Evolutionary transition of *doublesex* regulation in termites and cockroaches: from sex-specific splicing to male-specific transcription

Satoshi Miyazaki^{1*†}, Kokuto Fujiwara^{2†}, Keima Kai², Yudai Masuoka^{2,3}, Hiroki Gotoh⁴, Teruyuki Niimi^{5,6}, Yoshinobu Hayashi⁷, Shuji Shigenobu^{6,8} Kiyoto Maekawa^{2*}

¹ Graduate School of Agriculture, Tamagawa University, Machida, Tokyo 194-8610, Japan

² Faculty of Science, Academic Assembly, University of Toyama, Gofuku, Toyama, 930-8555, Japan

³ Institute of Agrobiological Sciences, NARO (National Agriculture and Food Research Organization), Tsukuba, Ibaraki 305-8634, Japan

⁴ Department of Biological Science, Faculty of Science, Shizuoka University, Suruga-ku, Shizuoka, 422-8529, Japan

⁵ Division of Evolutionary Developmental Biology, National Institute for Basic Biology, Okazaki, Aichi, 444-8585 Japan

⁶ Department of Basic Biology, School of Life Science, The Graduate University for Advanced Studies, SOKENDAI, Okazaki, Aichi, 444-8585 Japan

⁷ Department of Biology, Keio University, Yokohama, Kanagawa, 223-8521, Japan

⁸ NIBB Research Core Facilities, National Institute for Basic Biology, Okazaki, Aichi, 444-8585 Japan

* Correspondence e-mail: smiyazaki@agr.tamagawa.ac.jp, kmaekawa@sci.u-toyama.ac.jp

† contributed equally

Abstract

The sex determination gene *doublesex* (*dsx*), which encodes a transcription factor with two domains: oligomerisation domain 1 (OD1) and OD2, is conserved among insects. The sex-specific Dsx splicing isoforms regulate the transcription of target genes and trigger sex differentiation in all holometabolous insects examined to date. However, in some hemimetabolous insects, *dsx* is less conserved and not spliced sexually. Here, to elucidate evolutionary changes in *dsx* in the gene structure and its regulatory manner in termites, we searched genome and/or transcriptome databases for the OD1 and OD2 of *dsx* in six termite species and their sister group (woodroach). Molecular phylogenetic analysis identified OD1 sequences of termites and a woodroach clustered with *dsx* of holometabolous insects and regarded them as *dsx* orthologues. In the woodroach, a *dsx* orthologue containing OD2 was spliced sexually, as previously shown in other insects. However, OD2 were not found in all termite *dsx* orthologues. These orthologues were encoded by only a single exon in three termites with genome information; they were not alternatively spliced, but transcribed in a male-specific manner in two species examined. Evolution of *dsx* regulation from sex-specific splicing to male-specific transcription might be occurred at the early stage of social evolution in termites.

40 Introduction

41 In insects, sex determination is cell-autonomously controlled by a cascade composed of several genes,
 42 where the *doublesex* (*dsx*) encoding sex-specific transcription factors is located at the bottom ¹. In the
 43 cascade of fruit fly (*Drosophila melanogaster*), the primary signal, which is the number of X
 44 chromosome ² is sequentially transduced into the sex-specific splicing of *Sex-lethal*, *transformer*, and
 45 *dsx*, resulting in the sex-specific transcription of target genes responsible for their sex differentiation
 46 ^{3,4}. Although upstream genes encoding splicing regulators in this cascade differ among insect taxa, *dsx*
 47 is conserved among them, as well as in some crustaceans and chelicerates that are non-insect
 48 arthropods ^{1,5-7}. Moreover, the *dsx* orthologue is sexually spliced in all holometabolous insects
 49 examined to date ^{3,4}, thus, such *dsx* regulation is believed to be a general feature of the cascade in all
 50 insects. However, in hemimetabolous insects, the phylogenetic position of which is the interval
 51 between holometabolous insects and non-insect arthropods, *dsx* is sexually spliced in two of four
 52 species examined recently, but spliced alternatively and not sex-specifically in the others ^{8,9}.
 53 Furthermore, in some crustaceans and mites, *dsx* does not produce sex-specific splicing isoforms and
 54 is instead expressed in a sex-specific manner, which controls their male differentiation ^{5,7}. Therefore,
 55 in contrast to the knowledge based on holometabolous insects, the sex-specific regulation of *dsx* would
 56 be evolutionarily labile among insects and non-insect arthropods, and it could be diversified in

57 hemimetabolous insects.

58 Price et al. ⁶ searched public databases of insect genomes and transcriptomes for *dsx*
59 homologues, based on the presence of two domains conserved throughout holometabolous insects:
60 DNA-binding DM domain and dimerisation domain, referred to as oligomerisation domain (OD) 1 and
61 OD2, respectively. OD1 is shared among not only *dsx* but also *doublesex/mab-3 related transcription*
62 *factor (DMRT)* genes that are conserved among metazoans and involved in their sexual development
63 ^{10,11}, whereas OD2 is specific to *dsx* ¹². They failed to obtain any evidence of the presence of *dsx* in
64 several hemimetabolous insects, including termites (infraorder Isoptera or epifamily Termitoidae)
65 belonging to the monophyletic group within the cockroach clade ^{13,14}. In the lower (primitive) termite
66 *Zootermopsis nevadensis*, neither domain was found, and only OD1 was found in the higher (derived)
67 termite *Nastitermes takasagoensis* ⁶. The genome sequence of the German cockroach *Blattella*
68 *germanica* has been published, and a *dsx* homologue has been identified in the genome (gene ID:
69 PSN43312.1 ¹⁵). *Blattella dsx* was shown to contain both OD1 and OD2 (referred to as “DM domain”
70 and “Dsx Dimerization domain”, respectively, in Wexler et al. ⁹) and to be sexually spliced as in
71 holometabolous insects ⁹. Based on these results, we hypothesised that the *dsx* sequence (especially
72 OD1 and/or OD2) diversified during termite evolution, and then *dsx* sequence diversification affected
73 its regulatory manner. These hypotheses could be examined by a more comprehensive search of

complete genome sequences and transcriptomic data in termites and in their sister group, woodroach (*Cryptocercus* spp.)¹⁶⁻²¹.

The present study aimed to elucidate evolutionary changes in *dsx* in the gene structure and its regulatory manner in the clade of cockroaches and termites. First, we exhaustively searched genome and/or transcriptome databases of the woodroach *Cryptocercus punctulatus* and six termite species (*Reticulitermes speratus*, *Hodotermopsis sjostedti*, *Cryptotermes secundus*, and *Macrotermes natalensis*, as well as *Z. nevadensis* and *N. takasagoensis*) for the OD1 and OD2 of *dsx*, and then examined whether the searched sequences were *dsx* homologues based on the molecular phylogeny and synteny. Second, we performed gene expression analysis based on quantitative RT-PCR and published transcriptome data, for one woodroach species and two termite species. Third, to identify putative regulatory factors in the sex-specific transcription of termite *dsx*, transcription factor binding sites (TFBSs) were identified in the promoter region of the *dsx* orthologue of *R. speratus*. The expression levels of putative transcription factors were compared between the sexes based on published transcriptome data. On the basis of the results obtained, we concluded that *dsx* regulation shifted from sex-specific splicing to male-specific transcription at the early stage of social evolution in termites.

91 Results

92 Searches for *doublesex* orthologues in termites and woodroach

93 We performed BLAST searches using the translated OD1 sequence of *Blattella dsx* (45 amino acids)
 94 as a query against genome and/or transcriptome databases of six termite species and one woodroach
 95 species, after which two to four OD1 sequences were hit in each species. Then, BLAST searches were
 96 performed using the translated OD2 sequence of *Blattella dsx* (45 amino acids) as a query. A single
 97 sequence from the woodroach was hit (Cpun_comp8195_c0_seq1, Table 1), but no sequence was hit
 98 from any termites. Rapid amplification cDNA ends (RACE) PCR was performed using primers
 99 specific for Cpun_comp8195_c0_seq1 (Table S1), and a single full-length transcript was obtained only
 100 from females. Then, using a reverse primer specific to the male-specific exon of *Blattella dsx* (located
 101 at terminal codon and 3'UTR, Table S1), another 3' end of coding sequence was amplified from males.
 102 As these transcripts contained an OD1 upstream of OD2 (Fig. 1), it was designated as *C. punctulatus*
 103 *dsx* (*Cpun_dsx*). The amplified sequences downstream of OD2 were predicted to be the sex-specific
 104 exon(s) of *Cpun_dsx*. The determined nucleotide and putative amino acid sequences of the *Cpun_dsx*
 105 splicing variants are available in the DDBJ/EMBL/GenBank databases (accession no. xxxxx and
 106 xxxxxx).

