Hemimetabolous genomes reveal molecular basis of termite eusociality

Mark C. Harrison, Evelien Jongepier, Hugh M. Robertson, Nicolas Arning, Tristan Bitard-Feildel, Hsu Chao, Christopher P. Childers, Huyen Dinh, Harshavardhan Doddapaneni, Shannon Dugan, Johannes Gowin, Carolin Greiner, Yi Han, Haofu Hu, Daniel S. T. Hughes, Ann-Kathrin Huylmans, Carsten Kemen, Lukas P. M. Kremer, Sandra L. Lee, Alberto Lopez-Ezquerra, Ludovic Mallet, Jose M. Monroy-Kuhn, Annabell Moser, Shwetha C. Murali, Donna M. Muzny, Saria Otani, Maria-Dolors Piulachs, Monica Poelchau, Jiaxin Qu, Florentine Schaub, Ayako Wada-Katsumata, Kim C. Worley, Qiaolin Xie, Guillem Ylla, Michael Poulson, Richard A. Gibbs, Coby Schal, Stephen Richards, Xavier Belles, Judith Korb and Erich Bornberg-Bauer

Around 150 million years ago, eusocial termites evolved within the cockroaches, 50 million years before eusocial Hymenoptera, such as bees and ants, appeared. Here, we report the 2-Gb genome of the German cockroach, *Blattella germanica*, and the 1.3-Gb genome of the drywood termite *Cryptotermes secundus*. We show evolutionary signatures of termite eusociality by comparing the genomes and transcriptomes of three termites and the cockroach against the background of 16 other eusocial and non-eusocial insects. Dramatic adaptive changes in genes underlying the production and perception of pheromones confirms the importance of chemical communication in the termites. These are accompanied by major changes in gene regulation and the molecular evolution of caste determination. Many of these results parallel molecular mechanisms of eusocial evolution in Hymenoptera. However, the specific solutions are remarkably different, thus revealing a striking case of convergence in one of the major evolutionary transitions in biological complexity.

Eusociality, the reproductive division of labour with overlapping generations and cooperative brood care, is one of the major evolutionary transitions in biology. Although rare, eusociality has been observed in a diverse range of organisms, including shrimps, mole rats and several insect lineages. A particularly striking case of convergent evolution occurred within the hemimetabolous Hymenoptera and in the hemimetabolous termites (Isoptera), which are separated by over 350 Myr of evolution. Termites evolved within the cockroaches around 150 Myr ago, towards the end of the Jurassic period, about 50 Myr before the first bees and ants appeared. Therefore, identifying the molecular mechanisms common to both origins of eusociality is crucial to understanding the fundamental signatures of these rare evolutionary transitions. While the availability of genomes from many eusocial and non-eusocial hymenopteran species has allowed extensive research into the origins of eusociality within ants and bees, a paucity of genomic data from cockroaches and termites has precluded large-scale investigations into the evolution of eusociality in this hemimetabolous clade.

The conditions under which eusociality arose differ greatly between the two groups. Termites and cockroaches are hemimetabolous and so show a direct development, while holometabolous hymenopterans complete the adult body plan during metamorphosis. In termites, workers are immatures and only reproductive castes are adults, while in Hymenoptera, adult workers and queens represent the primary division of labour. Moreover, termites are diploid and their colonies consist of both male and female workers, and usually a queen and king dominate reproduction. This is in contrast to the haplodiploid system found in Hymenoptera, in which all workers and dominant reproductives are female. It is therefore intriguing that strong similarities have evolved convergently within the termites and the hymenopterans, such as differentiated castes and a nest life with reproductive division of labour. The termites can be subdivided into wood-dwelling and foraging termites. The former belong to the lower termites and produce simple, small colonies with totipotent workers that can become reproductives. Foraging termites (some lower and all higher termites) form large, complex societies, in which worker castes can be irreversible. For this reason, higher, but not lower, termites can be classed as superorganismal. Similarly, within Hymenoptera, varying levels of eusociality exist.

Here we provide insights into the molecular signatures of eusociality within the termites. We analysed the genomes of two lower and one higher termite species and compared them to the genome of the German cockroach.
of the German cockroach, Blattella germanica, as a closely related non-eusocial outgroup. Furthermore, differences in expression between nymphs and adults of the cockroach were compared to differences in expression between workers and reproducitives of the three termites, to gain insights into how expression patterns changed along with the evolution of castes. Using 15 additional insect genomes to infer background gene family turnover rates, we analysed the evolution of gene families along the transition from non-social cockroaches to eusociality in the termites. In this study, we concentrated particularly on two hallmarks of insect eusociality, as previously described for Hymenoptera, with the expectation that similar patterns occurred along with the emergence of termites. These are the evolution of a sophisticated chemical communication, which is essential for the functioning of a eusocial insect colony\(^1,13,16\), and major changes in gene regulation along with the evolution of castes\(^2,9\). We also tested whether transposable elements spurred the evolution of gene families that were essential for the transition to eusociality.

**Evolution of genomes, proteomes and transcriptomes**

We sequenced and assembled the genome of the German cockroach, B. germanica (Ectobiidae), and of the lower, drywood termite Cryptotermes secundus (Kalotermitidae; for assembly statistics, see Supplementary Table 1). The cockroach genome (2.0 Gb) is considerably larger than all three termite genomes. The genome size of C. secundus (1.30 Gb) is comparable to the higher, fungus-growing termite Macrotermes natalensis (1.31 Gb, Termitidae)\(^3\), but more than twice as large as the lower, dampwood termite Zootermopsis nevadensis (562 Mb, Termopsidae)\(^4\). The smaller genomes of termites compared to the cockroach are in line with previous size estimations based on C-values\(^5\). The proteome of B. germanica (29,216 proteins) is also much larger than in the termites, where we find the proteome size in C. secundus (18,162) to be similar to those of the other two termites (M. natalensis: 16,140; Z. nevadensis: 15,459; Fig. 1). In fact, the B. germanica proteome was the largest among all 21 arthropod species analysed here (Fig. 1). Strong evidence of the higher number of genes, both coding and non-coding, is present in B. germanica, although the first IR group for signs of species-specific positive selection. Within the Blattodea-specific intronless IRs, we found two codon positions under significant selection for the higher termite M. natalensis (codeml site models 7 and 8; P = 5.4 × 10\(^{-5}\)). Within a subgroup of five sequences, this was even more strongly pronounced with seven codon positions under significant positive selection for M. natalensis (P < 1.7 × 10\(^{-10}\)). The positively evolving codons are situated within the two ligand-binding lobes of the receptors (Fig. 3c), showing that a diversification of ligand specificity has occurred along with the transition to higher eusociality and a change from wood-feeding to fungus-farming in M. natalensis. Only two IRs were differentially expressed between nymphs and adult females in B. germanica. Underlining a change in expression along with the evolution of castes, we found 35 IRs to be differentially expressed between workers and queens in Z. nevadensis, 11 in C. secundus and 10 in M. natalensis (Fig. 2 and Supplementary Table 10). The possible

