Evolution of Termite Symbiosis Informed by Transcriptome-Based Phylogenies

Evolution of termite symbiosis informed by transcriptome-based phylogenies

Ales Bucek[1,2,3,9*], Jan Šobotník[2], Shulin He[2,4], Mang Shi[5], Dino P. McMahon[4,6], Edward C. Holmes[5], Yves Roisin[7], Nathan Lo[8], Thomas Bourguignon[1,2,9,*]

[1]Okinawa Institute of Science & Technology Graduate University, 1919–1 Tancha, Onna-son, Okinawa, 904–0495, Japan; [2]Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Kamycka 129, 16521, Prague, Czech Republic; [3]Institute of Organic Chemistry and Biochemistry, Flemingovo nám. 2, 166 10 Prague, Czech Repubic; [4]Institute of Biology, Freie Universität Berlin, Königin-Luise-Strasse 1-3, 14195, Berlin, Germany; [5]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia; [6]Department for Materials and Environment, BAM Federal Institute for Materials Research and Testing, Unter den Eichen 87, 12205 Berlin, Germany; [7]Evolutionary Biology and Ecology, CP 160/12, Université Libre de Bruxelles, Avenue F.D. Roosevelt 50, B-1050 Brussels, Belgium; [8]School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia

[9]Lead Contact

*Correspondence:

bucek.ales@gmail.com (A.B)

Thomas.bourguignon@oist.jp (T.B.)

Keywords: fungiculture; gut symbionts; insect evolution; Isoptera; molecular clock

SUMMARY

Termitidae comprises ~80% of all termite species [1], and play dominant decomposer roles in tropical ecosystems [2,3]. Two major events during termite evolution were the loss of cellulolytic gut protozoans in the ancestor of Termitidae, and the subsequent gain in the termitid subfamily Macrotermitinae of fungal nutritional symbionts cultivated externally in 'combs' constructed within the nest [4,5]. How these symbiotic transitions occurred remain unresolved. Phylogenetic analyses of mitochondrial data previously suggested that Macrotermitinae is the earliest branching termitid lineage, followed soon after by Sphaerotermitinae [6], which cultivates bacterial nutritional symbionts on combs inside its nests [7]. This has led to the hypothesis that comb building was an important evolutionary step in the loss of gut protozoa in ancestral termitids [8]. We sequenced genomes and transcriptomes of 55 termite species, and reconstructed phylogenetic trees from up to 4065 orthologous genes of 68 species. We found strong support for a novel sister group relationship between the bacterial comb-building Sphaerotermitinae and fungus comb-building Macrotermitinae. This key finding indicates that comb building is a derived trait within Termitidae, and that the creation of a comb-like 'external rumen' involving bacteria or fungi may not have driven the loss of protozoa from ancestral termitids, as previously hypothesized. Instead, associations with gut prokaryotic nutritional symbionts, combined with dietary shifts from wood to other plant-based substrates, may have played a more important role in this symbiotic transition. Our phylogenetic tree provides a platform for future studies of comparative termite evolution and the evolution of symbiosis in this taxon.

RESULTS AND DISCUSSION

The ecological success of Termitidae was associated with major changes in digestive symbiont composition

Termites are a small insect clade, comprising about 3000 described species [1]. Termites have an enormous impact on terrestrial ecosystems, especially in the tropics, where they are the most important macroscopic decomposers of organic matter [2,3,9–12]. All termites descend from a wood-feeding ancestor, and eight out of nine termite families digest wood in association with bacteria, archaea, and lignocellulolytic protozoans [4]. The family Termitidae represents a notable exception, because its ancestors lost their protozoans but retained diverse communities of bacteria and archaea in their gut [4,13]. Two lineages of Termitidae also acquired external nutritional symbionts: the Macrotermitinae, that cultivate *Termitomyces* fungi in comb structures made within their nests [14,15], and the Sphaerotermitinae, that build bacterial combs of unknown taxonomic composition [7]. These changes in digestive symbiotic communities allowed Termitidae to diversify their diet, with many species feeding on microepiphytes, leaf litter, grass, humus and soil [16,17], and to become the most diverse group of modern termites, comprising roughly 80% of described termite species [1].

The key symbiotic transitions that paved the way to the ecological success of modern termites can be understood within a phylogenetic framework. Previous molecular phylogenetic studies demonstrated that termites are cockroaches [18,19], as was first suggested by Handlirsch and Desneux more than a century ago [20]. Termites form the sister group of the wood-roach genus *Cryptocercus*, with which they share lignocellulolytic gut protozoans [5,21]. Previous phylogenetic work also resolved the position of the Termitidae, showing that it is a highly derived lineage nested within Rhinotermitidae [6,22–28]. However, the branching pattern among basal Termitidae subfamilies, including Macrotermitinae, Sphaerotermitinae and Foraminitermitinae, and their position relative to other Termitidae, was not consistent among previous phylogenetic analyses [6,23,24,27,29], preventing formulation of robust hypotheses about the symbiotic transitions that led to modern Termitidae. In this study, we used single-copy protein-coding

genes obtained from transcriptomes and low coverage draft genomes to reconstruct a robust phylogenetic tree of termites. Our tree resolved the early evolutionary events that occurred in the Termitidae.

Transcriptome data resolve relationships among major termite lineages

We sequenced transcriptomes of 53 termite species and low coverage draft genomes of two termite species. We combined this data set with publicly available genomes and transcriptomes of 13 termite species and seven dictyopteran outgroups (see Data S1A). Our concatenated matrices included up to 4,065 single-copy orthologous protein-coding genes (OGs) spanning over 7.7 million nucleotide positions, and comprising 17-47% gaps and ambiguities. We estimated 22 maximum likelihood phylogenetic trees using a combination of orthologous gene inference methods, partitioning schemes, and models of nucleotide and amino acid substitution. We also carried out analyses with the exclusion of genes with a high proportion of missing data, and exclusion of third codon positions (Data S1B). This approach enabled us to test the robustness of our phylogenetic analyses.