107 Phylogenetic analysis of the OD1 nucleotide sequences revealed that all *dsx* reported in

108 holometabolous insects belonged to a single clade (*dsx* clade) (Fig. 2). This clade contained a single
109 sequence containing OD1, but no OD2, from each of the four termites *H. sjostedti* (c35221¹⁶), *R.*
110 *speratus* (not annotated, scaffold 6²⁰: 6307974..6309410), *N. takasagoensis* (G5ZWOJF02FLJ2Z¹⁶),
111 and *M. natalensis* (Mnat_08109¹⁸); two sequences from *Cryptotermes secundus* (XM_023861307 and
112 XM_023858380¹⁵) (Fig. 2, Table 1). No sequence from *Z. nevadensis* was detected within this clade.
113 The other OD1 sequences from insects formed clades specific to the *DMRT11*, *DMRT93*, and *DMRT99*
114 orthologues (Fig. 2). To examine whether the *dsx*-clade sequences in termites were *dsx* orthologues,
115 synteny analyses for *dsx* were conducted based on the proximity of the *dsx* and *prospero* (*pros*), which
116 is well conserved among seven different holometabolous and hemimetabolous insect orders⁹. The
117 gene encoding the *dsx*-clade sequence was located 83 kb and 80 kb close to the *pros* orthologues
118 (RS012493²² and MN005277²³) in scaffold 6 of *R. speratus* and scaffold 295 of *M. natalensis*
119 (Mnat_08109), respectively²⁰. Therefore, these *dsx*-clade genes were designated as the termite
120 orthologues of *dsx* (*Rspe_dsx* and *Mnat_dsx*, Table 1). In *C. secundus*, both *dsx*-clade sequences,
121 XM_023861307 and XM_023858380, and the *pros* orthologue (XM_023851644) were located in
122 scaffold 829 (length: 88.5 kb), 635 (length: 1.3 Mb), and 511 (length: 2.2 Mb), respectively. Because
123 the scaffold lengths (especially for scaffold 829) might be insufficient for the synteny analysis based
124 on proximity (17–245 kb⁹), evidence of synteny in *C. secundus* could not be obtained. The synteny in

H. sjostedti and *N. takasagoensis* could not be examined because of the unavailability of their genome data. On the basis of the OD1 sequence similarity (Fig. 2), we designated these as *dsx* orthologues (hereafter referred to as *Csec_dsx1*, *Csec_dsx2*, *Hsjo_dsx*, and *Ntak_dsx*, Table 1).

Splicing patterns and expression of *doublesex* homologues in termites and woodroach

To confirm the sex-specific splicing of *Cpun_dsx*, expression levels of the predicted sex-specific splicing isoforms were quantified by quantitative RT-PCR using primers specific to the predicted isoforms (Fig. 1 and Table S2). The predicted female- and male-specific isoforms was confirmed to be expressed abundantly in females ($t = -6.68, p < 0.01$, generalised linear model [GLM]) and males ($t = 7.88, p < 0.01$, GLM), respectively (Fig. 3), indicating that *Cpun_dsx* was spliced in a sex-specific manner, as shown in holometabolous insects and the German cockroach *B. germanica*^{9,15}. Next, to obtain the splicing isoforms of *Hsjo_dsx*, *Rspe_dsx*, and *Ntak_dsx*, we performed 3'-RACE PCRs using gene-specific primers (Table S2) and amplified only the single fragment downstream of OD1 for each species (Fig. 1). The amplified downstream sequences of each species were consistent with those obtained by the aforementioned BLAST search. Although *dsx* is composed of ca. five exons, of which two to three posterior exons are sexually spliced in almost all insects examined previously²⁴, each of the predicted full-length transcripts of *Rspe_dsx*, *Mnat_dsx*, *Csex_dsx1*, and *Csec_dsx2* was encoded

by a single exon in their genome. Based on the transcriptome data in *R. speratus*²¹, *Rspe_dsx* was expressed only in male reproductives (primary kings), soldiers, and workers, but not in females, and in both heads ($t = 3.99$, $p < 0.01$, GLM) and the other parts (referred to as just “body”) ($t = 10.97$, $p < 0.001$, GLM) (Fig. 4A). Quantitative RT-PCR also showed male-specific expression patterns even in the nymphs ($t = 5.69$, $p < 0.01$, GLM, Fig. 4B) and eggs ($t = 3.10$, $p < 0.05$, GLM, Fig. 4C). In addition, *Ntak_dsx* was expressed only in males, regardless of caste ($t = 3.70$, $p < 0.01$, GLM, Fig. 4D). These results indicated that termite *dsx* was not alternatively spliced, but transcribed in a male-specific manner.

Predicted binding sites and expression patterns of putative regulatory factors for male-specific transcription of *Rspe_dsx*

To examine the proximate mechanism for male-specific transcription of termite *dsx*, we searched for putative transcription factors that bind to the promoter region of *Rspe_dsx*, and those expressed in a male- or female-specific manner. First, 1 kb upstream of the transcription start site of *Rspe_dsx* was extracted as a transcriptional regulatory region, as defined by Toyota et al.²⁵. Next, *de novo* motif discovery was performed for the promoter using hypergeometric optimisation of motif enrichment (HOMER) v4.11²⁶, based on the equipped motif library of known transcription factors in insects.

159 Although 32 putative TFBSs were detected in the promoter, 20 homologues out of the 32 transcription
 160 factors existed in the *Reticulitermes* genome, of which five were duplicated in the genome (Table 2).
 161 Based on the transcriptome data in *R. speratus*²¹, however, none of them were expressed in a sex-
 162 specific manner in both heads and bodies, except for the *vielfaltig* (*vfl*) orthologue (GLM, Table 2).
 163 The expression levels of the *vfl* orthologue were significantly different between sexes only in the bodies
 164 of reproductives, but similar patterns were not observed in other castes (Fig. S1).

165

Discussion

The *dsx* orthologue of the woodroach was spliced in a sex-specific manner, similar to that of the German cockroach and holometabolous insects (Fig. 5). Surprisingly, however, termite *dsx* homologues did not exhibit sex-specific splicing, and their transcription was regulated in a male-specific manner. Male- or female-biased expression of *dsx* has been reported in some arthropods, for example, water fleas ^{5,25}, red claw crayfish ²⁷, Chinese shrimp ²⁸ (Crustacea), and mites ⁷ (Chelicerata), suggesting that sex-biased *dsx* transcriptions are an ancestral state in Arthropoda ⁹. In contrast, *dsx* produces sex-specific splice isoforms in most hemimetabolous and holometabolous insects examined to date ^{6,9,24} except for the lice *Pediculus humanus* and the hemipteran *Bemisia tabaci*. These two hemimetabolous insects have *dsx* homologues with non-sex-specific isoforms ^{8,9}. Consequently, the sex-specific alternative splicing of *dsx* would be acquired early in the evolution of insects, and secondarily lost in some hemimetabolous insects ⁹, including termites. Within the cockroach and termite clade, after the ancestral termites had derived from the common ancestor of *Cryptocercus* and termites (around 140.6 MYA, ranging from 112.6 to 170.5 MYA ²⁹), it is suggested that the sex-specific *dsx* splicing was lost, and the male-specific *dsx* transcription was acquired alternatively in termites (Fig. 5). No *dsx* orthologues in *Z. nevadensis* are mystery at present, and further exhaustive search should be needed.

