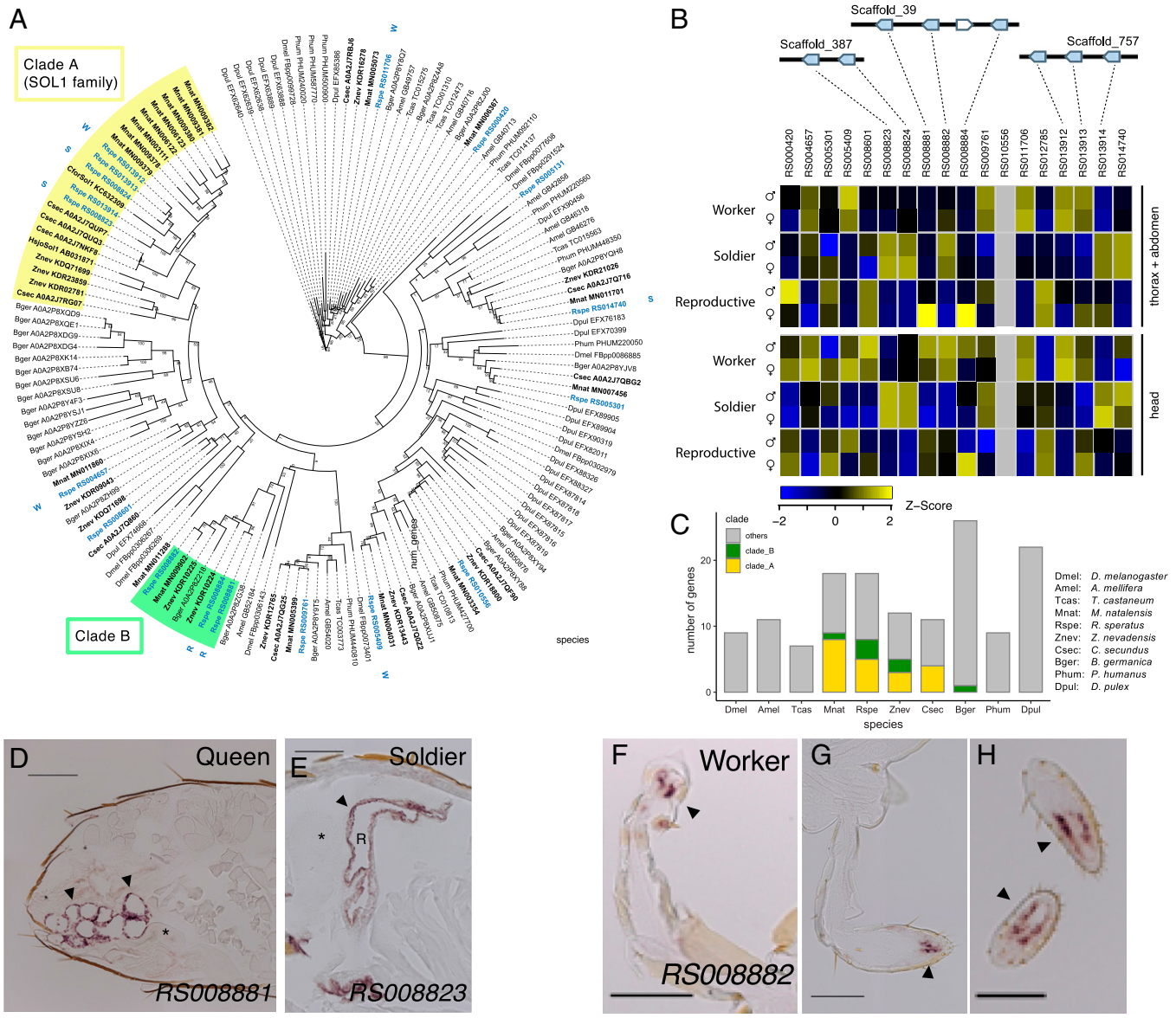

Fig. 3. Lipocalin genes in *R. speratus*. (*A*) Maximum likelihood tree of lipocalin homologs based on the amino acid sequences obtained with a log gamma model. Branches leading to clade A and clade B, which show gene family expansion specific to termite sublineages, are marked in yellow and green, respectively. Caste-DEGs (DEGs among castes) are marked as R, S, or W beside gene names, indicating biases toward the reproductive, soldier, or worker caste, respectively. (*B*) Lipocalin multigene clusters in the *R. speratus* genome and their relative expression levels among castes. The heatmap shows the Z-scores of the log(RPKM+1) values in the caste-specific transcriptome. (*C*) Comparison of the number of lipocalin subclasses among representative arthropods. (*D*) Vertical cryosection of the queen abdomen subjected to in situ hybridization with an antisense digoxigenin-labeled *RS008881* mRNA probe. The accessory gland cell layer is stained dark (arrowhead), in contrast to the other ovarian tissues, including the spermatheca (asterisk). (Scale bar, 0.2 mm.) (*E*) Photographs of in situ hybridization for *RS008823* mRNA in the soldier head. The front of the head is on the left side. The gland cell layer surrounding the frontal gland reservoir (R) is stained dark (arrowhead). The asterisk indicates the brain. (Scale bar, 0.1 mm.) (*F–H*) Vertical cryosection of the worker antenna (*F*) and horizontal cryosections of the worker labial palp (*G*, right palp) and maxillary palp (*H*, the last segment of the left [*Upper*] and right [*Lower*] palp) subjected to in situ hybridization for *RS008882* mRNA. Tissues around some sensilla are stained dark (arrowhead). (Scale bar, 0.1 mm.) Photographs of cryosections hybridized with sense probes (negative controls) are shown in *SI Appendix*, Fig. S5.

diversification of caste-biased expression. This gene expansion and regulatory diversification of lipocalins may have facilitated the evolution of the molecules involved in signaling during caste development and between individuals through social interactions.

To address the caste-specific function of the lipocalin paralogs, the expression patterns of several selected caste-biased lipocalin genes were examined by in situ hybridization (Fig. 3 *D–H* and *SI Appendix*, Fig. S5). *RS008881*, a queen-biased lipocalin gene, was found to be expressed exclusively in the accessory glands of the ovary (Fig. 3*D*). The next gene on the same scaffold, *RS008882*, was shown to be specifically expressed in worker antennae and maxillary/labial palps (Fig. 3 *F–H*). *RS008823*, a soldier-biased gene, was expressed exclusively in the frontal gland cells of the soldier heads (Fig. 3*E*). Note that ovaries and frontal glands develop during postembryogenesis in a caste-specific manner (i.e., ovaries in queens and frontal glands in soldiers) in the pathway of caste differentiation in *R. speratus*. Antennae and maxillary/labial palps are not caste specific but are crucial sensory organs, especially for blind termite immatures, such as workers. Given that animal lipocalins generally work as carrier proteins (63), there is a possibility that focal termite lipocalins bind and convey some molecules to targets from caste-specific organs

[e.g., egg-recognition pheromone and soldier defensive and/or inhibitory substances (64–66)], or participate in sensory reception, as observed for odorant-binding proteins (67).