A maximum likelihood phylogenetic tree based on manually-curated alignments of 462 OGs is represented in Figure 1. This tree was highly congruent with the other 21 maximum likelihood trees we inferred, as 56 of 67 internal branches were identical among the 22 trees, with ultrafast bootstrap support >95% (Figure 1). Our trees were also largely congruent with previously published phylogenies based on mitogenomes [6,26,27,30,31]. Congruent placement of lineages among all our trees and previously obtained mitogenome trees include: Mastotermitidae as the earliest branching termite lineage; the sister group relationship between Kalotermitidae and Neoisoptera; the sister-group relationship between Stylotermitidae and other Neoisoptera; and the polyphyletic nature of Rhinotermitidae, within which the monophyletic Serritermitidae and Termitidae are nested (Figure 1). Within Rhinotermitidae, our analyses strongly supported *Reticulitermes* + *Heterotermes* + *Coptotermes* as the sister group of Termitidae, as has been found in most previous studies [6,22,23,25,27], and suggested the paraphyly of *Heterotermes* with respect to *Coptotermes* [28]. Within the Termitidae, the monophyly of all subfamilies was supported, except for that of the paraphyletic Termitinae, as found previously [6,23,25,27] (Figure 1, Data S2A).

One notable incongruence with previous phylogenies was the positions of early diverging termitid lineages: Macrotermitinae, Sphaerotermitinae, and Foraminitermitinae [6,23,24,27,29] (Figure 2). All our phylogenetic trees unequivocally show that Sphaerotermitinae and Macrotermitinae form a clade, while Foraminitermitinae is recovered as sister group of Sphaerotermitinae + Macrotermitinae in some trees with third codon positions included, or sister to non-Sphaerotermitinae and non-Macrotermitinae Termitidae in all other trees (Figure 1, Data S2A). We used the approximately unbiased test on all 15 possible topological combinations of Macrotermitinae, Foraminitermitinae, Sphaerotermitinae, and the clade composed of all other Termitidae subfamilies, while other branches were left unchanged. We carried out the analyses on the manually-curated alignments of 462 orthologous genes, both on data sets with and without third codon positions (Data S2B). All topologies different from that presented in Figure 1 were rejected ($p < 0.05$), except for one alternative topology (Foraminitermitinae sister to Sphaerotermitinae + Macrotermitinae, which was not rejected when using the dataset including third codon positions ($p = 0.354$)) (Data S1C). Our study is the first to unambiguously resolve the position of Sphaerotermitinae, although that of Foraminitermitinae remains unresolved.

Basal lineages of Termitidae show high levels of gene tree discordance

Species trees and gene trees are often characterized by high levels of incongruence in lineages that went through rapid diversification [32–34], as is likely to have been the case for Termitidae. We used the coalescent-based gene tree summary method ASTRAL to infer a species tree taking into account discordance among gene trees [35,36]. We used the matrices composed of 4065 genes and 462 genes, with and without third codon positions, and reconstructed a total of eight phylogenetic trees presented as a summary-support ASTRAL tree (Data S2C; Data S1D). The ASTRAL tree was highly congruent with maximum likelihood trees based on concatenated data sets (Figure 1, Data S2C). Fifty three of 67 nodes were resolved with posterior probabilities > 0.99. The concordance among gene trees for these nodes was high, with 77-92% of gene tree quartets congruent with the ASTRAL species trees (Data S1D). Some nodes, however, exhibited high levels of local gene tree discordance. Many of these nodes were resolved neither by the maximum likelihood method, nor by the ASTRAL tree reconstruction method (Figure 1). A

few nodes were resolved with high support by both methods, but exhibited high levels of gene tree discordance, with less than 50% of gene tree quartets matching the ASTRAL species tree. Nodes with high levels of discordance include those containing Sphaerotermitinae and Foraminitermitinae representatives (Data S2C and S2D). These discordances among gene trees are suggestive of substantial amounts of introgression and/or incomplete lineage sorting among the ancestral representatives of early branching Termitidae lineages, which possibly explains the sister relationship of Sphaerotermitinae with non-Macrotermitinae and non-Foraminitermitinae Termitidae in mitochondrial genome phylogenies [27].

Timeframe of termite evolution

We used fossils of 12 termites and one mantis to calibrate the maximum likelihood tree inferred from 462 manually-curated gene alignments (Data S1E). We carried out the analyses on the trees derived from data sets with and without third codon positions. The timetree inferred from the data set with third codon positions yielded estimates up to 19.1 million years (Ma) younger than that without third codon positions (Data S2E). Here, we show the results of the latter, which have wider credibility intervals (95% CI), overlapping most of the credibility intervals obtained in the former (Figure 1). The time estimates of our tree without third codon positions diverged by less than ten million years from mitochondrial genome timetrees [6,27,28], but by up to 25 million years from a recently published cockroach transcriptome-based timetree [37]. However, our confidence intervals typically overlapped with those of the latter tree. One possible explanation for the divergences in median ages between the two trees is taxon sampling differences: the latter was focused on cockroaches (including only 6 termites), whilst our study was focused on termites (68 termites) [37]. If a shift in substitution rate occurred as a result of the evolution of eusociality [38], increased sampling of termites might influence divergence date estimation to a greater degree. A second explanation for the differences in divergence times found between our study and that of a recent study [37] is that we used nucleotide data, while the latter used amino acid data. However, our use of protein sequences resulted in even greater differences, as high as 40 million years (Data S2F). Similarly, the exclusion in our analysis of fossil calibrations not used in a previous study [37] did not substantially change the estimates shown in Figure 1 (Data S2G, S2H, S2I, and S2J).