The termite and cockroach genomes contain a higher level of repetitive DNA compared to the hymenopterans we analysed (Fig. 1). C. secundus and B. germanica genomes both contain 55% repetitive content (Supplementary Table 7), which is higher than in both Z. nevadensis (28%) and the higher termite M. natalensis (46%; Fig. 1)\(^6\). As also found in Z. nevadensis and M. natalensis\(^6\), LINEs and especially the subfamily BovB were the most abundant transposable elements (TEs) in the B. germanica and C. secundus genomes, indicating that a proliferation of LINEs may have occurred in the ancestors of Blattodea (cockroaches and termites).

We hypothesized that these high levels of TEs may be driving the high turnover in gene family sizes within the termites and B. germanica\(^7\). Expanded gene families indeed had more repetitive content within 10-kb flanking regions in all three termites (P < 1.3 × 10\(^{-4}\); Wald t-test; Supplementary Tables 8 and 9), in particular in the higher termite M. natalensis. In contrast, gene family expansions were not correlated with TE content in flanking regions for B. germanica. These results suggest that a major expansion of LINEs at the root of the Blattodea clade contributed to the evolution of gene families within termites, probably via unequal crossing-over\(^8\); however, the expansions in B. germanica were not facilitated by TEs. It can therefore be speculated that the large expansion of LINEs within Blattodea allowed the evolution of gene families that ultimately facilitated the transition to eusociality.

**Expansion and positive selection of ionotropic receptors**

Insects perceive chemical cues from toxins, pathogens, food and pheromones with three major families of chemoreceptors, the odorant (ORs), gustatory and ionotropic (IRs) receptors\(^9\). ORs, in particular, have been linked to colony communication in eusocial Hymenoptera, where they abound\(^8,10\). Interestingly, as previously detected for Z. nevadensis\(^4\), the OR repertoire is substantially smaller in B. germanica and all three termites compared to hymenopterans. IRs, on the other hand, which are less frequent in hymenopterans, are strongly expanded in the cockroach and termite genomes (Fig. 3 and Supplementary Fig. 3). Intronless IRs, which are known to be particularly divergent\(^7\), show the greatest cockroach- and Blattodea-specific expansions (Fig. 3a, B. germanica, Cockroach- and Group D-IRs). By far the most IRs among all investigated species were found in B. germanica (455 complete gene models), underlining that the capacity for detecting many different kinds of chemosensory cues is crucial for this generalist that thrives in challenging, human environments. In line with a specialization in diet and habitat, the total number of IRs is lower within M. natalensis (141; C. secundus: 135; M. natalensis: 75). Nevertheless, IRs are more numerous in termites than in all other analysed species (except Nasonia vitripennis: 111). This is strikingly similar to the pattern for ORs in Hymenoptera, which are also highly numerous in non-eusocial outgroups as well as in eusocial sister species\(^11,12,23,25\).

We scanned each IR group for signs of species-specific positive selection. Within the Blattodea-specific intronless IRs, we found two codon positions under significant selection for the higher termite M. natalensis (codeml site models 7 and 8; P = 5.4 × 10\(^{-5}\)). Within a subgroup of five sequences, this was even more strongly pronounced with seven codon positions under significant positive selection for M. natalensis (P < 1.7 × 10\(^{-10}\)). The positively evolving codons are situated within the two ligand-binding lobes of the receptors (Fig. 3c), showing that a diversification of ligand specificity has occurred along with the transition to higher eusociality and a change from wood-feeding to fungus-farming in M. natalensis. Only two IRs were differentially expressed between nymphs and adult females in B. germanica. Underlining a change in expression along with the evolution of castes, we found 35 IRs to be differentially expressed between workers and queens in Z. nevadensis, 11 in C. secundus and 10 in M. natalensis (Fig. 2 and Supplementary Table 10). The possible
role of IRs in pheromonal communication has been highlighted both in the cockroach *Periplaneta americana* and in *Drosophila melanogaster*, where several IRs show sex-biased expression.

One group of ORs (orange clade in Fig. 3b) is evolving under significant positive selection at codon positions within the second transmembrane domain in *M. natalensis* (*codeml* site model; \( P = 1.1 \times 10^{-11} \)) and *C. secundus* (*P = 5.6 \times 10^{-16} \); Fig. 3d). Such a variation in the transmembrane domain can be related to ligand-binding specificity, as has been shown for a polymorphism in the third transmembrane domain in *D. melanogaster*. Spatial variation in the transmembrane domain for an OR in *D. melanogaster* adds further support for an adaptive evolution of chemoreceptors, in line with the greater need for a sophisticated colony communication in the termites.

Similar to IRs, a higher proportion of ORs were differentially expressed between workers and queens in the three termites compared to ants, where these genes are under strong purifying selection, possibly related to their greater diversification of worker castes (major and minor workers, major and minor soldiers). Although desaturases are often discussed in the context of CHC production and chemical communication, their biochemical roles are quite diverse, and the positive selection we observe for *M. natalensis* may, at least in part, be related to their rather different ecology of foraging and fungus-farming rather than nest-mate recognition.