Cellulases. Lignocellulose degradation in termites is achieved by a diverse array of CAZymes produced by the host and its intestinal symbionts. The repertoire of CAZyme families in the genome of *R. speratus* did not show considerable differences from those of other nonxylophagous insects, such as honeybees and fruit flies (*SI Appendix*, Fig. S6). However, we found gene family expansion and expressional diversification of glycoside hydrolase family (GH) 1 and GH9 members (Fig. 4*A* and *SI Appendix*, Table S7). The majority of GH1 and GH9 members are β-glucosidases (BGs; EC 3.2.1.21) and endo-β-1,4-glucanases (EGs; EC 3.2.1.4), respectively, which are essential for cellulose digestion in termites (68).

We identified 16 GH1 paralogs (*SI Appendix*, Table S7). Such gene expansion of GH1 is also observed in the genome of other termites, but the reason for this expansion remains elusive (69). Although the phylogenetic tree divided these GH1 paralogs into seven distinct groups (clades A to G in Fig. 4*A*), most of them were tandemly located in the genome of *R. speratus* (Fig. 4 *B* and *C*). The predominantly expressed BG gene was *RS004136*, and the expression of this gene was clearly biased toward the body (thorax + abdomen) in reproductives and workers (Fig. 4*B*). This gene formed a rigid clade with bona fide BGs reported from the salivary glands or midgut of termites (clade A in Fig. 4*A*) (70), suggesting that this gene is involved in cellulose digestion in *R. speratus*. Indeed, in situ hybridization analysis showed that *RS004136* was specifically expressed in the salivary glands of workers (Fig. 4 *D* and *E* and *SI Appendix*, Fig. S7*A*). Other GH1 members showed a wide variety of expression patterns across castes and body parts (Fig. 4*B*). Some of these genes might have diversified their functions, other than wood digestion that are related to termite sociality, as observed for egg-recognition pheromones (71). A typical example of such diversification was provided by *RS004624*, which was expressed specifically in the abdomens of queens (Fig. 4*B*). The peptide sequence of this gene showed a monophyletic relationship with that of Neofem2 of *C. secundus* (clade G in Fig. 4*A*), which is a queen-recognition pheromone that probably functions in the suppression of reproductive emergence (72). In situ hybridization showed that *RS004624* was specifically expressed in the accessory glands of queen ovaries (Fig. 4 *F* and *G* and *SI Appendix*, Fig. S7*B*), suggesting that *RS004624* is involved in enzymatic activities in queen-specific glands. Together with the results for a queen-biased lipocalin (*RS008881*), this finding indicates that the queen accessory glands may produce some queen-specific pheromones. Like lipocalins, GH1 paralogs are also typical examples of multigene family members participating in caste-specific tasks that may have been acquired by gene duplication resulting in neo- or subfunctionalization.

We found four paralogs of GH9 in *R. speratus* (*SI Appendix*, Fig. S8 and Table S7). Although several insect GH9 EGs have acquired the ability to hydrolyze hemicellulose (73), the neo- or subfunctionalization of termite EGs has yet to be clarified. Intriguingly, we found that the GH9 member *RS006396* was weakly but uniformly expressed across all termite body parts and castes. This result suggests that some GH9 members also perform a function other than that of cellulase, as is the case for GH1.

Lysozymes. The immune system of termites is of particular interest because the group living of termites with nonsclerotized and nonpigmented epidermis and microbe-rich habitats puts them at high risk for pathogenic infections (74). Thus, defense against pathogenic microbes is important for termites. In the *R. speratus* genome, we identified 251 immune-related genes (*SI Appendix*, Table S8). The repertoire and number of immune-related genes of *R. speratus* showed no large differences compared to those of other insect species with the notable exception of lysozymes (*SI Appendix*, Fig. S9). Lysozymes are involved in bacteriolysis through the hydrolysis of β-1,4-linkages in the peptidoglycans present in bacterial cell walls, and three distinct types of lysozymes, chicken- or conventional-type (c-type), goose-type (g-type), and invertebrate-type (i-type) lysozymes, have been found in animals (75). We identified 13 and 3 genes encoding c-type and i-type lysozymes, respectively, and the number of lysozyme genes was larger than those in other insects (*SI Appendix*, Fig. S9 and Table S9). Phylogenetic analysis revealed that c-type lysozymes underwent extensive gene duplications in the sublineage leading to *R. speratus* (*SI Appendix*, Fig. S10*A*). Six c-type lysozymes formed a tandem array on scaffold_859 (*SI Appendix*, Fig. S10*B*), probably generated by repeated tandem gene duplication events. Interestingly, most of the c-type lysozyme genes showed caste-biased expression. Three genes (*RS014698*, *RS100022*, and *RS100023*) exhibited high expression levels compared to those of other lysozyme genes and were expressed in a soldier-specific manner, while *RS100026* was expressed in a worker-specific manner, and *RS100024* and *RS100025* were highly expressed in both workers and soldiers (*SI Appendix*, Fig. S10*B* and Table S9). The differential expression patterns of the lysozyme genes of *R. speratus* may correspond to the division of labor among castes in terms of colony-level immunity.

It is also possible that duplicated lysozymes may have functions in processes other than immunity. A previous study indicated that the salivary glands of *R. speratus* secrete c-type lysozymes to digest bacteria ingested by termites through social feeding behavior (76). The same lysozyme genes are also expressed in the queen ovaries and eggs and play a role in egg recognition as proteinaceous pheromones in *R. speratus* (66, 71). We could not find identical sequences of these lysozyme genes in our gene models, but these sequences were most closely related to *RS002400* with 88% nucleotide identity, which occupied the basal position of the lineage-specific gene expansion (*SI Appendix*, Fig. S10*A*).

GGPP synthase. Whole-genome comparison of *R. speratus* with *Z. nevadensis* and *M. natalensis* revealed a 270-kb *R. speratus*–specific fragment in scaffold_31, while the rest of this scaffold showed very high syntenic conservation among the three termites (Fig. 5*A*). We found that the *R. speratus*–specific region was encompassed by a tandemly duplicated gene cluster composed of 13 genes encoding GGPP synthase (Fig. 5*B* and *SI Appendix*, Table S10). GGPP synthase catalyzes the consecutive condensation of an allylic diphosphate with three molecules of isopentenyl diphosphate to produce GGPP, an essential precursor for the biosynthesis of diterpenes, carotenoids, and retinoids (77–79). The extensive duplication of GGPP synthase paralogs observed in *R. speratus* is unusual because the genomes of other insects surveyed have only a single copy of GGPP synthase gene. The phylogenetic analysis of GGPP synthase homologs revealed two clusters, corresponding to a possibly ancestral group (including *RS007484*) and an apical group (including the other paralogs identified) (Fig. 5*C*). The latter cluster also contained some GGPP synthase paralogs obtained from the termitid *Nasutitermes takasagoensis* (56).