We estimated the age of modern termites at 140.6 Ma (112.6–170.5 Ma 95% CI), suggesting that termites evolved 10 Ma before their oldest known fossil representatives [1]. We estimated the split of the drywood termite family Kalotermitidae and Neoisoptera at 120.9 Ma (96.4–147.0 Ma 95% CI), and the divergence between Termitidae and their sister clade at 64.9 Ma (51.5–79.2 Ma 95% CI). These time estimates are 15 to 25 million years (Myr) older than the oldest known fossils for these lineages [39–41]. The age of crown Termitidae was estimated at 50.1 Ma (39.9–61.1 Ma 95% CI), rapidly followed by the split of Termitidae into four lineages, Foraminitermitinae, Sphaerotermitinae, Macrotermitinae, and a clade including all other Termitidae subfamilies, all of which diverged within 3.0 Myr. Our timetree confirms that Termitidae, which represents the bulk of modern termite diversity [1], achieved ecological dominance during the past 50 Myr [6,27,42,43], about 100 Myr after the origin of termites.

The loss of lignocellulolytic protozoa in Termitidae was compensated by gut bacteria

The topology of our phylogenetic trees has important implications for models of coevolution between Termitidae and their symbionts, as it provides evidence that the loss of protozoa was originally compensated for by prokaryotic gut microbes, and that the construction of comb structures was subsequently acquired in the ancestor of Macrotermitinae and Sphaerotermitinae. All modern Macrotermitinae practice fungiculture, and grow *Termitomyces* within fungal combs inside their nests [14]. Our phylogenetic analyses show that Macrotermitinae are sister to Sphaerotermitinae, and are therefore more derived than previously acknowledged [23,27]. This phylogenetic position suggests that *Termitomyces* were acquired once in the ancestor of Macrotermitinae and have never been lost since then. In a similar way, the phylogenetic position of Sphaerotermitinae also suggests that the bacterial symbionts they cultivate on combs [7] were acquired once in the ancestor of Sphaerotermitinae and have been retained since then (although further studies on the taxonomic composition of comb bacteria are required to test this). Alternative hypotheses, such as acquisition of either fungal or bacterial symbionts in the common ancestor of Macrotermitinae + Sphaerotermitinae and subsequent replacement of the symbionts in the common ancestor of one of the subfamilies, are less parsimonious because they require additional loss events. An alternative scenario involving early replacement of protozoa by novel symbionts cultivated on combs, coined the ‘external

rumen' [8], based on the position of Macrotermitinae as the earliest branching termitid lineage, is less parsimonious, as it implies one additional event: the loss of external rumen in the rest of Termitidae. Therefore, transcriptome-based phylogenies suggest that the cellulolytic protozoa of lower termites were originally replaced by gut prokaryotes in the ancestor of Termitidae.

One potential explanation for the loss of protozoa is an early origin of soil-feeding in the ancestor of all modern Termitidae, depriving cellulolytic protozoa of cellulose and driving them to extinction. We reconstructed the evolution of diet on the phylogenetic tree represented in Figure 1 using a maximum likelihood model [44]. We recovered a 25% probability that the termitid ancestor was a soil-feeder (Data S2K). This probability rose to 38% in the case of Foraminitermitinae being sister to Macrotermitinae + Sphaerotermitinae, as suggested by some of our phylogenetic analyses (Data S2L). Therefore, ancestral state reconstructions neither support nor reject a soil-feeding ancestor for all modern Termitidae. Alternative evidence for the hypothesis could come from fossils of stem Termitidae. The oldest know fossil of Termitidae, the tiny *Nanotermes*, might be one such fossil, but its affinity with modern Termitidae remains to be clarified, and its morphological description lacks key characters, such as mandible shape, preventing any prediction of its diet [41]. An alternative hypothesis is that the loss of protozoa was caused by a shift in diet from wood to leaf litter.

Conclusion

Our study uses phylogenomic approaches to provide a robust backbone of termite evolution. Our phylogenetic trees support the result of previous studies [6,22,23,25,27], and show that several taxonomic groups are not monophyletic, calling for a nomenclatorial revision of these higher-ranked taxa. Nonnatural groups include: Rhinotermitidae, within which Serritermitidae and Termitidae are nested; Termitinae, which also includes Cubitermitinae, Nasutitermitinae and Syntermitinae; and *Heterotermes* which is paraphyletic with respect to *Coptotermes*. Our phylogenetic trees also resolve key nodes that suggests a reinterpretation of termite and digestive symbiont evolution. Our results imply that (1) fungiculture is a derived trait, unique to Macrotermitinae, that evolved several million years after the loss of gut protozoa; (2) the construction of combs to externally cultivate nutritional symbionts evolved once in the ancestor of Sphaerotermitinae and

Macrotermitinae and was retained since then; and, therefore, (3) gut prokaryotes replaced gut protozoa as the key digestive symbionts in the last common ancestor of Termitidae.

ACKNOWLEDGMENTS

We are grateful to David Sillam-Dussès and Petr Stiblík for sampling assistance, and to Jan Křeček for help with species identification. This work was supported by subsidy funding of the Okinawa Institute of Science and Technology, by project CIGA No. 20184306 (Czech University of Life Sciences, Prague), and grant "EVA4.0", No. CZ.02.1.01/0.0/0.0/16_019/0000803 financed by OP RDE. NL was supported by an Australian Research Council Future Fellowship (FT160100463). Computational resources were provided by the Okinawa Institute of Science and Technology and by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the program "Projects of Large Research, Development, and Innovations Infrastructures".