**CHC-producing enzymes have evolved caste-specificity**

Despite their different ancestry, both termites and eusocial hymenopterans are characterized by the production of caste-specific cuticular hydrocarbons (CHCs), which are often crucial for regulating reproductive division of labour and chemical communication. Accordingly, we find changes in the termites in three groups of proteins involved in the synthesis of CHCs: desaturases (introduction of double bonds), elongases (extension of C-chain length), and CYP4G1 (last step of CHC biosynthesis).

Desaturases are thought to be important for division of labour and social communication in ants. As previously described for ants, Desat B genes are the most abundant desaturase family in the termites and the cockroach (Supplementary Table 12), especially in *M. natalensis* where we found ten gene copies (significant expansion; \( P = 0.0003 \); Supplementary Table 5 and Supplementary Figure 4).

In contrast to ants, where these genes are under strong purifying selection, for the highly eusocial termite *M. natalensis*, we found significant positive selection within the Desat B genes (*codeml* site models 7 and 8; \( P = 1.1 \times 10^{-16} \)), indicating a diversification in function, possibly related to their greater diversification of worker castes (major and minor workers, major and minor soldiers). Although desaturases are often discussed in the context of CHC production and chemical communication, their biochemical roles are quite diverse, and the positive selection we observe for *M. natalensis* may, at least in part, be related to their rather different ecology of foraging and fungus-farming rather than nest-mate recognition.

Future experimental verification of the function of these genes will help better understand these observed genomic and transcriptomic patterns.
Underlining an increased importance of CHC communication in termites, the expression patterns of elongases (extension of C-chain length) differ considerably in the termites compared to the cockroach (Fig. 2 and Supplementary Table 14). In contrast to B. germanica, in which elongases are both nymph- (five genes) and adult-biased (four genes), only one or two elongase genes in each termite are queen-biased in their expression, while many are worker-biased. As with the desaturases, a group of M. natalensis elongases also reveal significant signals of positive selection (codeml branch-site test; $P = 4 \times 10^{-4}$), further indicating a greater diversification of CHC production in this higher termite.

The last step of CHC biosynthesis, the production of hydrocarbons from long-chain fatty aldehydes, is catalysed by a P450 gene, CYP4G1, in D. melanogaster. We found one copy of CYP4G1 in B. germanica, Z. nevadensis and C. secundus, but three copies in M. natalensis, reinforcing the greater importance of CHC synthesis in this higher termite. Corroborating the known importance of maternal CHCs in B. germanica, CYP4G1 is overexpressed in female adults compared to nymphs (Fig. 2 and Supplementary Table 15). In each of the termites, however, CYP4G1 is more highly expressed in workers (or kings in C. secundus) compared to queens (Fig. 2 and Supplementary Table 15), adding support that, compared to cockroach nymphs, a change in the dynamics and turnover of CHCs in termite workers has taken place.

**Changes in gene regulation in termites**

The development of distinct castes underlying division of labour is achieved via differential gene expression. Major changes in gene regulation have been reported as being central to the transition to eusociality in bees and ants. Accordingly, we found major changes in putative DNA methylation patterns (levels per 1-to-1 orthologue) among the termites compared to four other hemimetabolous insect species (Fig. 4a). This is revealed by CpG depletion patterns (CpGdepletion observed versus expected number of CpGs), a reliable predictor of DNA methylation, correlating more strongly between the termites than among any of the other analysed hemimetabolous insects (Fig. 4). In other words, within orthologous genes, predicted DNA methylation levels differ greatly between termites and other hemimetabolous species but remain conserved among termite species.

The predicted levels of DNA methylation correlated negatively with the caste-specificity of expression for each of the termites. This is confirmed by a positive correlation between Cpgsat (negative association with level of DNA methylation) and absolute log$_2$(fold change) of expression between queens and workers (Pearson’s $r = 0.32$ to $0.36; P < 2.2 \times 10^{-16}$). The caste-specific expression of putatively unmethylated genes in termites is reflected in the enrichment of GO terms related to sensory perception, regulation of transcription, signalling and development, whereas methylated genes are mainly related to general metabolic processes (Fig. 4b and Supplementary Table 16). These results show strong parallels to findings for eusocial Hymenoptera, reinforcing the very weak relationship between CHCs and queen versus worker expression between nymphs and adult females in B. germanica (Fig. 4b).

Our results argue in favour of a diminished role of DNA methylation in caste-specific expression within eusocial insects,
as recently shown\textsuperscript{38,44}. In fact, DNA methylation appears to be important for the regulation of housekeeping genes because predicted methylated genes are related to general biological processes (further supported by lower CpG\textsubscript{o/e} within 1-to-1 orthologues than in non-conserved genes)\textsuperscript{45}, while caste-specific genes are ‘released’ from this type of gene regulation. However, a recent study linked caste-specific DNA methylation to alternative splicing in Z. nevadensis\textsuperscript{46}.

Fig. 3 | Expansions, contractions and positive selection within iRs and ORs in termites. \textbf{a}, IR (a) and OR (b) gene trees of 13 insect species. In each tree, only well-supported clades (support values > 85) that include \textit{B. germanica} or termite genes are highlighted within the gene trees. The lengths of the coloured bars represent the number of genes per species within each of these clades. The red asterisk in \textit{a} denotes the putative root of intronless iRs. \textbf{c}, The upper schematic diagram depicts the 2D structure of an IR, containing ligand-binding lobes (S1 and S2), transmembrane regions (TM1-3) and the pore domain (P). Below, the sequence of the domains along the peptide is represented, showing that the sites, which are under significant positive selection (red bars; codeml site models 7 and 8) within Blattodea iRs for \textit{M. natalensis} (P < 1.7 x 10\textsuperscript{-10}; likelihood-ratio test, 5 sequences, 413 aligned codons), are all situated within the ligand-binding lobes and on or around the putative ligand-binding sites (asterisks)\textsuperscript{86}. \textbf{d}, The same representation for ORs, which include eight transmembrane regions. Positive selection was found for \textit{M. natalensis} (P = 1.1 x 10\textsuperscript{-10}; 5 sequences, 1,001 aligned codons) and \textit{C. secundus} (P = 5.6 x 10\textsuperscript{-16}; likelihood ratio test, 26 sequences, 1,913 aligned codons) of the orange clade, each at two codon positions within the second transmembrane region and at a third position within the carboxy-terminal extracellular region for \textit{M. natalensis}. 
**Fig. 4** CpG\(_{\text{ab}}\) of seven hemimetabolous insects. a. Principal component analysis (PCA) of predicted DNA methylation patterns among 2,664 1-to-1 orthologues, estimated via CpG\(_{\text{ab}}\). The spheres represent the positions of the species within the 3D PCA, with the distance between the spheres representing the similarity of CpG\(_{\text{ab}}\) between species at each orthologue; the curves are the distribution of CpG\(_{\text{ab}}\), with the dotted line showing CpG\(_{\text{ab}}\) enrichment. Number of enriched genes and total number of genes in topGO enrichment tests (low CpG\(_{\text{ab}}\) /high CpG\(_{\text{ab}}\) gene universe): B. germanica (3,291/1,842/11,409); termites (6,754/4,600/25,910). High CpG\(_{\text{ab}}\) indicates a low level of DNA methylation and vice versa.