Transcriptome data indicated that all of the GGPP synthase genes except for *RS007484*, which was a member of the ancestral group in the phylogenetic tree, showed caste-biased expression, and caste specificity varied across the paralogs (Fig. 5*B*). Specifically, *RS100010*, *RS007480*, *RS100012*, *RS100015*, *RS100016*, *RS100017*, and *RS007483* showed soldier-specific expression, while *RS007481*, *RS007482*, and *RS100013* showed reproductive-specific expression (Fig. 5*B*). Several GGPP synthase genes have been identified in some termite species and are known to function in a caste-specific manner [e.g., the soldiers of *N. takasagoensis* synthesize defensive polycyclic diterpenes for use in chemical defense via a process mediated by high expression of the GGPP synthase gene in the frontal gland (80)]. It has been reported that the soldiers of *Reticulitermes* have a frontal gland in which diterpenes are synthesized, although their biological role is not fully understood (81–83). Consequently, it is possible

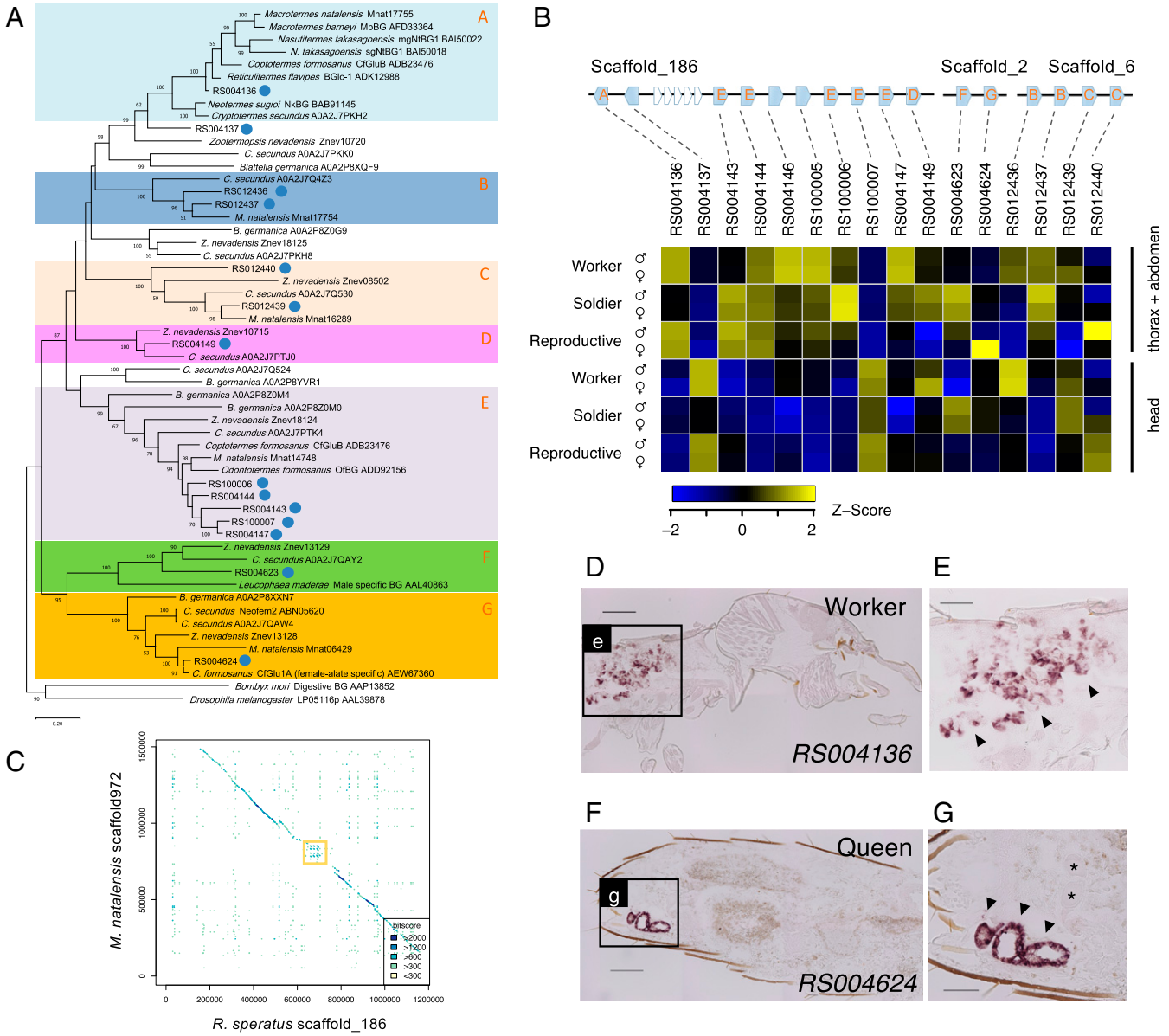

Fig. 4. GH1 in the *R. speratus* genome. (*A*) Maximum likelihood tree of GH1 genes based on the amino acid sequences obtained with an LG+G model. A total of 14 of 16 GH1 genes in *R. speratus* were used; two genes (*RS004146* and *RS100005*) were removed from the analysis due to incomplete retrieval of the coding sequences from gapped scaffolds. GH1 subclasses are colored and labeled A to G. (*B*) GH1 multigene clusters in the *R. speratus* genome and their expression levels. Letters A to G on the gene structures represent GH1 subclasses categorized in the phylogenetic tree in (*A*). The heatmap shows the Z-scores of the log(RPKM+1) values in the caste-specific transcriptome. (*C*) Synteny comparison around the GH1 multigene cluster region (orange rectangle) between the *R. speratus* and *M. natalensis* genomes. (*D*) Vertical cryosection of the worker thorax subjected to in situ hybridization with an antisense digoxigenin-labeled *RS004136* mRNA probe. The head is on the right side. (Scale bar, 0.2 mm.) (*E*) Magnified view of the worker thorax. The salivary gland cells are specifically stained dark (arrowhead). (Scale bar, 0.1 mm.) (*F*) Vertical cryosection of the queen abdomen subjected to in situ hybridization for *RS004624* mRNA. (Scale bar, 0.2 mm.) (*G*) Magnified view of the queen ovary. The accessory gland cell layer is stained dark (arrowhead), in contrast to the other ovarian tissues, including ovarioles with two oocytes (asterisks). (Scale bar, 0.1 mm.) See *SI Appendix*, Fig. S7 *A* and *B* for the negative controls of the in situ hybridization experiments (*D*–*G*).

that the soldier-specific GGPP synthases identified to date are involved in chemical defense. Indeed, in situ hybridization revealed that the soldier-specific GGPP synthase *RS100016* was expressed exclusively in the soldier frontal gland, as shown in a previous study (84) (Fig. 5*D* and *SI Appendix*, Fig. S7*C*). It is also possible that reproductive-specific GGPP synthases are involved in the metabolism of other diterpenes, such as the synthesis of pheromones; in particular, *RS007481* shows strong queen-specific expression in the thorax and abdomen and may play a role in the synthesis of queen substances.

Under the branch-site model of codon substitutions (85), significant positive selection was detected in five branches of the *R. speratus* GGPP synthase family tree (*SI Appendix*, Fig. S11): ancestral branches 1 and 2, and the branches leading to RS100017 (branch 3), RS100012 (branch 4), and RS007483 (branch 5). These results suggest that all GGPP synthase paralogs of *R. speratus* except the ancestral type *RS007484* have experienced positive selection and finally acquired novel roles in the production of defensive and/or pheromonal substances.