AUTHOR CONTRIBUTIONS

AB, JS and TB conceptualized the experiments. JS, SH, DPM, YR, NL and TB collected the samples. AB, SH, MS, and ECH performed lab experiments and/or generated data. AB analyzed the data, with significant input from TB. AB, NL and TB wrote the paper, with significant input of all other coauthors.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

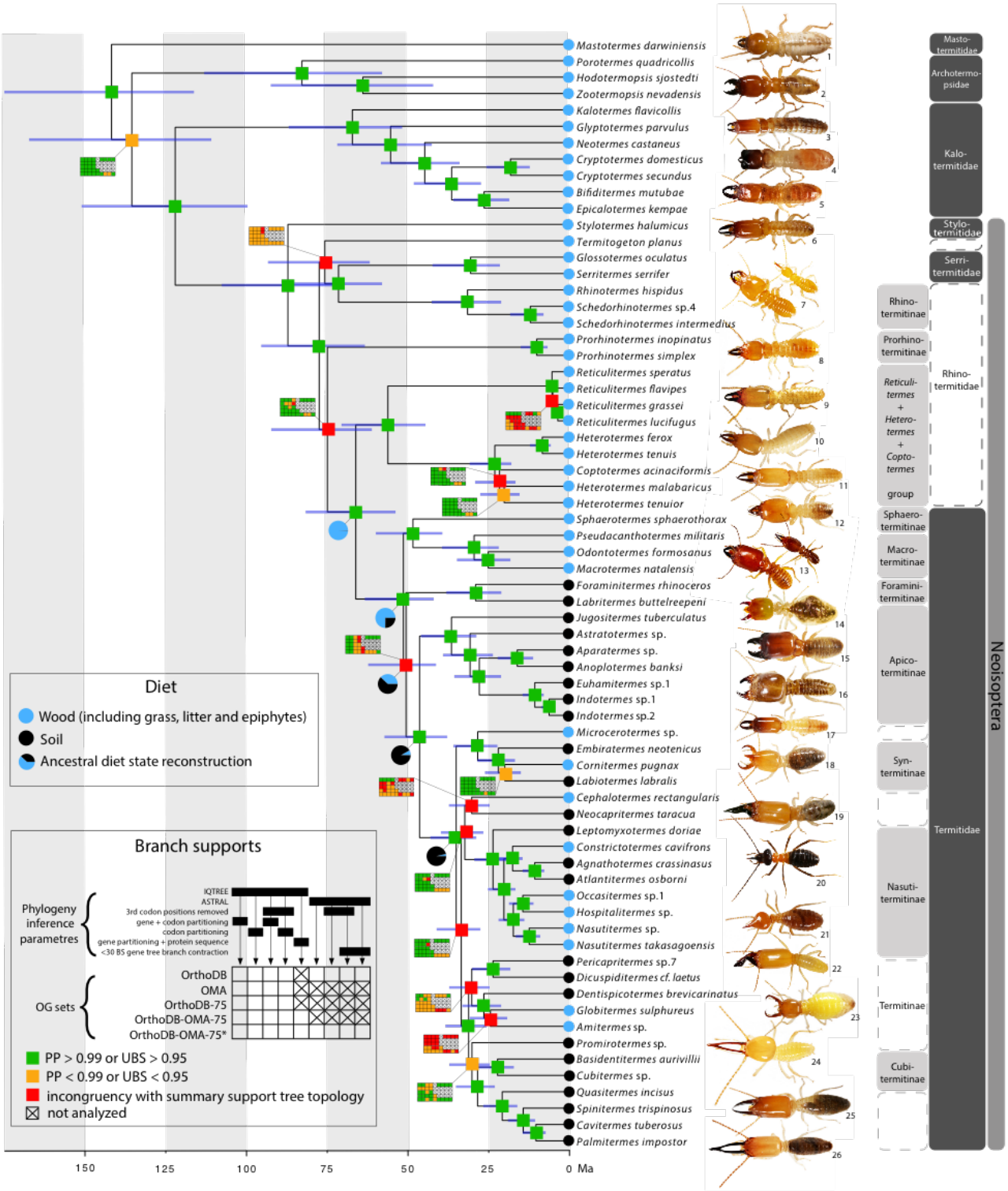

Figure 1. Time-calibrated phylogenetic tree of termites inferred from manually-curated alignments of 462 orthologous genes, without third codon positions, using maximum likelihood inference.

Internal tree nodes are labeled with colors summarizing the branch support from 22 maximum likelihood trees and eight ASTRAL trees. Color-coded matrices show supports

from each individual phylogenetic analysis for the branches that were not unequivocally supported by all analyses (see Data S1B and S1D for details on ortholog gene (OG) sets and phylogeny inference parameters used). Ultrafast bootstrap support (UBS) values were obtained using IQ-TREE and posterior probabilities (PP) were obtained using ASTRAL. The node bars represent the 95% confidence interval of age estimates. Tree tips are labeled with blue or black indicating a diet respectively consisting of wood (here including grass, litter and epiphytes) or soil. Selected internal nodes are labeled with pie charts showing the ancestral diet states (see Data S2K and S2L for all ancestral diet states). Termite soldiers represented alongside the tree are those of: 1, *Mastotermes darwiniensis,* 2, *Hodotermopsis sjostedti,* 3, *Glyptotermes* sp., 4, *Cryptotermes* sp., 5, *Epicalotermes kempae,* 6, *Stylotermes halumicus,* 7, *Rhinotermes hispidus,* 8, *Prorhinotermes canalifrons,* 9, *Reticulitermes flavipes,* 10, *Coptotermes formosanus,* 11, *Heterotermes tenuis,* 12, *Sphaerotermes sphaerothorax,* 13, *Pseudacanthotermes militaris,* 14, *Foraminitermes valens,* 15, *Jugositermes tuberculatus,* 16, *Indotermes sp.,* 17, *Microcerotermes sp.,* 18, *Embiratermes neotenicus,* 19, *Neocapritermes taracua,* 20, *Constrictotermes cavifrons,* 21, *Nasutitermes octopilis,* 22, *Pericapritermes sp.,* 23, *Globitermes sulphureus,* 24, *Promirotermes sp.,* 25, *Spinitermes trispinosus,* 26, *Palmitermes impostor*. The polyphyletic subfamily Termitinae and the polyphyletic family Rhinotermitidae are delimited with dashed boxes.