Major biological transitions are often accompanied by expansions of transcription factor (TF) families, such as genes containing zinc-finger (ZF) domains\(^{17}\). We also observed large differences in ZF families within the termites compared to B. germanica. Many ZF families were reduced or absent in termites, while different, unrelated ZF gene families were significantly expanded (Supplementary Tables 2–6). Queen-biased genes were significantly over-represented among ZF genes for each of the termites \((P < 2 \times 10^{-10}, \chi^2 \text{ test}; \text{Supplementary Table 17})\), indicating an important role of ZF genes in the regulation of genes related to caste-specific tasks and colony organization in the termites. This is in contrast to the significant under-representation of differentially expressed ZF genes within B. germanica \((P = 4.8 \times 10^{-5}, \chi^2 \text{ test})\). Interestingly, two other important TF families (bHLH and bZIP)\(^{15}\), which were not expanded in the termites, showed no caste-specific expression pattern \((P > 0.05)\), except bZIP genes, in which queen-biased genes were marginally over-represented for M. natalensis \((P = 0.049)\). These major upheavals in ZF gene families and their caste-specific expression show that major changes in TFs accompanied the evolution of termites, strikingly similar to the evolution of ants\(^{10}\).

**Evolution of genes related to moulting and metamorphosis**

Hemimetabolous eusociality is characterized by differentiated castes, which represent different developmental stages. This is in contrast to eusocial Hymenoptera, in which workers and reproductive are adults. While cockroaches develop directly through several nymphal stages before becoming reproductive adults, termite development is more phenotypically plastic, and workers are essentially immatures (Fig. 2). In wood-dwelling termites, such as C. secundus and Z. nevadensis, worker castes are non-reproductive immatures
that are potiotopent to develop into other castes, while in the higher termite *M. natalensis*, workers can be irreversibly defined instars. It is therefore clear that a major change during the evolution of termites occurred within developmental pathways. Accordingly, we found changes in expression and gene family size of several genes related both to moultung and metamorphosis.

In the synthesis of the moultung hormone, 20-hydroxyecdysone, the six Halloween genes (five cytochrome P450s and a Rieske-domain oxygenase) play a key role. Only one Halloween gene, Shade (Shd; CYP314A1), which mediates the final step of 20-hydroxyecdysone synthesis, is differentially expressed between the final nymphal stages and adult females in *B. germanica* (Fig. 2 and Supplementary Table 18), consistent with its role in the nymphal or imaginal moult. In the three termites, the Halloween genes show varying caste-specific expression (Fig. 2 and Supplementary Table 18), showing that ecdysone plays a significant role in the regulation of caste differences. Ecdysteroid kinase genes (ECK), which convert the insect moultung hormone into its inactive state, ecdysone 22-phosphate, for storage, are only overexpressed in female adults compared to nymphs in *B. germanica* (16/51 genes, Fig. 2 and Supplementary Table 19). In termites, however, where the gene copy number is reduced (18 to 20 per species), these important moultung genes appear to have evolved worker-specific functions (Fig. 2 and Supplementary Table 19).

Whereas 20-hydroxyecdysone promotes moultung, juvenile hormone (JH) represses imaginal development in pre-adult instars. JH is important in caste differentiation in eusocial insects, including termites. Haemolymph JH-binding proteins (JHBPs), which transport JH to its target tissues, are reduced within the termites (21 to 33 genes) but significantly expanded in *B. germanica* (51 copies; P = 0.0018; Supplementary Table 6). Thirteen of the JHBP genes are overexpressed in adult females and only 8 in nymphs in *B. germanica* (Fig. 2 and Supplementary Table 20). In *Z. nevadensis* and *M. natalensis*, on the other hand, JHBPs are significantly more worker-biased (*P < 0.01*, χ² test; Supplementary Table 20 and Fig. 2).

In *C. secundus*, expression is more varied, with four worker-biased, seven king-biased and two queen-biased genes (Fig. 2 and Supplementary Table 20). These changes in copy number and caste-specific expression of genes involved in moultung and metamorphosis within termites compared to the German cockroach demonstrate that changes occurred in the control of the developmental pathway along with the evolution of castes. However, this interpretation needs to be experimentally verified.

**Conclusions**

These results, considered alongside many studies on eusociality in Hymenoptera, provide evidence that major changes in gene regulation and the evolution of sophisticated chemical communication are fundamental to the transition to eusociality in insects. Strong changes in DNA methylation patterns correlated with broad-scale modifications of expression patterns. Many of these modified expression patterns remained consistent among the three studied termite species and occurred within protein pathways essential for eusocial life, such as CHC production, chemosensory receptors, cytochrome P450s and genes involved in the synthesis of the moulting hormone, 20-hydroxyecdysone.