The TY Family, a Novel Gene Family Restricted to Termites. Numerous studies have shown that novel genes (e.g., TRGs) play important roles in the evolution of novel social phenotypes in hymenopteran social insects (9, 86, 87). We found that termite-shared

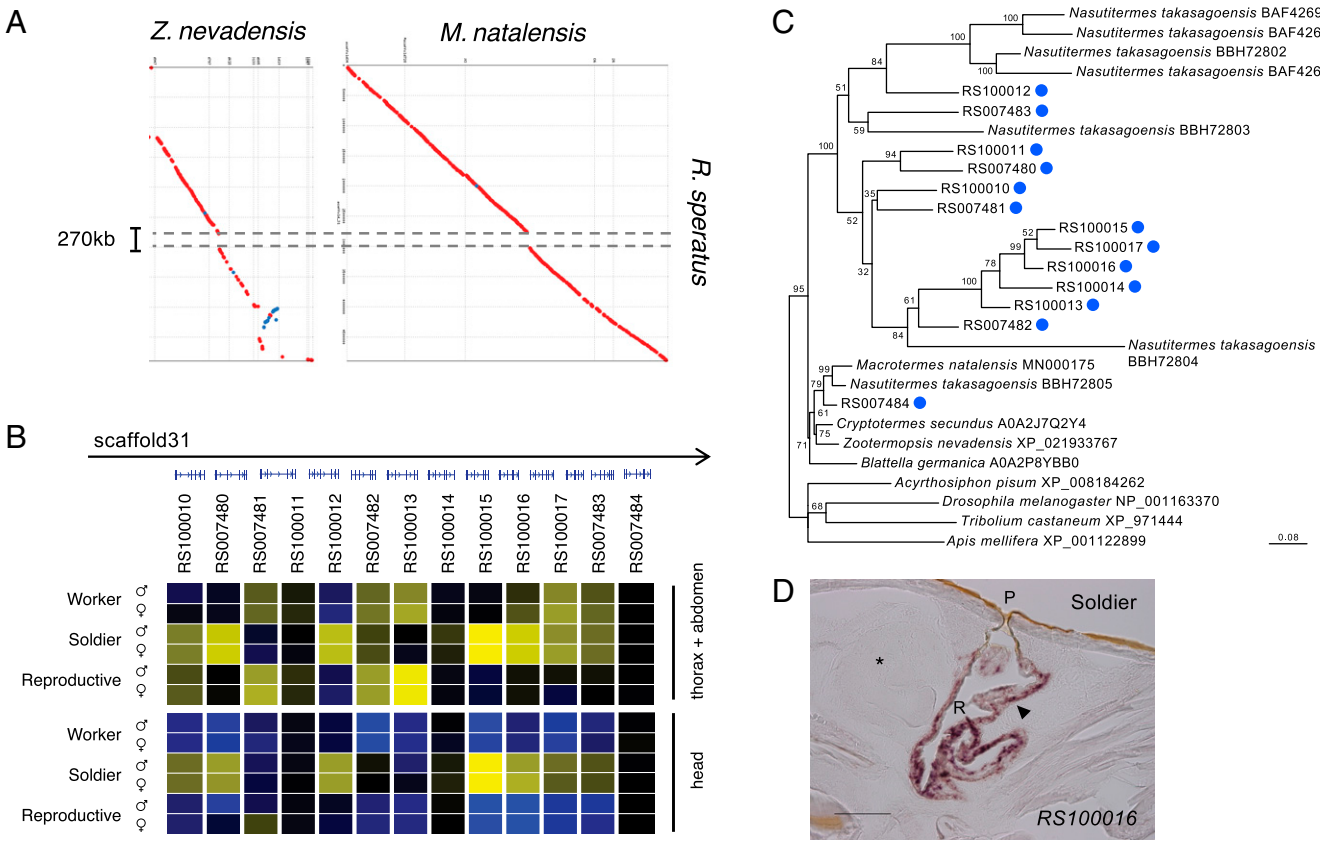

Fig. 5. GGPP synthase homologs in the *R. speratus* genome. (*A*) Synteny comparison around GGPP synthase loci among three termites, *R. speratus*, *M. natalensis*, and *Z. nevadensis*. *R. speratus*–specific insertions were found, where GGPP synthase paralogs were tandemly duplicated in the *R. speratus* genome. (*B*) Genomic location and gene expression of *R. speratus* GGPP synthase homologs. The heatmap shows the expression levels calculated as the mean-centered log(RPKM+1). Yellow indicates high expression, while blue denotes low expression. Black represents the mean level of expression among the castes. Note that the heatmap of *RS007484* is almost entirely black for all samples, which indicates that its expression was invariable among castes, while most of the other paralogs showed caste-biased expression. (*C*) Maximum likelihood tree of GGPP synthase homologs with an LG+G model. *R. speratus* genes are marked with blue circles. (*D*) Vertical cryosection of a soldier head subjected to in situ hybridization for *RS100016* mRNA. The front of the head is on the left side. The gland cell layer surrounding the frontal gland reservoir (R) is stained dark (arrowhead). The asterisk indicates the brain. The frontal pore (P) discharging the frontal gland secretion is also observed. (Scale bar, 0.1 mm.) See *SI Appendix*, Fig. S7C for the negative control experiment.

TRGs showed strong enrichment for caste-DEGs. A striking example of caste-biased TRGs is a tandem array of three novel genes (Fig. 6*A* and *SI Appendix*, Table S11), *RS001196*, *RS001197*, and *RS001198*, for which no significant homologs were found in any organisms outside of termite clades in our sequence similarity search, although the German cockroach, *Blattella germanica*, seems to have a single copy of a homolog of these genes. These three genes were expressed at extremely high levels (up to 250,000 reads per kilobase per million [RPKM]) in *R. speratus*, which constituted ~30% of the worker head transcriptome, and their expression was strongly biased across the three castes (Fig. 6*A*). Each gene encodes a short peptide of ~60 amino acids in length that contained a secretion signal peptide in the N-terminal region followed by a middle part rich in charged amino acid residues and a C-terminal region rich in polar amino acids, with an unusually high number of tyrosine residues (Fig. 6*B*). Here, we refer to this novel class of peptides as the tyrosine-rich peptide family (TY family). The three TY genes shared modest sequence similarity with each other, suggesting that they are paralogs derived by tandem duplication. Duplicated TY family orthologs were also found in the genomes of *Z. nevadensis*, *C. secundus*, and *M. natalensis* (*SI Appendix*, Fig. S12). We estimated pairwise evolutionary rates (the ratio of nonsynonymous to synonymous substitutions, K_a/K_s) between *R. speratus* and *Z. nevadensis* for *RS001196*, *RS001197*, and *RS001198*. The K_a/K_s of each gene was 0.15, 0.16, and 0.03, respectively, indicating that they evolved under strong purifying selection and suggesting a conserved function in the termite lineage.