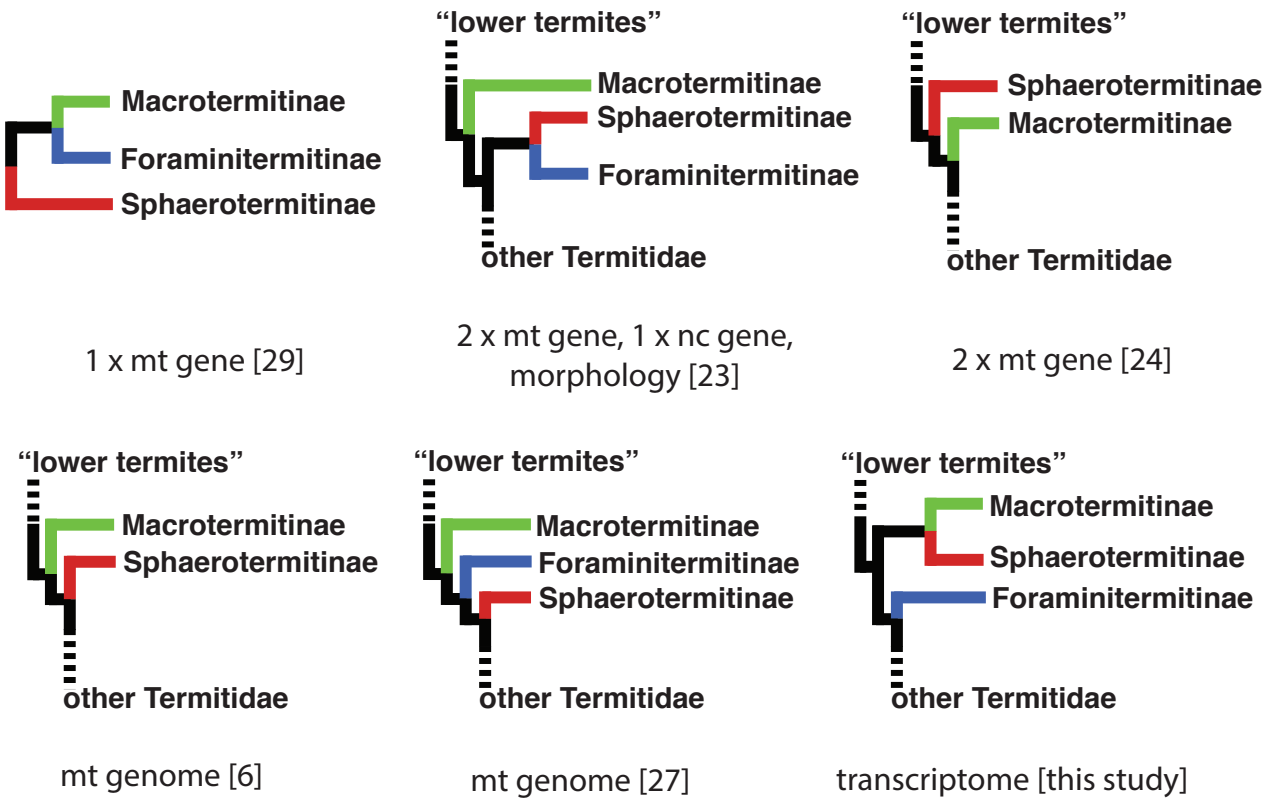

Figure 2. Overview of phylogenetic hypotheses on relationships among Macrotermitinae, Sphaerotermitinae, and Foraminitermitinae.

The literature reference and the type of data used for phylogenetic inference is indicated, i.e. mitochondrial (mt) genes and genomes, nuclear (nc) genes, morphology, and single-copy gene orthologs inferred from transcriptomes. "Lower termites" = paraphyletic termite group including all termites with the exception of Termitidae. "other Termitidae" = Termitidae subgroup excluding Macrotermitinae, Sphaerotermitinae and Foraminitermitinae.

STAR METHODS

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests may be directed to and will be fulfilled by the lead contact Thomas Bourguignon (thomas.bourguignon@oist.jp). Ales Bucek (bucek.ales@gmail.com) may also be contacted for further information. This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We obtained transcriptomes from 55 termite species and low coverage genomes from two termite species. In each case, RNA (or DNA) was obtained from worker heads. Specimens used in this study were stored at low temperature in RNA-later® or in TRIzol. Details on individual sample collection can be found in Data S1A.