Although *dsx* consists of approximately five exons in most insects ²⁴, two *dsx* paralogs (*dsx1* and *dsx2*) in water fleas consist of four and two exons, respectively ⁵. Its homologue in Chinese shrimp consists of two exons ²⁸, and those in mites ⁷ and termites consist of only a single exon (Fig. 5). Moreover, the present study showed no homologous OD2 in termite *dsx* (Fig. 5), suggesting that OD2 has indeed been lost or has diverged beyond our ability to detect it ⁶. OD2 is conserved in *dsx* orthologues with sex-specific splicing in holometabolous insects, as well as in those with sex-specific transcription in water fleas ^{5,25} and Chinese shrimp ²⁸ but not in red claw crayfish ²⁷. Both OD1 and OD2 are related to bind target DNA as a dimeric DNA binding unit in the fruit fly *D. melanogaster*: OD1 enables Dsx to bind target DNA and dimerises weakly in a DNA-dependent manner ¹², whereas OD2 is necessary for dimerisation and enhances DNA recognition of OD1 ³⁰. In addition, fly OD2 contains sex-specific spliced sequences, which may be involved in the formation of sex-specific units for transcriptional regulation of downstream target genes through either its sex-specific interactions with the transcriptional machinery or its sex-specific DNA binding ¹². The decreasing numbers of *dsx* exons in some arthropods and termites and the loss (or divergence) of OD2 in termites would be associated with the evolution of *dsx* regulation, although it remains unknown whether such evolutionary changes in the gene structure were the cause or consequence of the loss of sex-specific splicing.

The sex determination cascade has never been examined in termites. Almost all orthologues of sex determination genes reported in holometabolous insects (e.g., *transformer*, *transformer-2*, and *Sex-lethal*) are conserved in the genome of *R. speratus*²². However, these orthologues encode splicing factors and are unlikely to directly regulate male-specific *dsx* transcription. Male-specific transcription should be regulated by a transcription factor expressed in a sex-specific manner in every caste, as shown in termite *dsx*. Although we found 27 putative TFBSs in the *Rspe_dsx* promoter region, no orthologues of 27 transcription factors showed sex-specific expression patterns similar to those of *Rspe_dsx*. However, our motif search was based on the binding motifs found in the fruit fly *D. melanogaster*. In addition, unidentified binding motifs could be present out of the transcriptional regulatory region examined in this study. Further investigations are needed to determine whether the unidentified transcription factors expressed in a sex-specific manner would regulate male-specific *dsx* transcription in termites.

The transition of *dsx* regulation in termites is associated with the social evolution of cockroaches and termites (Fig. 5). Cooperative brood care by parents and generation overlap should have already been acquired in the last common ancestor of *Cryptocercus* and termites³¹. Therefore, the male-specific transcription of termite *dsx* would play a role in the regulation of reproductive division of labour, especially the fortress defence, with an altruistic sterile “soldier” caste that had occurred at the

origin of Isoptera ³². Some species, especially in the most derived family (Termitidae), have sex-specific or sex-biased soldier ratios and soldier differentiation pathways ³³. For example, all soldiers are females in most species of Termitinae and Macrotermitinae, whereas all soldiers are males in most of the examined species in Nasutitermitinae. Soldier differentiation requires high juvenile hormone titres in workers, and the strongly biased soldier-sex ratio might be caused by the differences in juvenile hormone titres (and probably related gene expression levels) between male and female workers ^{34,35}. In addition, reproductive caste differentiation is regulated in a sex-specific manner (Oguchi et al. ³⁶). Some “queen genes” with high expression levels in female reproductives are identified in some species (reviewed in Korb ³⁷). Furthermore, the upregulated genes in male reproductives, which are probably involved in male fertility, are also identified in species with available genome sequences ^{15,19,22}. Termite *dsx* could regulate this sexually dimorphic expression by upregulating or downregulating the expression of its target genes in a male-specific manner, resulting in reproductive and non-reproductive division of labour.

231 **Materials and Methods**

232 ***Insects***

233 Seven mature colonies of *R. speratus* were collected in Toyama Prefecture in 2016 and 2019. One *H.*
 234 *sjostedti* mature colony was collected in Yakushima Island, Kagoshima Prefecture in 2015. Two *N.*
 235 *takasagoensis* mature colonies were found on Ishigakijima Island, Okinawa Prefecture, in 2017 and
 236 2018. They were brought back to the laboratory and kept at ca. 25°C in constant darkness. Three
 237 families of *C. punctulatus* were collected at Mountain Lake Biological Station, Giles County, VA in
 238 April 2015-2017³⁸ and kept at 15°C in constant darkness. Testes and ovaries were dissected from three
 239 males and three females, respectively, and stored at -80°C until RNA extraction for subcloning and
 240 expression analyses of *Cpun_dsx*.

241

242 ***Blast searches for dsx and dmrt homologues in termites and cockroaches***

243 tBlastX searches were performed against genome and/or transcriptome databases of six termite species
 244 and one woodroach species (Table 1), using SequenceServer³⁹. Either the OD1 (45 amino acids) or
 245 OD2 (45 amino acids) sequences of *B. germanica dsx* were used as queries. Sequences with an E value
 246 less than 10⁻⁵ were selected as candidates for *dsx* orthologues and used for phylogenetic analysis.

247

248 ***Construction of phylogenetic tree of OD1 sequences and identification of dsx orthologues***

249 To identify the *dsx* orthologue, a phylogenetic tree of OD1 sequences (135 bp with no gaps) was
 250 constructed according to previous studies ^{5,40,41}. We used 84 OD1-containing genes (Table S3 'OTU').
 251 The DMRT7 genes of vertebrates (mouse and cattle) were used as the outgroups. Phylogenetic
 252 relationships were inferred using Bayesian inference (BI), maximum likelihood (ML), and maximum
 253 parsimony (MP) methods. For BI, the most appropriate model of sequence evolution was determined
 254 using the model selection option implemented in MEGA version 7.0.21 ⁴², and the GTR + G model
 255 was selected. Using MrBayes version 3.2.6 ⁴³, a total of 100,000 trees were obtained (ngen =
 256 10,000,000, samplefreq = 100). The first 25% of these (25,000) were discarded as burn-ins, and a 50%-
 257 majority-rule consensus tree was produced. For ML, 1,000 bootstrap replicates were performed based
 258 on the same model of sequence evolution as BI in MEGA 7.0.21, with the default tree inference options.
 259 For MP analysis, all characters were included and weighted equally, and 1,000 bootstrap replicates
 260 were performed using MEGA 7.0.21. The Subtree-Pruning-Regrafting algorithm, in which the initial
 261 trees were obtained by the random addition of sequences (10 replicates), was used.

262

263 ***Subcloning of the candidate dsx orthologues***

264 Female and male alate reproductives (winged adults) of *R. speratus* were obtained from three colonies

collected in 2016. According to previous studies^{44,45}, incipient colonies were established using unrelated female and male alates. Total 30 eggs and primary reproductives (queen and king) were obtained from 20 colonies at 1.5 months after colony establishment. Total RNA of the eggs was extracted and treated with DNase I using the ReliaPrep RNA Tissue Miniprep System (Promega, USA). Total RNA of the primary queens or kings was separately extracted from the body (except for the head parts) of three individuals using ISOGEN II (Nippongene, Japan). Total RNA of female or male alates of *N. takasagoensis* (using the colony collected in 2017) was separately extracted from the body (except for the head parts) of five individuals using ISOGEN II. Total RNA from the secondary queen or king of *H. sjostedti* (using the colony collected in 2015) was separately extracted from the body (except for the head part) of one individual using ISOGEN II. Female and male *C. punctulatus* adults were dissected, and the gonads (ovaries or testes) were collected from each individual. Total RNA was extracted from the gonads using ISOGEN I and II, and then treated with RNase-free DNase I (Takara, Japan). The purity and quantity of the extracted RNA were measured using a NanoVue spectrophotometer (GE Healthcare Bio-Sciences, Japan).

The 5' and 3' ends of *Cpun_dsx* were amplified using the SMARTer RACE 5'/3'kit (TaKaRa, Shiga, Japan) and gene-specific primers designed for OD2 ("*Cpun_dsx* OD2 5'RACE" and "*Cpun_dsx* OD2 3'RACE", Table S1). Its male-specific exon was amplified using Advantage® 2 Polymerase Mix

(TaKaRa), the gene-specific primer “*Cpun_dsx* OD2 3'RACE”, and a reverse primer in the male-specific exon of *Blattella dsx*, which was located at terminal codon and 3'UTR (“*Cpun_dsx* male-specific exon-R”, Table S1). The 3' ends of *Rspe_dsx*, *Hsjo_dsx*, and *Ntak_dsx* were also amplified using the SMARTer RACE 5'/3' Kit and newly designed gene-specific primers (Table S1). The amplified fragments were subcloned into the pGEM-T easy vector system (Promega), and the nucleotide sequences of each fragment were determined using ABI Prism Big Dye Terminator v3.1 Cycle Sequencing Kit in conjunction with a 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). The newly identified gene sequences were deposited in the GenBank/EMBL/DBJ database (Table 1).