Facilitation of Caste Specification by Gene Duplication. Recent advances in sociogenomics in different social insects are improving our understanding of the genetic bases of social evolution, which include the co-option of genetic toolkits of conserved genes, changes in protein-coding genes, cis-regulatory evolution leading to genetic network reconstruction, epigenetic modifications, accumulation of transposable elements, and TRGs (16, 22, 51, 88). In addition to these components, our genomic and transcriptomic analyses of *R. speratus* highlighted the significance of gene duplication for caste specialization. Gene duplication is, in general, a key source of genetic innovation that plays a role in the evolution of phenotypic complexity through sub- or neofunctionalization (26). Regarding eusocial evolution in insects, Gadagkar (89) first pointed out the importance of gene duplication, indicating that "genetic release followed by diversifying evolution" made possible the appearance of multiple caste phenotypes in social insects. According to Gadagkar's hypothesis, duplicated genes can be released from the constraints of original selection, leading to new directional evolution for caste-specific functions (e.g., queen- or worker-trait genes). Many decades later, genomic analyses revealed many instances of gene family expansions related to chemical communication, insulin, and vitellogenin signaling pathways in hymenopteran insects and termites (20,

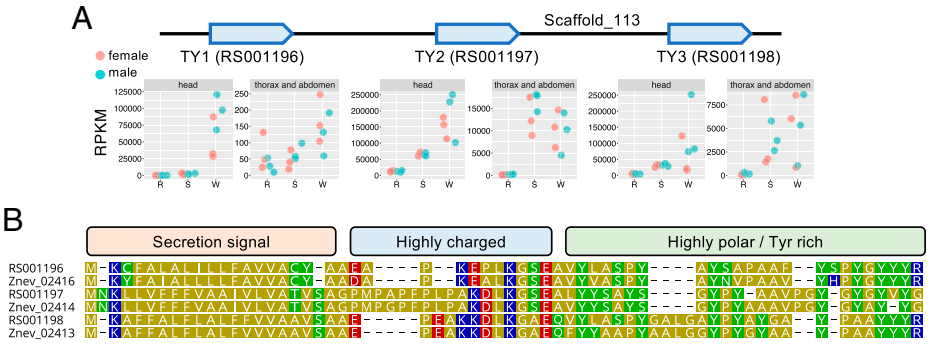

Fig. 6. The TY family, a novel secretion gene family identified from taxonomically termite-restricted genes. (*A*) Genomic locations and caste-biased expression patterns of TY family genes. R, S, and W indicate reproductive, soldier, and worker caste, respectively. (*B*) Multiple alignment of TY homologs of *R. speratus* and *Z. nevadensis*. Protein motifs and structural characteristics are represented.

21, 30–39), leading to the appreciation that gene duplications might have been commonly co-opted during social evolution in many social insects (23).

This study revealed that gene duplication associated with caste-biased gene expression is prevalent in the *R. speratus* genome. The list of duplicated genes encompasses a wide array of functional categories related to the social behaviors in termites as exemplified by transporters such as lipocalins (communication and physiological signaling; compare refs. 53 and 62), digestive enzymes such as CAZymes, immune-related genes such as lysozymes (social immunity), and metabolic enzymes such as GGPP synthase (social defense). This study also demonstrated that caste-specific expression patterns differed among in-paralogs. Although such paralogous genes were often observed in tandem in the genome, their expression patterns were often independent of one another, showing differential caste biases in many cases. Additionally, discordant caste biases in transcriptional expression were observed among closely related paralogs with similar coding sequences, as represented by the low correlation between phylogenetic position and caste specificity (Fig. 3*A*). Although the regulatory and evolutionary mechanisms underlying caste-biased expression patterns are elusive, these examples strongly suggest that gene duplications have facilitated caste specialization, leading to social evolution in termites.

After the gain of caste-biased gene regulation, subfunctionalization and/or neofunctionalization seems to have occurred, leading to caste-specific expression and caste-specialized functions. For example, in the case of the lipocalin family, lipocalin paralogs were generated by lineage-specific functional expansion in caste-specific organs or tissues: a queen-specific lipocalin (*RS008881*) was expressed specifically in the ovarian accessory glands, while a soldier-biased lipocalin (*RS008823*) was expressed exclusively in the frontal glands of soldier heads (Fig. 3 *D* and *E*). Taken together, these results lead us to hypothesize that, in termites, caste specification through gene duplication proceeded via following three steps: 1) gene family expansion by tandem gene duplication, 2) regulatory diversification leading to an expression pattern restricted to a certain caste, and 3) subfunctionalization and/or neofunctionalization of the gene products conferring caste-specific functions. As an exaptation of these steps, the case in which one (or several) of the multiple functions of pleiotropic genes is allocated and specialized to a duplicated gene copy might have led to caste-specific subfunctionalization (25, 26).

Recently, it was suggested that the evolution of phenotypic differences among castes of honeybees was associated with the gene duplication, based on the finding that duplicated genes presented higher levels of caste-biased expression than singleton genes (90). It was also shown that the level of gene duplication was correlated with social complexity in bees (superfamily Apoidea) (90). On the other hand, it was noted that a large number of genomics studies of hymenopteran species have not revealed such patterns (91, 92). Thus, it is still an open question whether gene duplication is a shared mechanism facilitating the evolution of caste systems in social insects with independent origins. Future studies targeting a broad range of taxa with different social systems will answer this question.

Materials and Methods

Insects. All mature colonies of *R. speratus* used for genome, RNA, and BS-seq were collected in Furudo, Toyama Prefecture, Japan (*SI Appendix*, Table S12). Detailed information on the samples is provided in *SI Appendix, Supplementary Methodology*.

Genome Sequencing and Assembly. We used female secondary reproductives (nymphoids I and II) for genome sequencing. Genomic DNA was isolated from each individual using a Genomic-tip 20/G (Qiagen). We examined five microsatellite loci to confirm whether the individuals were homozygous at these loci and shared the same genotype. We generated two paired-end libraries using a TruSeq DNA Sample Preparation Kit (Illumina) with insert sizes of ∼250 and ∼800 base pairs (bp) (*SI Appendix*, Table S13). Four mate-pair libraries with peaks at ∼3, ∼5, ∼8, and ∼10 kb were also generated using a Nextera Mate Pair Sample Preparation Kit (Illumina) (*SI Appendix*, Table S13). These libraries were sequenced using an Illumina HiSeq system with the 2 × 151-bp paired-end sequencing protocol. The reads of the paired-end and mate-pair libraries were assembled using ALLPATHS-LG (build No. 47878) (93) with the default parameters.

Gene Prediction. A protein-coding gene reference set of *R. speratus* was generated based on two main sources of evidence: aligned *R. speratus* transcripts and aligned homologous proteins of other insects and a set of ab initio gene predictions. The *R. speratus* RNA-seq reads were assembled de novo using Trinity (r2013_0814) (94) and then mapped to the genome using Exonerate (version 2.2.0) (95). We processed homology evidence at the protein level using the reference protein sequences of seven sequenced insects including *Z. nevadensis* and Blattodea protein sequences predicted from RNA-seq of *Periplaneta americana* and *N. takasagoensis* (detailed dataset information is provided in *SI Appendix, Supplementary Methodology*). These proteins were split-mapped to the *R. speratus* genome with Exonerate. These models were merged using the EvidenceModeler (r2012-06-25) (96), which yielded 15,584 gene models. A total of 74 genes were manually inspected and corrected. In particular, tandemly duplicated genes were prone to incorrect gene prediction with erroneous exon–exon connections across homologs. The final set of 15,591 genes was designated Rspe OGS1.0. The quality of OGS1.0 was evaluated by assessing two types of evidence: homology and expression data. Among the 15,591 genes, 12,996 (83.3%) showed hits in the National Center for Biotechnology Information nonredundant protein sequences database, 10,440 (70.0%) included known protein motifs defined in the Pfam database, and 14,302 (91.7%) showed evidence of expression with a threshold of RPKM = 1.0 in at least one sample of caste-specific RNA-seq data. In total, 15,577 genes (99.9%) showed evidence of the existence of homologs and/or expression.