METHOD DETAILS

RNA and DNA isolation and manipulation

This paper is the result of a collaborative effort aiming to reconstruct a termite phylogenetic tree (Figure 1, Data S2). It involved independent teams using customized in-house protocols. We used a total of four different procedures to extract and sequence RNA and DNA from 55 termite species (Data S1A):

Procedure 1 (RNA isolation and sequencing): Termites were collected in RNA-later®. Each sample was temporarily stored in the field at temperature varying between -20˚C and 4˚C, and then kept at -80˚C until RNA extraction. We dissected with a scalpel the heads of 2 to 15 individuals, predominantly workers, and transferred them in 400 µl of TRIzol Reagent (Invitrogen). Heads were homogenized using a Fisherbrand™ Disposable Pestle System (Fisher Scientific). We extracted total RNA using a standard phenol-chloroform procedure with TRIzol according to manufacturer's protocol. Extracted RNA samples were treated with TURBO DNase (Ambion) at 37˚C for 1h, and purified using the RNA Clean & Concentrator™-5 kit (Zymo Research). RNA samples were poly(A)+ enriched and fragmented. cDNA libraries were paired-end (2 x 125 bp) sequenced using an Illumina HiSeq 2500 platform.

Procedure 2 (RNA isolation and sequencing): Termites were collected and immediately stored at -80˚C until RNA extraction. Whole soldier and worker bodies were homogenized and RNA was isolated with RNeasy Plus Mini Kit (QIAGEN). rRNA was depleted with Ribo-Zero Gold rRNA Removal Kit (Illumina). cDNA libraries were prepared with a TruSeq Total RNA library Prep kit (Illumina) and paired-end (2 x 100 bp) sequenced using an Illumina HiSeq 2500 platform.

Procedure 3 (RNA isolation and sequencing): Termites were collected, immediately frozen in liquid nitrogen, and stored at -70 °C until RNA extraction. Multiple whole termite bodies were pooled in pre-cooled Trizol (Thermo Fisher Scientific), and homogenized twice with 5 mm steel beads (Qiagen) using a homogenizer (MP Biomedicals) at 2 M/s for 10 s. RNA was isolated according to the manufacturer's instructions with chloroform extraction and isopropanol precipitation, and dissolved in RNA storage solution (Ambion). Subsequently, total RNA was incubated with 2 units of TurboDNase (Ambion) for 30 min at 37 °C and purified using an RNAeasy Mini kit (Qiagen) according to the manufacturer's instructions. Quantity and quality of RNA were determined using a Qubit and Bioanalyzer 2100. Barcoded cDNA libraries were prepared using a NEXTflex™ Rapid Directional mRNA-seq kit (Bio Scientific) according to the manufacturer's instructions. Libraries were paired-end (2 x 75 bp) sequenced using the Illumina NextSeq500/550 platform.

Procedure 4 (DNA isolation and sequencing): We were unable to sequence transcriptomes of *Foraminitermes rhinoceros* and *Labritermes buttelreepeni* due to a lack of samples of sufficient quality. We instead sequenced low coverage genomes. Termites were collected in RNA-later®. Each sample was temporarily stored in the field at temperature varying between -20˚C and 4˚C, and then kept at -80˚C until DNA extraction. Five workers were homogenized, and DNA was isolated with the DNeasy Blood and Tissue extraction kit (Qiagen) using a QiaCUBE machine. DNA concentration was measured using Qubit, and libraries were prepared using Ultra FS II Library Preparation Kit (New England Biolabs) and the Unique Dual Indexing Kit (New England Biolabs). Libraries were pooled and sequenced using the Illumina HiSeq 4000 platform.

QUANTIFICATION AND STATISTICAL ANALYSIS

Transcriptome assembly

Read quality was evaluated using FastQC. Adapters were trimmed, and low-quality reads were trimmed and filtered with Trimmomatic v0.32 [45] using the following parameters: TRAILING:25 LEADING:25 SLIDINGWINDOW:4:20 AVGQUAL:20 MINLEN:50. Quality-filtered reads were *de novo* assembled using Trinity v2.4.0 with default parameters [46]. Assembly statistics were calculated with the TrinityStats.pl script from the Trinity package. The completeness of transcriptome assemblies was assessed with BUSCO [47], using 1658 single-copy orthologous genes for insects. The list of genes we used is available at: http://busco.ezlab.org/v2/datasets/insecta_odb9.tar.gz (July 2017). BUSCO was run using filtered termite transcriptome assemblies, only including the longest isoform for each gene. Publicly available transcriptomes and genomes from Dictyoptera were retrieved either assembled, or as raw reads, in which case assembly was carried out as described above. See Data S1A for further information on the statistics of the *de novo* assembled transcriptomes.

Draft genome assembly

To assemble low coverage genomes of *Foraminitermes rhinoceros* and *Labritermes buttelreepeni*, Illumina reads were first trimmed and quality-filtered with BBDuk from the BBMap v38.06 package using the command line "ktrim=r ktrim=l k=23 mink=6 hdist=1 tpe tbo maq=10 qtrim=rl trimq=20 minlength=35 restrictleft=50". Filtered and trimmed reads were then assembled with SOAPdenovo2 r241 [48] and SPAdes 3.12.0 [49], and the assembly completeness was assessed with BUSCO [47], as described above. Because the BUSCO statistics of SPAdes assemblies were substantially better, we used SPAdes assemblies for downstream analyses.

Orthology prediction

We used the best reciprocal hit search strategy to infer orthologous gene groups (OGGs) for each transcriptome and low coverage genome. First, we used the predicted complete proteomes of three reference species of Dictyoptera (*Blattella germanica*, *Zootermopsis nevadensis*, and *Macrotermes natalensis*) and two inference methods, one implemented in

the OMA software v2.0.0 [50], and one implemented in the OrthoDB server (accessed February 2017) [51] to infer reference OGGs. OMA inferred 2981 single-copy OGGs, and OrthoDB inferred 4065 single-copy OGGs. We used these reference OGGs to predict OGGs in each transcriptome and low coverage genome using the best reciprocal hit search strategy implemented in the Orthograph tool [52]. The analyses were repeated twice, once with the 2981 single-copy OGGs determined with OMA, and once with the 4065 single-copy OGGs determined with OrthoDB.