RNA-seq analysis

RNA-seq data of *R. speratus* were used to compare the expression levels of *Rspe_dsx* (Table 1) among three castes (workers, soldiers, and reproductives), two body parts (heads and the remaining parts), and two sexes (biological triplicates; NCBI BioProject Accession No. PRJDB5589). The filtered RNA-seq reads were mapped onto their genome assembly²⁰ using TopHat v2.1.021. Transcript abundances were then estimated using the featureCounts program of the Subread package (Liao et al. 2013). To compare gene expression levels among castes and between sexes, first, counts per million (CPM) were

calculated from the estimated transcript abundances. Genes with at least CPM of 1 in at least three samples were kept for subsequent analyses. CPM values were then normalized with the trimmed mean of M-values (TMM) algorithm in edgeR (Robinson et al. 2010). Differential gene expression analyses were performed separately for each body parts using a GLM with two factors, namely, caste and sex implemented in edgeR, and then, genes with false discovery rate (FDR) < 0.01 were identified as genes expressed in a sex-specific manner. RPKM (Reads Per Kilobase Million) values were calculated by dividing the CPM values by the length of the genes in kilobases.

Real-time quantitative PCR

Gene-specific primers were designed against *Rspe_dsx*, *Ntak_dsx*, and *Cpun_dsx* using Primer3Plus⁴⁶ for real-time quantitative PCR (Table S2). Total RNA of female and male nymphs of *R. speratus* (from the colony collected in 2019) was separately extracted from the whole body of 10 individuals using ISOGEN II. Total RNA of male (minor and medium) workers, female (medium and major) workers, and male soldiers of *N. takasagoensis* (from the colony collected in 2018) were separately extracted from the whole body of five individuals using ISOGEN II. Total RNAs derived from the gonads of adult woodroaches were also used. DNase treatment was performed using the same method described above. According to previous studies^{44,45}, incipient colonies (queen-king and two-queen colonies) of

316 *R. speratus* were established using alates derived from three colonies collected in 2019. Total RNA of
 317 eggs produced asexually and sexually, which develop into females and both sexes, respectively, were
 318 extracted from 30 eggs and treated with DNase I using ReliaPrep RNA Tissue Miniprep System
 319 (Promega). cDNAs were synthesised from the purified RNA using a High-Capacity cDNA Reverse
 320 Transcription Kit (Applied Biosystems). Biological replications (n = 3-5) were prepared for each
 321 category. Expression analyses were performed using Thunderbird SYBR qPCR Mix (Toyobo, Japan)
 322 with a Mini Option Real-time PCR system (Bio-Rad, Japan) and an Applied Biosystems 7500 Fast
 323 Real-Time PCR System (Applied Biosystems).

324 To determine a sustainable internal control gene for *R. speratus*, the expression levels of six
 325 genes [*EF1-α* (accession no. AB602838), *NADH-dh* (no. AB602837), *β-actin* (no. AB520714), *GstD1*
 326 (gene id RS001168), *EIF-1* (RS005199), and *RPS18* (RS015150)] were evaluated using the GeNorm
 327 ⁴⁷ and NormFinder ⁴⁸ software (Table S4). For *C. punctulatus* and *N. takasagoensis*, according to
 328 previous studies ^{38,49}, three genes [*β-actin* (nos. Cp_TR19468 and AB501107), *NADH-dh*
 329 (Cp_TR49774 and AB50119), and *EF1-α* (AFK49795 and AB501108)] were evaluated (Tables S5 and
 330 S6). All gene-specific primers were designed using Primer3Plus (Table S2). We confirmed the
 331 amplification of a single PCR product using dissociation curves and product sizes.

332 The expression levels were statistically analysed using GLMs with gamma error distribution

and log link function. The fixed effects were sex, caste, and their interaction. Multiple comparisons were performed using general linear hypothesis testing (glht function) with Tukey adjustment in the “multcomp” R package⁵⁰. These analyses were conducted using R ver. 4.0.3 (available at <http://cran.r-project.org/>).

Search for transcription factor binding motif

The *Reticulitermes* genome was loaded into the database of HOMER v4.11 using the “loadGenome.pl” program. The “findMotifsGenome.pl” program was executed to search for *Drosophila* motif collections using the “-mcheck” option against the promoter of *Rspe_dsx*, which was the 1.0 kbp upstream region from its transcription start site. The nucleotide sequences of *Drosophila* transcription factors that were detected in their binding sites were subjected to a BlastX search for orthologues against the protein database of *R. speratus*³⁸ using BLAST+ (2.10.1)⁵¹. The expression levels of these orthologues were extracted from their transcriptome data²⁷ and statistically analysed using GLMs with gamma error distribution and log link function, as mentioned above.

348 **Acknowledgement**

349 Chrisnine A. Nalepa kindly provided us with *Cryptocercus punctulatus*. We thank Shutaro Hanmoto
 350 for his assistance with laboratory work. We also thank Toru Miura and Masatoshi Matsunami for
 351 their constructive suggestions regarding this study. This study was partly supported by JSPS
 352 KAKENHI (Grant Numbers JP19K06860 to SM, JP20K06816 to YH, and JP19H03273 to KM) and
 353 by the NIBB Collaborative Research Program (Nos. 19-335 and 20-323).
 354 SM, HG, and KM planned and designed the experiments.

355

356 **Author contributions**

357 SM, KF, KK, and YM collected the samples. SM, KF, KK, and YM performed the
 358 experiments. SM analysed the data. SM, KF, and KM wrote the manuscript. All authors
 359 read and approved the final manuscript.

360

361 **Another Information**

362 All authors declare that they have no conflict of interest.

363

364

365

References

- 1 Verhulst, E. C., van de Zande, L. & Beukeboom, L. W. Insect sex determination: it all evolves around *transformer*. *Current Opinion in Genetics & Development* **20**, 376-383, doi:10.1016/j.gde.2010.05.001 (2010).
- 2 Erickson, J. W. & Quintero, J. J. Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in *Drosophila*. *PLOS Biology* **5**, e332, doi:10.1371/journal.pbio.0050332 (2007).
- 3 Gempe, T. & Beye, M. Function and evolution of sex determination mechanisms, genes and pathways in insects. *Bioessays* **33**, 52-60, doi:10.1002/bies.201000043 (2011).
- 4 Verhulst, E. C. & van de Zande, L. Double nexus—*Doublesex* is the connecting element in sex determination. *Briefings in Functional Genomics* **14**, 396-406, doi:10.1093/bfpg/elv005 (2015).
- 5 Kato, Y., Kobayashi, K., Watanabe, H. & Iguchi, T. Environmental sex determination in the branchiopod crustacean *Daphnia magna*: deep conservation of a *doublesex* gene in the sex-determining pathway. *PLOS Genetics* **7**, e1001345, doi:10.1371/journal.pgen.1001345 (2011).
- 6 Price, D. C., Egizi, A. & Fonseca, D. M. The ubiquity and ancestry of insect *doublesex*. *Scientific Reports* **5**, 13068, doi:10.1038/srep13068 (2015).
- 7 Pomerantz, A. F. & Hoy, M. A. Expression analysis of *Drosophila doublesex*, *transformer-2*, *intersex*, *fruitless-like*, and *vitellogenin* homologs in the parahaploid predator *Metaseiulus occidentalis* (Chelicerata: Acari: Phytoseiidae). *Experimental and Applied Acarology* **65**, 1-16, doi:10.1007/s10493-014-9855-2 (2015).
- 8 Guo, L. *et al.* Identification and characterization of *doublesex* in *Bemisia tabaci*. *Insect Molecular Biology* **27**, 620-632, doi:10.1111/imb.12494 (2018).
- 9 Wexler, J. *et al.* Hemimetabolous insects elucidate the origin of sexual development via alternative splicing. *eLife* **8**, e47490, doi:10.7554/eLife.47490 (2019).
- 10 Kopp, A. *Dmrt* genes in the development and evolution of sexual dimorphism. *Trends in Genetics* **28**, 175-184, doi:10.1016/j.tig.2012.02.002 (2012).
- 11 Bellefroid, E. J. *et al.* Expanding roles for the evolutionarily conserved *Dmrt* sex transcriptional regulators during embryogenesis. *Cellular and Molecular Life Sciences* **70**, 3829-3845, doi:10.1007/s00018-013-1288-2 (2013).
- 12 An, W., Cho, S., Ishii, H. & Wensink, P. C. Sex-specific and non-sex-specific oligomerization domains in both of the *doublesex* transcription factors from *Drosophila melanogaster*. *Molecular and Cellular Biology* **16**, 3106-3111, doi:10.1128/mcb.16.6.3106 (1996).
- 13 Inward, D., Beccaloni, G. & Eggleton, P. Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biology Letters* **3**, 331-335, doi:10.1098/rsbl.2007.0102 (2007).
- 14 Lo, N. *et al.* Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. *Current Biology* **10**, 801-804, doi:10.1016/s0960-9822(00)00561-3 (2000).
- 15 Harrison, M. C. *et al.* Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nature Ecology & Evolution* **2**, 557-566, doi:10.1038/s41559-017-0459-1 (2018).