Of particular interest is the regulation of genes involved in the synthesis of 20-hydroxyecdysone and juvenile hormone. The Halloween genes, which play a crucial role in the synthesis of 20-hydroxyecdysone, have undergone significant expansion in *B. germanica* compared to other termite species. This expansion is consistent with the higher level of eusociality in *B. germanica* compared to the other studied species. Other genes involved in the synthesis of 20-hydroxyecdysone and juvenile hormone, such as the JH-binding proteins (JHBPs), have also undergone changes in expression and gene family size. These changes in expression and gene family size are consistent with the higher level of eusociality in *B. germanica* compared to the other studied species.

The above results suggest that the evolution of eusociality in termites is a complex process involving changes in gene regulation and the evolution of sophisticated chemical communication. These changes are consistent with the transition to eusociality in insects and provide evidence for the evolution of eusocial life, such as CHC production, chemosensory receptors, cytochrome P450s and genes involved in the synthesis of the moulting hormone, 20-hydroxyecdysone.

**Methods**

**Genome sequencing and assembly.** Genomic DNA from a single *Blattella germanica* male from an inbred line (strain: American Cyanamid=Orlando Normal) was used to construct two paired-end (180-bp and 500-bp inserts) and one of the two mate-pair libraries (2-kb inserts). An 8-kb mate-pair library was constructed from a single female. The libraries were sequenced on an Illumina HiSeq2000 sequencing platform. The 413 Gb of raw sequence data were assembled with Allpaths LG, and then scaffolded and gap-filled using the in-house tools Allpaths LG v1.0 (https://www.hgsc.bcm.edu/software/allpaths-lg) and Atlas gap-fill v2.2. For *Cryptotermes secundus*, three paired-end libraries (250-bp, 500-bp and 800-bp inserts) and three mate-pair libraries (2-kb, 5-kb and 10-kb inserts) were constructed from genomic DNA that was extracted from the head and thorax of 1,000 individuals, originating from a single, inbred field colony. The libraries were sequenced on an Illumina HiSeq2000 sequencing platform. The *C. secundus* genome was assembled using SOAPdenovo (v2.04) with optimized parameters, followed by gapcloser (v1.10, released with SOAPdenovo) and kgf (v1.18, released with SOAPdenovo).

**Transcriptome sequencing and assembly.** For annotation purposes, 22 whole-body RNA-sequencing (RNA-Seq) samples from various developmental stages were obtained for *B. germanica*. For *C. secundus*, RNA-Seq libraries were obtained for three workers, four queens and four kings, based on degutted, whole-body extracts. In addition, we sequenced ten *Macrotermes natalensis* RNA-Seq libraries from three queens, one king and six pools of workers. All libraries were constructed using the Illumina (TruSeq) RNA-Seq kit.

For protein-coding gene annotation, *B. germanica* reads were assembled with de novo Trinity (version 2014-04-13). The *C. secundus* reads were assembled using Cufflinks on reads mapped with TopHat (version 2.2.1) and de novo Trinity and genome-guided Trinity on reads mapped with TopHat.

**Repeat annotation.** A custom *C. secundus* and *B. germanica* repeat library was constructed using a combination of homology-based and de novo approaches, including RepeatModeler/RepeatClassifier (http://www.repeatmasker.org/RepeatModeler/), LTRHarvest/LTRdigest and TransposonPSI (http://transposonpsi.sourceforge.net/). The ab initio repeat library was complemented with the RepBase (update 28 August 2016) and SINE repeat databases, filtered for redundancy with CD-hit and classified with RepeatClassifier. RepeatMasker (version open-4.0.6, http://www.repeatmasker.org) was used to mask the *C. secundus* and *B. germanica* genome. Repeat content for the other studied species (Fig. 1) was obtained from the literature.

**Protein-coding gene annotation.** The *B. germanica* genome was annotated with Maker (version 3.21.8), using the species-specific repeat library, *B. germanica* transcriptome data (22 whole-body RNA-Seq samples) and the Swiss-Prot/UniProt database (last accessed: 21 January 2016) plus the *C. secundus* and *Zootermopsis nevadensis* protein sequences for evidence-based gene model predictions. AUGUSTUS (version 3.2.4), GeneMark-ES Suite (version 4.2.1) and SNAP were used for ab initio predictions. *C. secundus* protein-coding genes were predicted using homology-based, ab initio and expression-based methods, and integrated into a final gene set with the RepBase/LTRHarvest/LTRdigest RepeatClassifier. RepeatMasker (version open-4.0.6, http://www.repeatmasker.org) was used to mask the *C. secundus* and *B. germanica* genome. Repeat content for the other studied species (Fig. 1) was obtained from the literature.
Differential gene expression. The C. secundus and M. natalensis RNA-Seq libraries were complemented with nine published Z. nevadensis libraries, yielding two to six libraries from workers, queens and kings for each termite. These were compared to six of the B. germanica libraries: two from fifth instar nymphs, two from sixth instar nymphs and two from adult females. Reads were mapped to the genome using HiSat2\(^{27}\). Read counts per gene were obtained using htseq-count and DESeq\(^{27}\) was used for differential expression analysis. Differential expression analysis between kings (males), queens (females) and workers (majors and minors combined for B. natalensis) was assessed for the termites. For B. germanica we evaluated the differential expression between adults and the two last nymphal stages combined, with the assumption that the final nymphal stages are homologous to termite workers and the adult females are homologous to termite queens. Genes were considered significantly differentially expressed if \(P < 0.05\) and \(\log(\text{fold change}) > 1\) in order to account for allometric differences as recommended in a previous study\(^{28}\).

Protein orthology. In addition to B. germanica, C. secundus, Z. nevadensis and M. natalensis, 16 other insect proteomes were included in our analyses: Locusta migratoria, Rhodnius prolixus, Ephemerida danica, Drosophila melanogaster, Aedes aegypti, Tribolium castaneum, Naosmia vitripennis, Polistes canadensis, Apis mellifera, Haploplethis saltator, Linepithema humile, Camponotus floridanus, Pogonomyrmex barbatus, Solenopsis invicta, Acromyrmex eucalyptus and Atta cephalotes; as well as for the centipede Strigamia maritima as an outgroup (for sources, see Supplementary Table 22). These proteomes were grouped into orthologous clusters with OrthoMCL\(^{81}\), with a granularity of 1.5.