Sequence alignment

We aligned protein sequences of each OGG using the software MAFFT v7.305 and the command line: --maxiterate 1000 --globalpair [53]. All sequences with internal stop codons or putative selenocysteine codons were removed. Protein alignments were back translated to nucleotide sequences using pal2nal v14 [54]. Alignments were concatenated with FASconCAT-G_v1.04.pl [55]. We manually curated and removed spurious OGs from a concatenated supermatrix of 462 OGGs that were retrieved in all 75 species analyzed in this study, both with OMA and with OrthoDB. During the curation, we paid special attention to the low coverage genome sequences of *F. rhinoceros* and *L. buttelreepeni*, from which we manually trimmed short stretches of nucleotides presenting no homology to other species. These short stretches were presumably short portions of introns adjacent to exons retrieved via Orthograph. In total, we generated five concatenated supermatrices: 1) a supermatrix including all OGGs inferred with OrthoDB, 2) a supermatrix including all OGGs inferred with OMA, 3) a supermatrix including OGGs inferred with OrthoDB that were retrieved in all 75 species analyzed in this study, 4) a supermatrix including all OGGs that were retrieved in all 75 species analyzed in this study, both with OMA and with OrthoDB, and 5) the latter supermatrix with manual curation (Data S1B).

Phylogenetic analyses

We used the maximum likelihood method implemented in IQ-TREE 1.6.7 [56] to estimate phylogenetic trees. We used the edge-proportional partition model [57], and the relaxed hierarchical clustering partition scheme [58]. Bootstrap resampling was carried out using the ultrafast bootstrap with 1000 replicates [59]. Each bootstrap tree was optimized using

the hill-climbing nearest neighbor interchange search. We determined the best-fit nucleotide substitution model with ModelFinder [60], using the Bayesian information criterion. For the largest data sets, calculation of the best-fit model was computationally too demanding, and we limited our ModelFinder search to five models: GTR, GTR+I, GTR+G, GTR+G+I, and GTR+G+I+R. All phylogenetic analyses were carried out with and without third codon positions. Summary-support trees were generated with TreeGraph 2 [61]. We mapped ultrafast bootstrap branch supports from all maximum likelihood trees onto the maximum likelihood tree inferred from 462 manually-curated OG alignments with third codon positions included. An overview of the trees inferred with IQ-TREE, including information on the parameters used, is available in Data S1B.

We also reconstructed phylogenetic trees using the coalescent-based method implemented in ASTRAL 5.5.9 which infers species trees by summarizing gene trees [62]. First, we estimated gene trees using IQ-TREE with parameters set as described above, both with and without third codon positions. We used OGGs inferred using OrthoDB, and ran the analyses on two data sets, on that with all OGGs, and on that including only OGGs retrieved in all 75 species analyzed in this study. To account for the low accuracy of gene tree estimation [62], we generated for each OGG set an additional gene tree with collapsed low-support branches. The branches with ultrafast bootstraps lower than 30 were collapsed using Newick Utilities 1.6 [63]. We ran ASTRAL with the option -t 2, which produces quartet branch supports and local posterior probabilities for all topologies. Summary-support trees were generated with TreeGraph [61] by mapping posterior probabilities and quartet supports from all ASTRAL trees onto the ASTRAL tree inferred from 462 manually-curated OG alignments, with third codon positions included. An overview of the trees reconstructed with ASTRAL, and the parameters used, is available in Data S1D.

Tree topology test

We tested the probability of alternative branching patterns for Foraminitermitinae and Sphaerotermitinae. Alternative tree topologies were tested using the approximately unbiased test [64] implemented in IQ-TREE 1.6.9 with 10,000 RELL replicates.

Molecular dating

We estimated a timescale of termite evolution using the program MCMCTREE implemented in the PAML 4.9g package [65,66]. We used the maximum likelihood trees inferred from 1) 462 manually-curated protein alignments and 2) 462 manually-curated nucleotide alignments, both with and without third codon positions, as input trees for MCMCTREE. We used Bayesian estimation of species divergence time with approximate likelihood calculation. We first calculated the gradient and Hessian of the log-likelihood (*usedata* = *3*) with the HKY+Gamma model of nucleotide sequence evolution and the JTT model of amino acid substitution for nucleotide and protein data, respectively. The calculated gradient and Hessian values were used for MCMC sampling of posterior distribution using the approximate likelihood method (*usedata* = *2*). We used an independent lognormal clock model (*clock* = *2*). Node age priors for the nodes without fossil calibrations were uniformly distributed between present time and root age (*BDparas* = *1 1 0*). We set the gamma-Dirichlet prior on the mean substitution rate for partitions at 0.05 substitutions per 100 Myr (*rgene_gamma* = *2 40 1*), and we set the rate variance parameter *sigma2_gamma* = *1 10 1*. The MCMC chain was run for 50,500,000 iterations and sampled every 100 iterations. The first 500,000 iterations were discarded as burn-in. We ran two replicates of the MCMC chain for each input tree. The probability for violation of the bound was set to 0.025.