- Hayashi, Y. *et al.* Construction and characterization of normalized cDNA libraries by 454 pyrosequencing and estimation of DNA methylation levels in three distantly related termite species. *PLOS ONE* **8**, e76678, doi:10.1371/journal.pone.0076678 (2013).
- Mitaka, Y., Kobayashi, K. & Matsuura, K. Caste-, sex-, and age-dependent expression of immune-related genes in a Japanese subterranean termite, *Reticulitermes speratus*. *PLOS ONE* **12**, e0175417, doi:10.1371/journal.pone.0175417 (2017).
- Poulsen, M. *et al.* Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proceedings of the National Academy of Sciences* **111**, 14500-14505, doi:10.1073/pnas.1319718111 (2014).
- Terrapon, N. *et al.* Molecular traces of alternative social organization in a termite genome. *Nature Communications* **5**, 3636, doi:10.1038/ncomms4636 (2014).
- Shigenobu, S. *Reticulitermes speratus* genome assembly. doi: 10.6084/m9.figshare.14267342 (2021).
- Shigenobu, S. *R. speratus* caste RNA-seq. doi: 10.6084/m9.figshare.14267498 (2021).
- Shigenobu, S. *Reticulitermes speratus* gene models OGS1.0. doi: 10.6084/m9.figshare.14267381 (2021).
- Shigenobu, S. *Macrotermes natalensis* gene models MnGM3. doi: <https://doi.org/10.6084/m9.figshare.1437026> (2021).
- Baral, S., Arumugam, G., Deshmukh, R. & Kunte, K. Genetic architecture and sex-specific selection govern modular, male-biased evolution of *doublesex*. *Science Advances* **5**, eaau3753, doi:10.1126/sciadv.aau3753 (2019).
- Toyota, K. *et al.* Molecular cloning of *doublesex* genes of four cladocera (water flea) species. *BMC Genomics* **14**, 239, doi:10.1186/1471-2164-14-239 (2013).
- Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
- Zheng, J. *et al.* Identification and functional analysis of the *doublesex* gene in the redclaw crayfish, *Cherax quadricarinatus*. *Gene Expression Patterns* **37**, 119129, doi:10.1016/j.gep.2020.119129 (2020).
- Li, S., Li, F., Yu, K. & Xiang, J. Identification and characterization of a *doublesex* gene which regulates the expression of insulin-like androgenic gland hormone in *Fenneropenaeus chinensis*. *Gene* **649**, 1-7, doi:10.1016/j.gene.2018.01.043 (2018).
- Bucek, A. *et al.* Evolution of termite symbiosis informed by transcriptome-based phylogenies. *Current Biology* **29**, 3728-3734.e3724, doi:10.1016/j.cub.2019.08.076 (2019).
- Zhu, L. *et al.* Sexual dimorphism in diverse metazoans is regulated by a novel class of intertwined zinc fingers. *Genes Dev* **14**, 1750-1764 (2000).
- Evangelista, D. A. *et al.* An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea). *Proceedings of the Royal Society B: Biological Sciences* **286**, 20182076, doi:10.1098/rspb.2018.2076 (2019).
- Korb, J., Buschmann, M., Schafberg, S., Liebig, J. & Bagnères, A.-G. Brood care and social evolution in termites.

- 33 Bourguignon, T., Hayashi, Y. & Miura, T. Skewed soldier sex ratio in termites: testing the size-threshold hypothesis. *Insectes Sociaux* **59**, 557-563, doi:10.1007/s00040-012-0252-8 (2012).
- 34 Toga, K. *et al.* Sexual difference in juvenile-hormone titer in workers leads to sex-biased soldier differentiation in termites. *Journal of Insect Physiology* **87**, 63-70, doi:10.1016/j.jinsphys.2016.02.005 (2016).
- 35 Watanabe, D., Gotoh, H., Miura, T. & Maekawa, K. Social interactions affecting caste development through physiological actions in termites. *Frontiers in Physiology* **5**, doi:10.3389/fphys.2014.00127 (2014).
- 36 Oguchi, K., Maekawa, K. & Miura, T. Regulatory mechanisms underlying the differentiation of neotenic reproductives in termites: partial release from arrested development. *Frontiers in Ecology and Evolution* **9**, doi:10.3389/fevo.2021.635552 (2021).
- 37 Korb, J. Genes underlying reproductive division of labor in termites, with comparisons to social hymenoptera. *Frontiers in Ecology and Evolution* **4**, doi:10.3389/fevo.2016.00045 (2016).
- 38 Masuoka, Y., Toga, K., Nalepa, C. A. & Maekawa, K. A crucial caste regulation gene detected by comparing termites and sister group cockroaches. *Genetics* **209**, 1225-1234, doi:10.1534/genetics.118.301038 (2018).
- 39 Priyam, A. *et al.* Sequenceserver: a modern graphical user interface for custom BLAST databases. *Molecular Biology and Evolution* **36**, 2922-2924, doi:10.1093/molbev/msz185 (2019).
- 40 Kato, Y. *et al.* Molecular cloning and sexually dimorphic expression of DM-domain genes in *Daphnia magna*. *Genomics* **91**, 94-101, doi:10.1016/j.ygeno.2007.09.002 (2008).
- 41 Oliveira, D. C. S. G. *et al.* Identification and characterization of the *doublesex* gene of *Nasonia*. *Insect Molecular Biology* **18**, 315-324, doi:10.1111/j.1365-2583.2009.00874.x (2009).
- 42 Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**, 1870-1874, doi:10.1093/molbev/msw054 (2016).
- 43 Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574, doi:10.1093/bioinformatics/btg180 (2003).
- 44 Maekawa, K., Ishitani, K., Gotoh, H., Cornette, R. & Miura, T. 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*. *Physiological Entomology* **35**, 52-58, doi:10.1111/j.1365-3032.2009.00711.x (2010).
- 45 Miyazaki, S. *et al.* Intracolony genetic variation affects reproductive skew and colony productivity during colony foundation in a parthenogenetic termite. *BMC Evolutionary Biology* **14**, 177, doi:10.1186/s12862-014-0177-0 (2014).
- 46 Untergasser, A. *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research* **35**, W71-W74, doi:10.1093/nar/gkm306 (2007).
- 47 Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**, research0034.0031, doi:10.1186/gb-2002-3-7-research0034 (2002).

(2002).

48 Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research* **64**, 5245-5250, doi:10.1158/0008-5472.can-04-0496 (2004).

49 Toga, K., Hojo, M., Miura, T. & Maekawa, K. Expression and function of a limb-patterning gene *Distal-less* in the soldier-specific morphogenesis in the nasute termite *Nasutitermes takasagoensis*. *Evolution & Development* **14**, 286-295, doi:10.1111/j.1525-142X.2012.00545.x (2012).

50 Hothorn, T., Bretz, F. & Westfall, P. Simultaneous inference in general parametric models. *Biometrical Journal* **50**, 346-363, doi:10.1002/bimj.200810425 (2008).

51 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421, doi:10.1186/1471-2105-10-421 (2009).

52 Bourguignon, T. *et al.* Transoceanic dispersal and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. *Molecular Biology and Evolution* **35**, 970-983, doi:10.1093/molbev/msy013 (2018).