IR and OR identification, phylogeny and structure. Ionotropic receptors (IRs) were identified using two custom hidden Markov models (HMMs) obtained with PfamScan\(^{38,39}\) and the termite genomes. Genes were considered significantly differentially expressed if \(P < 0.05\) and \(\log(\text{fold change}) > 1\) in order to account for allometric differences as recommended in a previous study\(^{28}\).

IRs and ORs were identified using two custom hidden Markov models (HMMs) obtained with PfamScan\(^{38,39}\) and the termite genomes. Genes were considered significantly differentially expressed if \(P < 0.05\) and \(\log(\text{fold change}) > 1\) in order to account for allometric differences as recommended in a previous study\(^{28}\).

Multiple sequence alignments of IRs and ORs were obtained with hmmalign\(^{82}\). IR and OR identification, phylogeny and structure.

Gene family expansions and contractions. For the analyses of gene family expansions and contractions, the hierarchical clustering algorithm MC-UPGMA\(^{8}\) was used, with a ProtO. level cutoff of 80 (ref. \(^{2}\)). Protein families were further divided into sub-families they contained more than 100 proteins in a single species, or more than an average of 35 proteins per species. Proteins were blasted against the RepeatMasker TE database (E-value < 10\(^{-10}\)) and clusters where > 50% of the proteins were identified as transposable elements were discarded. Clade- and species-specific protein family expansions and contractions, were identified with CAFE v3.0\(^{9}\) using the same protocol as in previous studies\(^{10,11}\) (see also Supplementary Information).

TE-facilitated expansions. The repeat content in the 10-kb flanking regions of B. germanica, C. secundus, Z. nevadensis and M. natalensis genes was calculated using bedtools\(^{12}\). Coding DNA sequences (CDSs) from neighbouring genes were removed and the repeat content was analysed using generalized linear mixed models (glimmix implemented in the R\(^{13}\) package MASS\(^{13}\) with binomial error distribution. Fixed predictors included gene family expansion, species ID and their interaction. Cluster ID was fitted as a random factor to avoid pseudo-replication. Significance was assessed on the basis of the Wald test (R package clogit\(^{14}\)) at \(a < 0.05\). Main and interaction effects for each of the genomic regions are listed in Supplementary Table 8. Model parameters are visualized in Supplementary Information.

Tests for positive selection. To test for positive selection within gene families of interest, site model tests and 8 were performed (model = 0; NSsites = 7 8) on species-specific CDS alignments, or branch-site test (model = 2; NSsites = 2; fix_omega = 1 for null model and 0 for alternative model) on multi-species alignments. Protein orthology. In addition to B. germanica, C. secundus, Z. nevadensis and M. natalensis proteins were identified as transposable elements were discarded. Clade- and species-specific protein family expansions and contractions, were identified with CAFE v3.0\(^{9}\) using the same protocol as in previous studies\(^{10,11}\) (see also Supplementary Information).

CpG depletion patterns and GO enrichment. To estimate DNA methylation, we compared observed to expected CpG counts within CDS sequences\(^{15,16}\). A low CpG\(_{nevadensis}\) indicates a high level of DNA methylation, as the cytosines of methylated CpGs often mutate to thymines. Expected CpG counts were calculated by dividing the product of cytosine and guanine counts by the sequence length. The principal component analysis in Fig. 4 was created using the R function pcorpcor on log-transformed CpG\(_{nevadensis}\) values for all 1 to 1 orthologues for the seven hemimetabolous species. These orthologues were extracted from the OrthoMCL results. The three-dimensional (3D) plot was created with the plot3d command from the R package rgl.

CpG-depleted (first quartile) and -enriched (fourth quartile) genes were tested for enrichment of Gene Ontology terms. Pfam protein domains were obtained for B. germanica, Z. nevadensis, C. secundus and M. natalensis protein sequences using PfamScan\(^{17}\). Corresponding GO terms were obtained with Pfam2GO. GO-term over-representation was assessed using the TopGO\(^{18}\) package in R. Enrichment analysis was performed using the weight algorithm selecting nodeSize = 10 to remove terms with fewer than ten annotated GO terms. After that, GO terms classified as significant (topGOFisher < 0.05) were visualized using the R package tagcloud (https://cran.r-project.org/web/packages/tagcloud/).

Life Science Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Code availability. All custom-made scripts used in these analyses are available at the following repository: https://github.com/ebbgroup/Genomic-comparisons-in-Blattodea.

Data availability. The genome assembly of Blattella germanica is archived on NCBI under the accession PRJNA203136. The genome assembly of Cryptotermes secundus is available on NCBI under the accession PRJNA381866. The additionally annotated genes for Z. nevadensis and M. natalensis are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.5ld14. Transcriptomic reads generated in this study are available in the SRA IR: germarica: PRJNA382128, C. secundus: PRJNA382129, M. natalensis: PRJNA382034.

Received: 16 August 2017; Accepted: 19 December 2017; Published online: 5 February 2018