Orthology Inference and Gene Duplication Analysis. Orthology determination among three termites: Orthologous genes among three termite species, *R. speratus*, *Z. nevadensis*, and *M. natalensis* (protein sequences of gene models RspeOGS1.0, ZnevOGSv2.229, and MnatOGS3, respectively), were determined by pairwise comparisons with InParanoid version 4.1, followed by a three-species comparison with MultiParanoid (97, 98). The *M. natalensis* gene set (MnatOGS3) was built in this study using a pipeline similar to that used for *R. speratus*.

Orthology analysis with arthropods: The orthology relationships of *R. speratus* genes (OGS1.0) with other arthropod genes were analyzed by referring to

the OrthoDB gene orthology database version 8 (87 arthropod species) (99). We grouped *R. speratus* genes with the OrthoDB ortholog group using a two-step clustering procedure. For each *R. speratus* protein, BLASTP was used to find similar proteins among the arthropod proteins, and the ortholog group of the top hit was provisionally assigned to the query *R. speratus* gene. Then, ortholog grouping was evaluated by comparing the similarity levels (BLAST bit scores) among members within the focal ortholog group. We retained the grouping if the BLAST bit score between the query *R. speratus* gene and the top arthropod gene was higher than the minimal score within the original cluster members. Among the 15,591 *R. speratus* OGS1.0 genes, 12,434 were clustered into 9,033 OrthoDB Arthropod ortholog groups. Gene duplication was assessed based on this clustering. If two or more members of one species were included in a single ortholog group, it was regarded as a multigene family.

RNA-seq. Old workers (sixth and seventh instars) and soldiers were collected from each colony. To collect primary reproductives, dealated adults were randomly chosen from each colony in accordance with the method described in the literature (100), and female–male pairs were mated (*SI Appendix*, Table S2). Kings and queens were sampled after 4 mo. Each individual was divided into the head and body (thorax + abdomen). We prepared RNA-seq libraries for 12 categories based on castes (reproductives, workers, and soldiers), sexes (males and females), and body parts (head and thorax + abdomen). Three biological replications of the 12 categories were performed from three different field colonies totaling 36 RNA-seq libraries (*SI Appendix*, Table S2). All Illumina libraries prepared using a TruSeq Stranded mRNA Library Prep kit were subjected to single-end sequencing of 101-bp fragments on HiSeq 2500. The cleaned reads were mapped onto the genome with TopHat version 2.1.0 guided by the OGS1.0 gene models. Transcript abundances were estimated using featureCounts and normalized with the trimmed mean of M-values algorithm in edgeR. DEGs among castes and between sexes were detected in each body part (head/thorax and abdomen) according to a GLM with two factors, namely, caste and sex using edgeR with the conditions set to a FDR < 0.01 and a log2-fold change of the expression level > 1.

To quantify caste specificity from the RNA-seq data, we used the tau score, which was originally developed to measure the tissue specificity of an expression profile (101). The index tau is defined as follows:

$$\tau = \frac{\sum_{i=1}^{N}(1 - x_i)}{N - 1},$$

where N is the number of categories (i.e., castes in this study), and x_i is the expression level normalized by the maximal component value. We used RPKM+1 values as representations of expression levels.

See *SI Appendix*, *Supplementary Methodology* for whole-genome alignment, annotation of gene models/repeat sequences, methylome analysis, and RNA in situ hybridization.

Data and Code Availability. Data from whole-genome sequencing, transcriptome sequencing, and methylome sequencing have been deposited in the DDBJ database under BioProject accessions PRJDB2984, PRJDB5589, and PRJDB11323, respectively. The analyzed data including genome assembly, gene prediction, annotation, and gene expression (Datasets S2–S5) are available through FigShare (https://doi.org/10.6084/m9.figshare.c.5483235). The *R. speratus* genome browser is available at http://www.termite.nibb.info/retsp/. Custom R and Ruby scripts were deposited into Github (https://github.com/termiteg/retsp_genome_paper).

ACKNOWLEDGMENTS. We thank R. H. Suzuki and A. Karasawa for experimental support, T. Nishiyama and M. Hasebe for discussion on genome analyses, N. Kanasaki and K. Kai for rearing insects, T. Shibata, S. Ohi, T. Aizu, H. Ishizaki, and H. Asao for next-generation sequencing (NGS), and K. Yamaguchi for NGS data management. Computations were partially performed on the supercomputers at the Data Integration and Analysis Facility, National Institute for Basic Biology. This study was funded by the Japan Society for the Promotion of Science/Ministry of Education, Culture, Sports, Science and Technology KAKENHI Grant Nos. 25128705, 24570022, 16K07511, JP19H03273, 22128008, 19K22294, and 221S0002 and NIBB Collaborative Research Programs (20-323).