Figure legends

Figure 1. Structures of *dsx* transcripts in woodroach and termites. Full-length coding sequence of *dsx* transcripts in one woodroach and five termite species are shown as transcribed from left to right. The OD1 is shown as a filled box, whereas the sexually shared, female-specific, and male-specific OD2 is shown as a grey, red, and blue box, respectively. Numbers on the right of boxes are amino acid lengths of the translated *dsx*. Positions of primers for quantifying *Cpun_dsx* expression by quantitative RT-PCR are shown in arrows: “F”, “Rf”, and “Rm” indicate the shared forward primer, reverse primer for the female-specific splicing variant, and reverse primer for the male-specific variant, respectively (Table S2).

Figure 2. Molecular phylogenetic tree of *dsx* and *DMRT* homologues. Bayesian tree of *dsx* and *DMRT* of insects and crustaceans was constructed based on the OD1 sequences (135 bp with no gaps). Numbers shown above branches represent the Bayes posterior probabilities. Bootstrap values (1,000 replicates) are shown below branches to indicate the level of support in the ML and MP methods, respectively. An asterisk indicates that a node is not supported by the ML and MP methods.

Figure 3. Expression patterns of *Cpun_dsx* in the female and male gonads of *C. punctulatus* adults. Left and right panels show the relative expression levels of its female- and male-specific isoform to those of *β-actin*, respectively. “F”, “Rf”, and “Rm” indicate the primers shown in Fig. 1. The expression levels were compared between sexes using GLM analyses. ** indicates $p < 0.01$.

Figure 4. Sex-specific expression patterns of *Rspe_dsx* and *Ntak_dsx*. (A) Reads per kilo base per million mapped reads (RPKM) of *Rspe_dsx* in each sex and in each caste of *R. speratus*, which were re-analysed RNA-seq data (DRA010978, biological triplicates²¹). (B) *Rspe_dsx* expression levels relative to those of *NADH* in female and male nymphs. (C) Relative *Rspe_dsx* expression in eggs produced asexually and sexually. The formers are fated to develop into females, whereas the sexual fate of the latter was not determined. (D) *Ntak_dsx* expression levels relative to those of *NADH* in each sex and in each caste of *N. takasagoensis*. In (A) and (D), the effects of sex, caste, and their interaction on gene expression levels were evaluated using GLM followed by multiple comparison using linear hypothesis testing with Tukey adjustment in the “multcomp” R package. Different letters indicate significant differences according to multiple comparison ($p < 0.05$). In (B) and (C), relative expression levels of *Rspe_dsx* were compared between sexes using GLM analyses. ** and *** indicates $p < 0.01$ and 0.001 in GLM, respectively.

Figure 5. Evolutionary changes in *dsx* regulations associated with social evolution in termites. The presence of OD1 and OD2, numbers of exons containing coding sequences (†), and regulatory manners of *dsx* homologues were compared among German cockroach, woodroach, and six termite species. “-” and “?” mean “not detected” and “unknown”, respectively. *: Data on *B. germanica* were cited from Wexler et al. ⁹. The tree topology was based on that of Bourguignon et al. ⁵² and Bucek et al. ²⁹.

Figure S1. RPKM of *vfl* orthologue in each sex and in each caste of *R. speratus*, which RNA-seq data were deposited in NCBI (DRA010978, biological triplicates ²¹). The effects of sex, caste, and their interaction on gene expression levels were evaluated using GLM.

Tables

Table 1. Results of blast search for OD1 against the genomes and/or transcriptomes of six termite species and their sister group (woodroach).

	Database	<i>dsx</i>	<i>Dmrt11</i>	<i>Dmrt93</i>	<i>Dmrt99</i>
<i>Reticulitermes speratus</i>	Genome (Rspe OGS1.0) transcriptome	<i>Rspe_dsx</i> (scaffold6: 6307974..6309410)	RS007930	RS006912	RS002870
<i>Zootermopsis nevadensis</i>	Genome (Znev OGS v2.229)	No hit	scaffold668	Znev05388*	Znev16235*
<i>Hodotermopsis sjostedti</i>	Transcriptome	<i>Hsjo_dsx</i> (c35221)	c18070	c38968	No hit
<i>Nastitermes takasagoensis</i>	Transcriptome	<i>Ntak_dsx</i> (G5ZWOJF02FLJ2Z*)	comp25194	comp174542	No hit
<i>Macrotermes natalensis</i>	Genome (Mnat_gene_v1.2.cds), transcriptome	<i>Mnat_dsx</i> (Mnat_08109)	No hit	Mnat_01812	Mnat_08410
<i>Cryptotermes secundus</i>	Genome (Csec_1.0), Transcriptome	<i>Csec_dsx1</i> (XM_023861307), <i>Csec_dsx2</i> (XM_023858380)	XM0238633381	XM0238639211	XM0238500541
<i>Cryptocercus punctulatus</i>	Transcriptome	<i>Cpun_dsx</i> † (Cpun_comp8195_c0_seq1)	comp1991	No hit	No hit

* is also described by Price et al. ⁶

† found by BLAST search using OD2 (but not OD1) sequences of *Blattella dsx* orthologues, followed by 5' RACE

Table 2. Transcription factors with putative binding sites detected in the *R. speratus* genome and their expression differences between sexes.

Gene symbol	Gene ID in <i>R.</i> <i>speratus</i>	FDR* in heads	FDR* in bodies**
<i>Trl</i>	RS014506	1.000	0.176
<i>grh</i>	RS004650	1.000	0.825
<i>Cf2-II</i>	RS004984	1.000	0.778
<i>ara</i>	RS009368	1.000	0.492
<i>Hr46</i>	RS006489	1.000	0.910
<i>br-Z4</i>	RS006084	1.000	0.970
<i>br-Z4</i>	RS006083	1.000	0.628
<i>brk</i>	RS014094	1.000	0.217
<i>br-Z3</i>	RS006084	1.000	0.970
<i>dve</i>	RS006834	1.000	0.836
<i>ttk</i>	RS007158	1.000	0.914
<i>ttk</i>	RS007159	1.000	0.931
<i>pros</i>	RS012493	1.000	0.890
<i>pros</i>	RS012492	1.000	0.217
<i>yfl</i>	RS003882	1.000	0.275×10^{-3}
<i>ap</i>	RS004033	1.000	0.913
<i>ap</i>	RS004477	1.000	0.633
<i>D</i>	RS010190	1.000	0.822
<i>prd</i>	RS008463	1.000	0.967
<i>Hsf</i>	RS003326	1.000	0.406
<i>su(Hw)</i>	RS012357	1.000	0.791
<i>br-Z2</i>	RS006084	1.000	0.970
<i>br-Z2</i>	RS006083	1.000	0.628
<i>Pnt</i>	RS003359	1.000	0.807
<i>sd</i>	RS008357	1.000	0.919

*: FDR based on GLM with sex using edgeR

**Body: the thorax and abdomen

Figure 1

woodroach

termites

Figure 2

Figure 3

Figure 4

Figure 5

Table S1. Primers used for RACE PCRs and subcloning.

species	primer name	primer sequence
<i>Cryptocercus punctulatus</i>	<i>Cpun_dsx</i> OD2 3'RACE	5'-CCCTTCCGCTCATCTATGTTGTGCTTC-3'
<i>Cryptocercus punctulatus</i>	<i>Cpun_dsx</i> OD2 5'RACE	5'-AGCACAACATAGATGAGCGGAAGGGCTT-3'
<i>Cryptocercus punctulatus</i>	<i>Cpun_dsx</i> male-specific exon-R	5'- GGAGGTGGAAGGTCAC TGCTA -3'
<i>Hodotermopsis sjostedti</i>	<i>Hsjo_dsx</i> OD2 3'RACE	5'-GGAAGTGAACCATCTCAAAGGCCACA-3'
<i>Reticulitermes speratus</i>	<i>Rspe_dsx</i> OD2 3'RACE	5'-CCAACCTGTGCCCCGGTGTGCGAAATCATT-3'
<i>Nastitermes takasagoensis</i>	<i>Ntak_dsx</i> OD2 3'RACE	5'-GCCCCGGTGTGCGAATTCATAAGGAACTAA-3'