References
1. Zahlbruck, E. & Maynard Smith, J. The major evolutionary transitions. Nature 374, 227–232 (1995).
2. Andersson, M. The evolution of eusociality. Annu. Rev. Ecol. Syst. 15, 165–189 (1984).
3. Wilson, E. O. The Insect Societies (Harvard University Press, Cambridge, 1971).
4. Rubenstein, D. R. & Abbot, P. (eds) Comparative Social Evolution (Cambridge University Press, Cambridge, 2017).
5. Misof, B. et al. Phyllogenomics resolves the timing and pattern of insect evolution. Science 346, 763–767 (2014).
6. Legrande, F. et al. Phylogeny of Dictyoptera: dating the origin of cockroaches, praying mantises and termites with molecular data and controlled fossil evidence. PLoS One 18, e0130027 (2015).
7. Bourguignon, T. et al. The evolutionary history of termites as inferred from 66 mitochondrial genomes. Mol. Biol. Evol. 32, 406–421 (2015).
8. Elsner, D., Kremer, L. P., Arning, N. & Bornberg-Bauer, E. Comparative genomic approaches to investigate molecular traits specific to social insects. Curr. Opin. Insect Sci. 16, 87–94 (2016).
9. Kapheim, K. M. et al. Genomic signatures of evolutionary transitions from solitary to group living. Science 348, 1139–1143 (2015).
10. Simola, D. F. et al. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. Genome Res. 23, 1235–1247 (2013).
11. Woodard, S. H. et al. Genes involved in convergent evolution of eusociality in bees. Proc. Natl. Acad. Sci. USA 108, 7472–7477 (2011).
12. Korb, J. & Hartfelder, K. Life history and development - a framework for understanding developmental plasticity in lower termites. Biol. Rev. 83, 395–313 (2008).
13. Roomans, J. J. & Gawe, R. Superorganismality and caste differentiation as points of no return: how the major evolutionary transitions were lost in transition. Biol. Rev. 93, 28–54 (2018).
14. Zhou, X. et al. Chemoreceptor evolution in Hymenoptera and its implications for the evolution of eusociality. Genome Biol. Evol. 7, 2407–2416 (2015).
15. Shibata, W. et al. Orco mutagenesis causes loss of antennal lobe glomeruli and impaired social behavior in ants. Cell 170, 727–735.e10 (2017).
16. Poulsen, M. et al. Complementary symbiont contributions to plant decomposition in a fungus-farming termite. Proc. Natl. Acad. Sci. USA 111, 14500–14505 (2014).
17. Terrapon, N. et al. Molecular traces of alternative social organization in a termite genome. Nat. Commun. 5, 3636 (2014).
18. Gregory, T. R. Animal Genome Size Database (accessed 25 November 2017); http://www.genomesize.com/.
19. Ferreira, P. G. 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).

20. Korsh, J. et al. A genomic comparison of two termites with different social complexity. *Front. Genet.* 6, 9 (2015).

21. Kazarian, H. H. Mobile elements: drivers of genome evolution. *Science* 303, 1622–1632 (2004).

22. Joseph, R. M. & Carlson, J. R. *Drosophila* chemoreceptors: a molecular interface between the chemical world and the brain. *Trends Genet.* 31, 682–695 (2015).

23. Brand, P. & Ramirez, S. R. The evolutionary dynamics of the odorant receptor gene family in carpiculate bees. *Genome Biol. Evol.* 9, 2023–2036 (2017).

24. Croset, V. et al. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet.* 6, e1001170 (2010).

25. Robertson, H. M., Gadau, J. & Wanner, K. W. The insect chemoreceptor superfamily of the parasitoid jewel wasp *Nasonia vitripennis*. *Insect Mol. Biol.* 19, 121–136 (2010).

26. Chen, Y., He, M., Li, Z.-Q., Zhang, Y.-N. & He, P. Identification and tissue expression profile of genes from three chemoreceptor families in an urban pest, *Periplaneta americana*. *Sci. Rep.* 6, 22094 (2016).

27. Koh, T.-W. et al. The *Insecta* superfamily of the parasitoid jewel wasp *Nasonia vitripennis*. *Insect Mol. Biol.* 19, 121–136 (2010).

28. Pellegrino, M., Steinbach, N., Stensmyr, M. C., Hansson, B. S. & Vosshall, L. B. A natural polymorphism alters odour and DEET sensitivity in an insect odorant receptor. *Nature* 478, 511–514 (2011).

29. Nichols, A. S. & Luetje, C. W. Transmembrane segment 3 of *Drosophila melanogaster* odorant receptor subunit 85b contributes to ligand-receptor interactions. *J. Biol. Chem.* 285, 11854–11862 (2010).

30. Oystaeyen, A. V. et al. Conserved class of queen pheromones stops social interactions. *Science* 344, 287–290 (2014).

31. Weil, T., Hoffmann, K., Kroiss, J., Strohm, E. & Korb, J. Scent of a queen–sex-specific DNA methylome of the termite *Macrotermes bellicosus*. *BMC Genom.* 14, 757–763 (2013).

32. Chen, Y., He, M., Li, Z.-Q., Zhang, Y.-N. & He, P. Identification and tissue expression profile of genes from three chemoreceptor families in an urban pest, *Periplaneta americana*. *Sci. Rep.* 6, 22094 (2016).

33. Koh, T.-W. et al. The *Drosophila* IR20a clade of ionotropic receptors are candidate taste and pheromone receptors. *Neuron* 83, 850–865 (2014).

34. Pellegrino, M., Steinbach, N., Stensmyr, M. C., Hansson, B. S. & Vosshall, L. B. A natural polymorphism alters odour and DEET sensitivity in an insect odorant receptor. *Nature* 478, 511–514 (2011).

35. Qiu, Y. et al. An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbons mediate discrimination of reproductives and nonreproductives in the ant *Myrmecia gulosa*. *Proc. Natl. Acad. Sci. USA* 100, 10341–10346 (2003).

36. Dallarac, R. et al. A 29 desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 97, 9449–9454 (2000).

37. Finck, J., Berdan, E. L., Mayer, F., Ronacher, B. & Geiselhardt, S. Divergence of cuticular hydrocarbons in two sympatric grasshopper species and the evolution of fatty acid synthases and elongases across insects. *Sci. Rep.* 6, 33695 (2016).

38. Qi, Y. et al. An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. *Proc. Natl. Acad. Sci. USA* 109, 14858–14863 (2012).

39.Fetcher, M. R., Cash, E. & Gadau, J. Evolution of the insect desaturase gene family with an emphasis on social Hymenoptera. *Mol. Biol. Evol.* 32, 456–471 (2015).

40. Bewick, A. J., Vogel, K. J., Moore, A. J. & Schmitz, R. J. Evolution of DNA methylation in the clonal raider ant brain. *Curr. Biol.* 26, 391–395 (2016).

41. Motro, S., Kucharski, R., Pittelkow, Y., Lockett, G. A. & Malteszka, R. Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. * BMC Genom.* 10, 472 (2009).