1. E. Szathmáry, J. M. Smith, The major evolutionary transitions. *Nature* **374**, 227–232 (1995).
2. J. Korb, J. Heinze, Major hurdles for the evolution of sociality. *Annu. Rev. Entomol.* **61**, 297–316 (2016).
3. E. O. Wilson, *The Insect Societies* (Belknap Press, 1971).
4. G. E. Robinson, C. M. Grozinger, C. W. Whitfield, Sociogenomics: Social life in molecular terms. *Nat. Rev. Genet.* **6**, 257–270 (2005).
5. C. M. Grozinger, Y. Fan, S. E. R. Hoover, M. L. Winston, Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Mol. Ecol.* **16**, 4837–4848 (2007).
6. R. Bonasio *et al.*, Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* **329**, 1068–1071 (2010).
7. P. G. Ferreira *et al.*, Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biol.* **14**, R20 (2013).
8. R. Bonasio *et al.*, Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr. Biol.* **22**, 1755–1764 (2012).
9. B. Feldmeyer, D. Elsner, S. Foitzik, Gene expression patterns associated with caste and reproductive status in ants: Worker-specific genes are more derived than queen-specific ones. *Mol. Ecol.* **23**, 151–161 (2014).
10. S. Patalano *et al.*, Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13970–13975 (2015).
11. R. Libbrecht, P. R. Oxley, L. Keller, D. J. C. Kronauer, Robust DNA methylation in the clonal raider ant brain. *Curr. Biol.* **26**, 391–395 (2016).
12. D. S. Standage *et al.*, Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. *Mol. Ecol.* **25**, 1769–1784 (2016).
13. D. F. Simola *et al.*, Epigenetic (re)programming of caste-specific behavior in the ant *Camponotus floridanus*. *Science* **351**, aac6633 (2016).
14. K. M. Glastad, B. G. Hunt, M. A. D. Goodisman, DNA methylation and chromatin organization in insects: Insights from the ant *Camponotus floridanus*. *Genome Biol. Evol.* **7**, 931–942 (2015).
15. S. M. Rehan, A. L. Toth, Climbing the social ladder: The molecular evolution of sociality. *Trends Ecol. Evol.* **30**, 426–433 (2015).
16. A. L. Toth, S. M. Rehan, Molecular evolution of insect sociality: An eco-evo-devo perspective. *Annu. Rev. Entomol.* **62**, 419–442 (2017).
17. Y. Roisin, J. Korb, "Social Organisation and the Status of Workers in Termites" in *Biology of Termites: A Modern Synthesis*, D. E. Bignell, Y. Roisin, N. Lo, Eds. (Springer Netherlands, 2011), pp. 133–164.
18. J. Korb, B. Thorne, "Sociality in Termites" in *Comparative Social Evolution*, D. R. Rubenstein, P. Abbot, Eds. (Cambridge University Press, 2017), pp. 124–153.
19. A. Bucek *et al.*, Evolution of termite symbiosis informed by transcriptome-based phylogenies. *Curr. Biol.* **29**, 3728–3734.e4 (2019).
20. N. Terrapon *et al.*, Molecular traces of alternative social organization in a termite genome. *Nat. Commun.* **5**, 1–12 (2014).
21. M. C. Harrison *et al.*, Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nat. Ecol. Evol.* **2**, 557–566 (2018).
22. J. Korb *et al.*, A genomic comparison of two termites with different social complexity. *Front. Genet.* **6**, 9 (2015).
23. J. Korb, Genes underlying reproductive division of labor in termites, with comparisons to social Hymenoptera. *Front. Ecol. Evol.* **4**, 45 (2016).
24. K. M. Glastad, K. Gokhale, J. Liebig, M. A. D. Goodisman, The caste- and sex-specific DNA methylome of the termite *Zootermopsis nevadensis*. *Sci. Rep.* **6**, 37110 (2016).
25. S. Ohno, *Evolution by Gene Duplication* (Springer, Berlin, Heidelberg, 1970).
26. H. Innan, F. Kondrashov, The evolution of gene duplications: Classifying and distinguishing between models. *Nat. Rev. Genet.* **11**, 97–108 (2010).
27. M. Lynch, J. S. Conery, The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).
28. M. Long, E. Betrán, K. Thornton, W. Wang, The origin of new genes: Glimpses from the young and old. *Nat. Rev. Genet.* **4**, 865–875 (2003).
29. G. C. Conant, K. H. Wolfe, Probabilistic cross-species inference of orthologous genomic regions created by whole-genome duplication in yeast. *Genetics* **179**, 1681–1692 (2008).
30. Y. Wurm *et al.*, The genome of the fire ant *Solenopsis invicta*. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5679–5684 (2011).
31. X. Zhou *et al.*, Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals sex-specific signatures of odor coding. *PLoS Genet.* **8**, e1002930 (2012).
32. X. Zhou *et al.*, Chemoreceptor evolution in Hymenoptera and its implications for the evolution of eusociality. *Genome Biol. Evol.* **7**, 2407–2416 (2015).
33. S. K. McKenzie, I. Fetter-Pruneda, V. Ruta, D. J. C. Kronauer, Transcriptomics and neuroanatomy of the clonal raider ant implicate an expanded clade of odorant receptors in chemical communication. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14091–14096 (2016).
34. A. W. Legan, C. M. Jernigan, S. E. Miller, M. F. Fuchs, M. J. Sheehan, Expansion and accelerated evolution of 9-exon odorant receptors in *Polistes* paper wasps. *Mol. Biol. Evol.* **38**, 3832–3846 (2021).
35. V. Chandra *et al.*, Social regulation of insulin signaling and the evolution of eusociality in ants. *Science* **361**, 398–402 (2018).
36. M. Corona *et al.*, Vitellogenin underwent subfunctionalization to acquire caste and behavioral specific expression in the harvester ant *Pogonomyrmex barbatus*. *PLoS Genet.* **9**, e1003730 (2013).
37. P. Kohlmeier, B. Feldmeyer, S. Foitzik, Vitellogenin-like A-associated shifts in social cue responsiveness regulate behavioral task specialization in an ant. *PLoS Biol.* **16**, e2005747 (2018).
38. S. Lin, J. Werle, J. Korb, Transcriptomic analyses of the termite, *Cryptotermes secundus*, reveal a gene network underlying a long lifespan and high fecundity. *Commun. Biol.* **4**, 384 (2021).