5 Table S2. Primers used for quantitative RT-PCRs.

species	primer name	prime sequence
<i>Cryptocercus punctulatus</i>	<i>Cpun_dsx</i> qPCR-F	5'- TGTGCTTCAGGTCTCACAATCG -3'
<i>Cryptocercus punctulatus</i>	<i>Cpun_dsx</i> qPCR-Rf	5'- TGAGTTCATAGCAAATGAATACATGTAATGT -3'
<i>Cryptocercus punctulatus</i>	<i>Cpun_dsx</i> qPCR-Rm	5'- CCATGGACCGCAGCTGTT -3'
<i>Cryptocercus punctulatus</i>	<i>Cpun_beta-actin</i> qPCR-F	5'- TTCCTGGGTATGGAATCCTG -3' *
<i>Cryptocercus punctulatus</i>	<i>Cpun_beta-actin</i> qPCR-R	5'- GTGTTGGCGTACAGGTCCTT -3' *
<i>Cryptocercus punctulatus</i>	<i>Cpun_EF-1alpha</i> qPCR-F	5'- CCTGGGCACAGAGATTTCAT -3' *
<i>Cryptocercus punctulatus</i>	<i>Cpun_EF-1alpha</i> qPCR-R	5'- GAACTCTCCCGTACCAGCAG -3' *
<i>Cryptocercus punctulatus</i>	<i>Cpun_NADH-dh</i> qPCR-F	5'- CGGTTTGAAATTGTTTTTATGGA -3' *
<i>Cryptocercus punctulatus</i>	<i>Cpun_NADH-dh</i> qPCR-R	5'- AAATTAGCACCCAACCCTGA -3' *
<i>Reticulitermes speratus</i>	<i>Rspe_dsx</i> qPCR-F	5'- ACCTCCGTACCAAACCAACCTG -3'
<i>Reticulitermes speratus</i>	<i>Rspe_dsx</i> qPCR-R	5'- TTACCGCCCTCCGCTACAAC -3'
<i>Reticulitermes speratus</i>	<i>Rspe_beta-actin</i> qPCR-F	5'- AGCGGGAAATCGTCCGTGA -3'
<i>Reticulitermes speratus</i>	<i>Rspe_beta-actin</i> qPCR-R	5'- CAATGGTGATGACCTGCCCCAT -3'
<i>Reticulitermes speratus</i>	<i>Rspe_EF1-alpha</i> qPCR-F	5'- GGTGATGCGGCTATTGTTAACC -3'
<i>Reticulitermes speratus</i>	<i>Rspe_EF1-alpha</i> qPCR-R	5'- GTGGTGGGAATTCTGAGAAAGATT -3'
<i>Reticulitermes speratus</i>	<i>Rspe_NADH-dh</i> qPCR-F	5'- GCTGGGGGGTTATTTCATCCAT -3'
<i>Reticulitermes speratus</i>	<i>Rspe_NADH-dh</i> qPCR-R	5'- GGCATACCACAAAGGGCAAAA -3'
<i>Reticulitermes speratus</i>	<i>Rspe_GstD1</i> qPCR-F	5'- GCTGTTGGTGTGGATTGAA -3'
<i>Reticulitermes speratus</i>	<i>Rspe_GstD1</i> qPCR-R	5'- GTATGCTGCGGGTTCATCTT -3'
<i>Reticulitermes speratus</i>	<i>Rspe_EIF-1</i> qPCR-F	5'- ATGGTAGGCTTGAAGCGATG -3'
<i>Reticulitermes speratus</i>	<i>Rspe_EIF-1</i> qPCR-R	5'- TTTGCATCCTGGTAGTCACG -3'
<i>Reticulitermes speratus</i>	<i>Rspe_RPS18</i> qPCR-F	5'- ACTCTCAGCTCACATCCAGT -3'
<i>Reticulitermes speratus</i>	<i>Rspe_RPS18</i> qPCR-R	5'- CCTCAGGCCCAATAATGTC -3'
<i>Nastitermes takasagoensis</i>	<i>Ntak_dsx</i> qPCR-F	5'- TAAAGGCCGTACCACACCAA -3'
<i>Nastitermes takasagoensis</i>	<i>Ntak_dsx</i> qPCR-R	5'- ACAATCGCGACAATGACACT -3'
<i>Nastitermes takasagoensis</i>	<i>Ntak_beta-actin</i> qPCR-F	5'- AGCGGGAAATCGTACGTGAC -3' †
<i>Nastitermes takasagoensis</i>	<i>Ntak_beta-actin</i> qPCR-R	5'- CAATGGTGATGACCTGGCCAT -3' †
<i>Nastitermes takasagoensis</i>	<i>Ntak_EF1-alpha</i> qPCR-F	5'- GGTGATGCGCTATTGTTAACC -3' †
<i>Nastitermes takasagoensis</i>	<i>Ntak_EF1-alpha</i> qPCR-R	5'- GTGGTGGGAATTCTGAGAAAGATT -3' †
<i>Nastitermes takasagoensis</i>	<i>Ntak_NADH-dh</i> qPCR-F	5'- GCTGGGGGCGTTATTTCATTCTA -3' †
<i>Nastitermes takasagoensis</i>	<i>Ntak_NADH-dh</i> qPCR-R	5'- GGCATGCCACAAAGAGCAAAA -3' †

6 * from Masuoka et al.³⁶, † from Hojo et al.⁵¹

Table S3. Information of OTU used for phylogenetic analysis.