42. Glastad, K. M., Gokhale, K., Liebig, J. & Goodman, M. A. D. The caste- and sex-specific DNA methylome of the termite *Zootermopsis nevadensis*. *Sci. Rep.* 6, 37110 (2016).
Acknowledgements

We thank O. Nieuwkoop for allowing use of the unpublished E. danica genome, J. Gadad and C. Smith for comments and advice on the manuscript, and J. Schmitte for assistance with analyses and proofreading the manuscript. J.K. thanks Charles Darwin University (Australia), especially S. Garnett and the Horticulture and Aquaculture team, for providing logistic support to collect C. secundus. The Parks and Wildlife Commission, Northern Territory, the Department of the Environment, Water, Heritage and the Arts gave permission to collect (Permit number 36401) and export (Permit WTe2610-6997) the termites. USDA is an equal opportunity provider and employer. M.C.H. and E.J. are supported by DFG grant BO2544/11-1 to E.-B. B. and M.-D.P. are supported by University of Osnabrück and DFG grant KO1895/16-1. X.B. and M.-D.P. are supported by Spanish Ministerio de Economía y Competitividad (CGL2012-36251 and CGL2015-64727P to X.B., and CGL2016-76011-R to M.-D.P.), including FEDER funds, and by Catalan Government (2014 SGR 619). C.S. is supported by grants from the US Department of Homeland and Urban Development (NCHU-0017-13), the National Science Foundation (IOS-1557864), the Alfred P. Sloan Foundation (2013-5-35 MBE), the National Institute of Environmental Health Sciences (P30ES025128) to the Center for Human Health and the Environment, and the Blanton J. Whitmire Endowment. M.P. is supported by a Villum Kann Rasmussen Young Investigator Fellowship (VKRI0101).

Author contributions

E.B. conceived, managed and coordinated the project; M.C.H., E.J. and H.M.R. are the first authors. J.K. conceived and managed the C. secundus sequencing project, and coordinated termite-related analyses; S.R. conceived and managed the B. germanica sequencing project; S.R., S.D., S.L.I., H.C., H.V.D., H.D.Y., I.J.Q., S.C.M., S.D.T.H., K.C.W., D.M.M. and R.A.G. carried out B. germanica library construction, genome sequencing and assembly; C.S. and A.-W.-K. provided biological material through full-sib mating for B. germanica; X.B. and C.S. co-managed the B. germanica analysis; M.F. and C.P.C. implemented Web Apollo data traces; S.O. and M.P. provided biological material for M. natalensis; C.G., J.G., J.M.M.-K., A.M., F.S., H.H. and J.K. coordinated and carried out DNA and RNA sequencing for C. secundus; M.-D.P., X.B. and G.Y. coordinated transcriptome sequencing of B. germanica; L.M. performed automated gene prediction on C. secundus; E.J. and N.A. improved assembly and annotation for B. germanica & C. secundus, and compared and analysed genome sizes and quality. E.J., N.A. and L.P.M.K. analysed TeXs; M.C.H. analysed Cpg patterns and signatures of selection; T.B.-E., E.J., C.K., L.P.M.K. and A.-I.-E. performed orthology and phylogenetic analyses; L.P.M.K., E.J., H.M.R. and M.C.H. analysed gene family evolution; A.L.-E. and J.J. M.L. carried out orthoMCL clustering; H.M.R. corrected gene models for chemoreceptors; C.K. and E.J. corrected gene models for desaturases and elongases; A.-K.H. and M.C.H. corrected gene models for cytochrome P450s; E.B.-B. and M.C.H. drafted and wrote the manuscript; X.B., M.-D.P. and J.K. contributed to sections of the manuscript; E.J., L.P.M.K., A.-E.-C. and M.C.H. wrote and organized the Supplementary Information; L.P.M.K., N.A., A.-L.E., M.C.H. and E.B.-B. prepared figures for the manuscript. All authors read, corrected and commented on the manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-017-0459-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to X.B. or J.K. or E.B.-B.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.
Experimental design

1. Sample size
   Describe how sample size was determined.
   For our differential expression analyses the sample size is predetermined by the number of genes, since we were comparing full transcriptomes between conditions.

2. Data exclusions
   Describe any data exclusions.
   For the kings of the termite Macrotermes natalensis, the sequencing of several samples failed, leading to only two replicates. We therefore did not conduct or report the results of any statistical tests with these samples.

3. Replication
   Describe whether the experimental findings were reliably reproduced.
   For the differential expression analyses we only reported results for which at least 3 replicates were available. For DESeq2, the package with which we calculated differential expression, it is standard practice to work with 3 or more replicates.

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.
   This is not relevant. The experimental groups were determined by the caste membership of an individual.

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   not relevant

   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters
   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

   □ Confirmed

   □ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
   □ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
   □ A statement indicating how many times each experiment was replicated
   □ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
   □ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
   □ The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
   □ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
   □ Clearly defined error bars

   See the web collection on statistics for biologists for further resources and guidance.
Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

- The software used is described in detail within the methods and the supplementary material. These are:
  - Genome assembly: Allpaths LG, SOAPdenovo, gapcloser, kgf
  - Transcriptome assembly: Trinity, Cufflinks, TopHat
  - Repeat annotations: RepeatModeler/RepeatClassifier, LTRharvest/LTRdigest, TransposonPSI, CD-hit, Repeat Classifier, RepeatMasker
  - Annotation: Maker, AUGUSTUS, GeneMark-ES Suite, SNAP, GeneWise, PASA, GLEAN, CPC, OrfPredictor
  - Differential gene expression: HiSat2, DESeq2
  - Protein orthology: OrthoMCL
  - IR and OR identification: HMMER suite, MAFFT
  - Gene family expansions and contractions: MC-UPGMA, CAFE
  - Test for positive selection: codeml of the PAML suite
  - GO enrichment: pfam2GO, topGO.

Many custom-made scripts available on request.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

- no unique materials

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

- not applicable

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

- n/a

b. Describe the method of cell line authentication used.

- n/a

c. Report whether the cell lines were tested for mycoplasma contamination.

- n/a

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

- n/a

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

- worker, queen and kings of the two termite species: Cryptotermes secundus and Macrotermes natalensis
- Nymphs (5th and 6th instars) and adult females of Blattella germanica

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

- n/a