39. L. P. M. Kremer, J. Korb, E. Bornberg-Bauer, Reconstructed evolution of insulin receptors in insects reveals duplications in early insects and cockroaches. *J. Exp. Zoolog. B Mol. Dev. Evol.* **330**, 305–311 (2018).
40. E. Jongepier *et al.*, Remodeling of the juvenile hormone pathway through caste-biased gene expression and positive selection along a gradient of termite eusociality. *J. Exp. Zoolog. B Mol. Dev. Evol.* **330**, 296–304 (2018).
41. K. Krishna, D. A. Grimaldi, V. Krishna, M. S. Engel, Treatise on the Isoptera of the world. (Bulletin of the American Museum of Natural History, no. 377) (2013).
42. D. J. G. Inward, A. P. Vogler, P. Eggleton, A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology. *Mol. Phylogenet. Evol.* **44**, 953–967 (2007).
43. T. Bourguignon *et al.*, The evolutionary history of termites as inferred from 66 mitochondrial genomes. *Mol. Biol. Evol.* **32**, 406–421 (2015).
44. E. L. Vargo, C. Husseneder, Biology of subterranean termites: Insights from molecular studies of *Reticulitermes* and *Coptotermes*. *Annu. Rev. Entomol.* **54**, 379–403 (2009).
45. W. V. Harris, "Termites of the palearctic region" in *Biology of Termites*, K. Krishna, F.M. Weesner, Eds. (Academic Press, New York), pp. 295–313 (1970).
46. F. M. Weesner, "Termites of the nearctic region" in *Biology of Termites*, K. Krishna, F.M. Weesner, Eds. (Academic Press, New York), pp. 477–522 (1970).
47. S. Govorushko, Economic and ecological importance of termites: A global review. *Entomol. Sci.* **22**, 21–35 (2019).
48. K. Matsuura *et al.*, Queen succession through asexual reproduction in termites. *Science* **323**, 1687–1687 (2009).
49. S. Koshikawa, S. Miyazaki, R. Cornette, T. Matsumoto, T. Miura, Genome size of termites (Insecta, Dictyoptera, Isoptera) and wood roaches (Insecta, Dictyoptera, Cryptocercidae). *Naturwissenschaften* **95**, 859–867 (2008).
50. M. Seppey, M. Manni, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227–245 (2019).
51. S. T. C. Chak, S. E. Harris, K. M. Hultgren, N. W. Jeffery, D. R. Rubenstein, Eusociality in snapping shrimps is associated with larger genomes and an accumulation of transposable elements. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2025051118 (2021).
52. M. Poulsen *et al.*, Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14500–14505 (2014).
53. T. Miura, Developmental regulation of caste-specific characters in social-insect polyphenism. *Evol. Dev.* **7**, 122–129 (2005).
54. H. F. Nijhout, Development and evolution of adaptive polyphenisms. *Evol. Dev.* **5**, 9–18 (2003).
55. T. Wu, G. K. Dhami, G. J. Thompson, Soldier-biased gene expression in a subterranean termite implies functional specialization of the defensive caste. *Evol. Dev.* **20**, 3–16 (2018).
56. M. Hojo, S. Shigenobu, K. Maekawa, T. Miura, G. Tokuda, Duplication and soldier-specific expression of geranylgeranyl diphosphate synthase genes in a nasute termite *Nasutitermes takasagoensis*. *Insect Biochem. Mol. Biol.* **111**, 103177 (2019).
57. M. E. Scharf, D. Wu-Scharf, B. R. Pittendrigh, G. W. Bennett, Caste- and development-associated gene expression in a lower termite. *Genome Biol.* **4**, R62 (2003).
58. M. M. Steller, S. Kambhampati, D. Caragea, Comparative analysis of expressed sequence tags from three castes and two life stages of the termite *Reticulitermes flavipes*. *BMC Genomics* **11**, 463 (2010).
59. T. Weil, M. Rehli, J. Korb, Molecular basis for the reproductive division of labour in a lower termite. *BMC Genomics* **8**, 198 (2007).
60. D. R. Flower, The lipocalin protein family: Structure and function. *Biochem. J.* **318**, 1–14 (1996).
61. T. Miura *et al.*, Soldier caste-specific gene expression in the mandibular glands of *Hodotermopsis japonica* (Isoptera: termopsidae). *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13874–13879 (1999).
62. H. Yaguchi *et al.*, A lipocalin protein, Neural Lazarillo, is key to social interactions that promote termite soldier differentiation. *Proc. Biol. Sci.* **285**, 20180707 (2018).
63. M. Ruiz, D. Sanchez, C. Correnti, R. K. Strong, M. D. Ganfornina, Lipid-binding properties of human ApoD and Lazarillo-related lipocalins: Functional implications for cell differentiation. *FEBS J.* **280**, 3928–3943 (2013).
64. Y. Mitaka *et al.*, Caste-specific and sex-specific expression of chemoreceptor genes in a termite. *PLoS One* **11**, e0146125 (2016).
65. D. Watanabe, H. Gotoh, T. Miura, K. Maekawa, Social interactions affecting caste development through physiological actions in termites. *Front. Physiol.* **5**, 127 (2014).
66. K. Matsuura, T. Tamura, N. Kobayashi, T. Yashiro, S. Tatsumi, The antibacterial protein lysozyme identified as the termite egg recognition pheromone. *PLoS One* **2**, e813 (2007).
67. P. Pelosi, J.-J. Zhou, L. P. Ban, M. Calvello, Soluble proteins in insect chemical communication. *Cell. Mol. Life Sci.* **63**, 1658–1676 (2006).
68. H. Watanabe, G. Tokuda, Cellulolytic systems in insects. *Annu. Rev. Entomol.* **55**, 609–632 (2010).
69. G. Tokuda, Plant cell wall degradation in insects: Recent progress on endogenous enzymes revealed by multi-omics technologies. *Adv. In Insect Phys.* **57**, 97 (2019).
70. J. Ni, G. Tokuda, Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. *Biotechnol. Adv.* **31**, 838–850 (2013).
71. K. Matsuura, T. Yashiro, K. Shimizu, S. Tatsumi, T. Tamura, Cuckoo fungus mimics termite eggs by producing the cellulose-digesting enzyme β-glucosidase. *Curr. Biol.* **19**, 30–36 (2009).
72. J. Korb, T. Weil, K. Hoffmann, K. R. Foster, M. Rehli, A gene necessary for reproductive suppression in termites. *Science* **324**, 758 (2009).
73. M. Shelomi, B. Wipfler, X. Zhou, Y. Pauchet, Multifunctional cellulase enzymes are ancestral in Polyneoptera. *Insect Mol. Biol.* **29**, 124–135 (2020).
74. Q. Gao, G. J. Thompson, Social context affects immune gene expression in a subterranean termite. *Insectes Soc.* **62**, 167–170 (2015).
75. L. Callewaert, C. W. Michiels, Lysozymes in the animal kingdom. *J. Biosci.* **35**, 127–160 (2010).
76. A. Fujita, I. Shimizu, T. Abe, Distribution of lysozyme and protease, and amino acid concentration in the guts of a wood-feeding termite, *Reticulitermes speratus* (Kolbe): Possible digestion of symbiont bacteria transferred by trophallaxis. *Physiol. Entomol.* **26**, 116–123 (2001).
77. West, C. A, Biosynthesis of diterpenes. *Biosynthesis of Isoprenoid Compounds* **1**, 375–412 (1981).
78. K. Ogura, T. Koyama, Enzymatic aspects of isoprenoid chain elongation. *Chem. Rev.* **98**, 1263–1276 (1998).
79. K. C. Wang, S. Ohnuma, Isoprenyl diphosphate synthases. *Biochim. Biophys. Acta* **1529**, 33–48 (2000).
80. M. Hojo, T. Matsumoto, T. Miura, Cloning and expression of a geranylgeranyl diphosphate synthase gene: Insights into the synthesis of termite defence secretion. *Insect Mol. Biol.* **16**, 121–131 (2007).
81. G. D. Prestwich, The chemicals of termite societies (Isoptera). *Sociobiology* **14**, 175–191 (1988).
82. L. J. Nelson, L. G. Cool, B. T. Forschler, M. I. Haverty, Correspondence of soldier defense secretion mixtures with cuticular hydrocarbon phenotypes for chemotaxonomy of the termite genus *Reticulitermes* in North America. *J. Chem. Ecol.* **27**, 1449–1479 (2001).
83. A. Quintana *et al.*, Interspecific variation in terpenoid composition of defensive secretions of European *Reticulitermes* termites. *J. Chem. Ecol.* **29**, 639–652 (2003).
84. M. Hojo, K. Toga, D. Watanabe, T. Yamamoto, K. Maekawa, High-level expression of the Geranylgeranyl diphosphate synthase gene in the frontal gland of soldiers in *Reticulitermes speratus* (Isoptera: Rhinotermitidae). *Arch. Insect Biochem. Physiol.* **77**, 17–31 (2011).
85. Z. Yang, R. Nielsen, Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**, 908–917 (2002).
86. B. R. Johnson, N. D. Tsutsui, Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* **12**, 164 (2011).
87. S. Sumner, The importance of genomic novelty in social evolution. *Mol. Ecol.* **23**, 26–28 (2014).
88. C. R. Smith, A. L. Toth, A. V. Suarez, G. E. Robinson, Genetic and genomic analyses of the division of labour in insect societies. *Nat. Rev. Genet.* **9**, 735–748 (2008).
89. R. Gadagkar, The evolution of caste polymorphism in social insects: Genetic release followed by diversifying evolution. *J. Genet.* **76**, 167–179 (1997).
90. L. M. Chau, M. A. D. Goodisman, Gene duplication and the evolution of phenotypic diversity in insect societies. *Evolution* **71**, 2871–2884 (2017).
91. D. R. Rubenstein *et al.*, Coevolution of genome architecture and social behavior. *Trends Ecol. Evol.* **34**, 844–855 (2019).
92. A. J. Moore, K. M. Benowitz, From phenotype to genotype: The precursor hypothesis predicts genetic influences that facilitate transitions in social behavior. *Curr. Opin. Insect Sci.* **34**, 91–96 (2019).
93. S. Gnerre *et al.*, High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1513–1518 (2011).
94. M. G. Grabherr *et al.*, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
95. G. S. C. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
96. B. J. Haas *et al.*, Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
97. M. Remm, C. E. Storm, E. L. Sonnhammer, Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052 (2001).
98. A. Alexeyenko, I. Tamas, G. Liu, E. L. L. Sonnhammer, Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics* **22**, e9–e15 (2006).
99. E. V. Kriventseva *et al.*, OrthoDB v8: Update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res.* **43**, D250–D256 (2015).
100. K. Maekawa, K. Ishitani, H. Gotoh, R. Cornette, T. Miura, Juvenile hormone titre and vitellogenin gene expression related to ovarian development in primary reproductives compared with nymphs and nymphoid reproductives of the termite *Reticulitermes speratus*. *Physiol. Entomol.* **35**, 52–58 (2010).
101. I. Yanai *et al.*, Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).