Clade	Gene	Taxon	Organism	DNA accession No.
dsx	<i>dsx</i>	Blattodea	<i>Blattella germanica</i>	PYGN01000615
	<i>Cpun_dsx</i>	Blattodea	<i>Cryptocercus punctulatus</i>	*
	<i>Csec_dsx1</i>	Isoptera	<i>Cryptotermes secundus</i>	XM_023861307
	<i>Csec_dsx2</i>	Isoptera	<i>Cryptotermes secundus</i>	XM_023858380
	<i>Hsjo_dsx</i>	Isoptera	<i>Hodotermopsis sjostedti</i>	Hsjo_m.41619, c35221
	<i>Mnat_dsx</i>	Isoptera	<i>Macrotermes natalensis</i>	Mnat_08109
	<i>Niak_dsx</i>	Isoptera	<i>Nasutitermes takasagoensis</i>	G5ZWOJF02FLJ2Z
	<i>Rspe_dsx</i>	Isoptera	<i>Reticulitermes speratus</i>	*
	<i>dsx1a</i>	Crustacea: Diplostraca	<i>Daphnia magna</i>	AB569296
	<i>dsx1b</i>	Crustacea: Diplostraca	<i>Daphnia magna</i>	AB569297
	<i>dsx2</i>	Crustacea: Diplostraca	<i>Daphnia magna</i>	AB569298
	<i>dsx</i>	Coleoptera	<i>Tribolium castaneum</i>	XM_966683
	<i>dsx</i>	Hymenoptera	<i>Apis florea</i>	XM_003691527
	<i>dsx</i>	Hymenoptera	<i>Apis mellifera</i>	NM_001111255
	<i>dsx</i>	Hymenoptera	<i>Bombus impatiens</i>	XM_003489174
	<i>dsx</i>	Hymenoptera	<i>Bombus terrestris</i>	XM_003395718
	<i>dsx</i>	Hymenoptera	<i>Nasonia vitripennis</i>	XM_008207201
	<i>dsx</i>	Lepidoptera	<i>Bombyx mori</i>	NM_001111345
	<i>dsx</i>	Lepidoptera	<i>Ostrinia scapularis</i>	AB548678
	<i>dsx</i>	Diptera	<i>Aedes aegypti</i>	DQ440534
	<i>dsx</i>	Diptera	<i>Anastrepha fraterculus</i>	DQ494334
	<i>dsx</i>	Diptera	<i>Anopheles obliqua</i>	AY948421
	<i>dsx</i>	Diptera	<i>Anopheles gambiae</i>	XM_560052
	<i>dsx</i>	Diptera	<i>Antheraea assama</i>	GU930278
	<i>dsx</i>	Diptera	<i>Bactrocera oleae</i>	AJ547622
	<i>dsx</i>	Diptera	<i>Drosophila melanogaster</i>	NM_169202
	<i>dsx</i>	Diptera	<i>Musca domestica</i>	AY461854
DMRT11	<i>DMRT11E</i>	Crustacea: Diplostraca	<i>Daphnia magna</i>	AB361069
	<i>DMRT11B</i>	Blattodea	<i>Blattella germanica</i>	Bger_genome_scaffold:Scaffold1935:103730-103596*
	<i>DMRT11B</i>	Blattodea	<i>Cryptocercus punctulatus</i>	Cpun_m.20374, comp1991
	<i>DMRT11B</i>	Isoptera	<i>Hodotermopsis sjostedti</i>	Hsjo_m.15983, c18070
	<i>DMRT11B</i>	Isoptera	<i>Zootermopsis nevadensis</i>	scaffold668:816294-816428*
	<i>DMRT11B</i>	Isoptera	<i>Nasutitermes takasagoensis</i>	TR57579, comp25194
	<i>DMRT11B</i>	Isoptera	<i>Reticulitermes speratus</i>	RS007930
	<i>DMRT11E</i>	Diptera	<i>Drosophila melanogaster</i>	NM_078591
	<i>DMRT11E</i>	Diptera	<i>Drosophila pseudoobscura</i>	XM_001355494
DMRT93	<i>DMRT93B</i>	Crustacea: Diplostraca	<i>Daphnia magna</i>	AB361070
	<i>DMRT93B</i>	Blattodea	<i>Blattella germanica</i>	Bger_genome_scaffold:Scaffold353:1209444-1209310*
	<i>DMRT93B</i>	Isoptera	<i>Hodotermopsis sjostedti</i>	Hsjo_m.51385, c38968
	<i>DMRT93B</i>	Isoptera	<i>Macrotermes natalensis</i>	Mnat_01812
	<i>DMRT93B</i>	Isoptera	<i>Nasutitermes takasagoensis</i>	comp174542
	<i>DMRT93B</i>	Isoptera	<i>Reticulitermes speratus</i>	RS006912
	<i>DMRT93B</i>	Isoptera	<i>Zootermopsis nevadensis</i>	Znev_05388
	<i>DMRT93B</i>	Hemiptera	<i>Acyrtosiphon pisum</i>	XM_001950295
	<i>DMRT93B</i>	Coleoptera	<i>Tribolium castaneum</i>	XM_966511
	<i>DMRT93B</i>	Hymenoptera	<i>Apis mellifera</i>	XM_392966
	<i>DMRT93B</i>	Hymenoptera	<i>Bombus terrestris</i>	XM_003403144
	<i>DMRT93B</i>	Hymenoptera	<i>Nasonia vitripennis</i>	XM_001603289
	<i>DMRT93B</i>	Diptera	<i>Aedes aegypti</i>	XM_001649562
	<i>DMRT93B</i>	Diptera	<i>Anopheles gambiae</i>	XM_321748
	<i>DMRT93B</i>	Diptera	<i>Drosophila melanogaster</i>	NM_079704
	<i>DMRT93B</i>	Diptera	<i>Drosophila pseudoobscura</i>	XM_001360022
DMRT99	<i>DMRT99B</i>	Crustacea: Diplostraca	<i>Daphnia magna</i>	AB361071
	<i>DMRT99B</i>	Blattodea	<i>Blattella germanica</i>	Bger_genome_scaffold:Scaffold1558:10787-10653 *
	<i>DMRT99B</i>	Isoptera	<i>Cryptotermes secundus</i>	XM_0238500541
	<i>DMRT99B</i>	Isoptera	<i>Macrotermes natalensis</i>	Mnat_08410
	<i>DMRT99B</i>	Isoptera	<i>Reticulitermes speratus</i>	RS002870
	<i>DMRT99B</i>	Isoptera	<i>Zootermopsis nevadensis</i>	Znev_16235
	<i>DMRT99B</i>	Hemiptera	<i>Acyrtosiphon pisum</i>	XM_001949304
	<i>DMRT99B</i>	Coleoptera	<i>Tribolium castaneum</i>	XM_970582
	<i>DMRT99B</i>	Hymenoptera	<i>Apis florea</i>	XM_012486429
	<i>DMRT99B</i>	Hymenoptera	<i>Bombus terrestris</i>	XM_003396307
	<i>DMRT99B</i>	Diptera	<i>Anopheles gambiae</i>	XM_310668
	<i>DMRT99B</i>	Diptera	<i>Drosophila melanogaster</i>	NM_079704
	<i>DMRT99B</i>	Diptera	<i>Drosophila pseudoobscura</i>	XM_001357729
DMRT1	<i>DMRT1</i>	Vertebrate: Primate	<i>Homo sapiens</i>	AY442914
	<i>DMRT1</i>	Vertebrate: Anura	<i>Xenopus laevis</i>	NM_001096500
	<i>DMRT1</i>	Vertebrate: Beloniformes	<i>Oryzias latipes</i>	AF319994
	<i>DMY</i>	Vertebrate: Beloniformes	<i>Oryzias latipes</i>	AB071534
	<i>DMRT1</i>	Vertebrate: Tetraodontiformes	<i>Takifugu rubripes</i>	NM_001037949
DMRT2	<i>DMRT2</i>	Vertebrate: Primate	<i>Homo sapiens</i>	NM_006557
	<i>DMRT2</i>	Vertebrate: Beloniformes	<i>Oryzias latipes</i>	AF319992
	<i>DMRT2</i>	Vertebrate: Tetraodontiformes	<i>Takifugu rubripes</i>	NM_001037946
DMRT3	<i>DMRT3</i>	Vertebrate: Primate	<i>Homo sapiens</i>	NM_021240
	<i>DMRT3</i>	Vertebrate: Tetraodontiformes	<i>Takifugu rubripes</i>	NM_001037946
	<i>DMRT3</i>	Vertebrate: Cypriniformes	<i>Danio rerio</i>	AY621083
DMRT4	<i>DMRT41</i>	Vertebrate: Primate	<i>Homo sapiens</i>	NM_022160
	<i>DMRT4</i>	Vertebrate: Anura	<i>Xenopus laevis</i>	AY648303
	<i>DMRT4</i>	Vertebrate: Beloniformes	<i>Oryzias latipes</i>	AB055958
	<i>DMRT4</i>	Vertebrate: Tetraodontiformes	<i>Takifugu rubripes</i>	NM_001037948
DMRT5	<i>DMRT42</i>	Vertebrate: Primate	<i>Homo sapiens</i>	NM_032110
	<i>DMRT5</i>	Vertebrate: Anura	<i>Xenopus laevis</i>	DQ329358
	<i>DMRT5</i>	Vertebrate: Beloniformes	<i>Oryzias latipes</i>	AB083691
	<i>DMRT5</i>	Vertebrate: Tetraodontiformes	<i>Takifugu rubripes</i>	NM_001037950
DMRT7	<i>DMRT7</i>	Vertebrate: Glires	<i>Mus musculus</i>	NM_027732
	<i>DMRT7</i>	Vertebrate: Cetartiodactyla	<i>Bos taurus</i>	NM_001038182

*not yet registered to the NCBI

7 Table S4. Stability values of internal control gene candidates of *R. speratus* using GeNorm and NormFinder.

	GeNorm (stability value)	NormFinder (stability value)
<i>Rspe_beta-actin</i>	0.385	0.264
<i>Rspe_EF1-alpha</i>	0.246	0.106
<i>Rspe_NADH-dh*</i>	0.148	0.051
<i>Rspe_GstD1</i>	0.158	0.089
<i>Rspe EIF-1</i>	0.148	0.061
<i>Rspe_RPS18</i>	0.161	0.093

8 * *Rspe_NADH-dh* was selected by GeNorm and NormFinder due to the lowest stability values.

9

10

11 Table S5. Stability values of internal control gene candidates of *C. punctulatus* using GeNorm and NormFinder.

	GeNorm (stability value)	NormFinder (stability value)
<i>Cpun_beta-actin*</i>	2.365	0.621
<i>Cpun_EF1-alpha</i>	2.665	1.306
<i>Cpun_NADH-dh</i>	3.237	2.075

12 **Cpun_beta-actin* was selected by GeNorm and NormFinder due to the lowest stability values.

13

14

15 Table S6. Stability values of internal control gene candidates of *N. takasagoensis* using GeNorm and NormFinder.

	GeNorm (stability value)	NormFinder (stability value)
<i>Ntak_beta-actin</i>	0.908	0.505
<i>Ntak_EF1-alpha*</i>	0.757	0.237
<i>Ntak_NADH-dh</i>	0.983	0.603

16 **Ntak_EF1-alpha* was selected by GeNorm and NormFinder due to the lowest stability values.

17

18

Figure S1