Fossil calibrations

We used 12 termite fossils and one mantis fossil as internal calibrations (see Data S1E). In all cases, we used the youngest fossil age estimates reported on the fossilworks database [67]. *Juramantis initialis*, the only non-termite fossil used in this study, was used to calibrate Dictyoptera [68]. *J. initialis* was described on the basis of a fragment of wing and placed within the Mantodea. The features used for this placement were interpreted as specific to Mantodea [68]; however, they are features widespread in Blattodea [69], suggesting that the placement of *J. initialis* within Mantodea might be erroneous. Its placement within Dictyoptera is, however, unambiguous. We used *Valditermes brenanae* and *Archeorhinotermes rossi* to calibrate the nodes corresponding to Isoptera +

Cryptocercus and Kalotermitidae + Neoisoptera, respectively. The use of these two fossil calibrations is supported by the analyses of [43], and additional justifications are provided by [37]. We used *Nanotermes isaacae* to calibrate the node corresponding to Termitidae + sister group. *N. isaacae* is a definite member of the Termitidae, possibly representing a stem group termitid or a crown group termitid [41]. The nine remaining calibrations are all based on fossils from Oligocene or Miocene deposits. Six fossils were described from the Dominican amber: *Constrictotermes electroconstrictus*, *Microcerotermes insularis*, *Termes primitivus*, *Amitermes lucidus*, *Anoplotermes* sensu lato, and *Dolichorhinotermes dominicanus*. The first four fossils are all ascribed to modern genera [70,71], while the taxonomic position of the last two fossils is not as clear. Eleven species of *Anoplotermes* sensu lato preserved in Dominican amber have been described by Krishna and Grimaldi (2009) [70], all of which belong to the South American *Anoplotermes*-group, but their generic status is unclear, as is the case for most extant species of the South American *Anoplotermes*-group [72]. We used these fossils to calibrate the node corresponding to all Apicotermitinae with the exclusion of *Jugositermes*. *D. dominicanus* is a clear member of the *Rhinotermes*-complex [73], which we used to calibrate *Rhinotermes* + sister group. Note that additional species of *Dolichorhinotermes* are known from the Mexican amber [74]. *Reticulitermes antiquus* is known from the Baltic amber. Its generic assignment is clearly established, but its relationship with modern *Reticulitermes* species is unknown [75]. In our views, it represents a stem group *Reticulitermes*, and we therefore used it to calibrate the node corresponding to *Reticulitermes* + sister group. *Coptotermes sucineus* was described by Emerson (1971) from alate specimens preserved in Mexican amber [76]. Specimens assigned to *Heterotermes*, a paraphyletic genus within which *Coptotermes* is nested [28], were present in the same piece of amber [76]. The fossil of *C. sucineus* therefore represents an ideal calibration for *Heterotermes* + *Coptotermes*. *Macrotermes pristinus* was described by Charpentier (1843) as *Termes pristinus* [77], and later on assigned to *Macrotermes* by Snyder (1949) [78]. We use it as a calibration of the node *Macrotermes* + *Odontotermes*.

In addition, we used two alternative sets of fossil calibrations with 10 termite fossils and one mantis fossil (see Data S1F and S1G). Both fossil sets excluded *Reticulitermes antiquus* and *Nanotermes* fossils which were excluded also by Evangelista et al. [37]. The first alternative set used 47.8 Ma as a maximum age constraint of first fossil Termitidae,

i.e. the age of *Nanotermes* fossil (Data S1F), while the second used 94.3 Ma as maximum age constraint, i.e. the age of *Archeorhinotermes rossi* fossil (Data S1G).

Ancestral state reconstruction of diet

We reconstructed the ancestral state of termite diet using the maximum likelihood trees inferred from 462 manually-curated alignments. We carried out the analyses on the trees estimated with and without third codon positions. We considered two states, wood-feeders and soil-feeders, determined based on data from the literature [16,17]. Wood-feeders equated to Donovan et al.'s feeding-group I and II (thus including termite species feeding on grass, leaf litter, and epiphytes), and soil-feeders equated to feeding-group III and IV [16]. We carried out the reconstruction using the "ace" function available in the R package phytools [79]. The model was a maximum Likelihood model with equal rate of transition between states. All analyses were carried out in R version 3.2.0.

DATA AND SOFTWARE AVAILABILITY

The raw reads generated for 55 termite species using shotgun genome and transcriptome sequencing are available in NCBI Short Read Archive as BioProject PRJNA560101 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA560101/) under sample accessions SAMN12568907- SAMN12568961. Low coverage genome assembly and transcriptome assemblies are available upon request.

Supplemental data legends

Data S1. Summary of the phylogenetic analyses carried out in this study. Related to STAR Methods. (A) List of samples, collection localities, statistics of transcriptome and genome assemblies, and results of orthology predictions. (B) List of the maximum likelihood analyses performed with IQTREE. (C) Results of the approximately unbiased tree topology test. The null hypothesis of rejecting tree topology ($p < 0.05$) was tested for 15 alternative tree topologies (see Data S2B), using the dataset composed of 462 manually-curated orthologous genes, with and without third codon position. logL, tree likelihood; p-AU, p-value; rejected topologies are marked with the symbol '-', and non-rejected topologies are marked with the symbol '+'. (D) List of coalescent-based gene tree

reconstructed with ASTRAL. (E) Fossils used as calibrations for estimating divergence times. (F) First alternative fossil set used as calibrations for estimating divergence times. (G) Second alternative fossil set used as calibrations for estimating divergence times.

Data S2. Termite phylogenetic trees reconstructed in this study. Related to STAR Methods. (A) Summary support maximum likelihood tree. The maximum likelihood tree summarizes branch bootstrap supports calculated from 22 different maximum likelihood analyses mapped onto tree topology reconstructed using 462 manually-curated orthologous genes, with third codon positions included. Individual maximum likelihood analyses differed in inclusion/exclusion of third codon positions, partitioning strategies, and subsets of orthologous genes used. Values in square brackets indicate topologies conflicting with that of the represented tree. See Data S1B for details on phylogenetic reconstruction parameters. (B) The 15 alternative topologies tested using the approximately unbiased test. We tested all possible combinations of topological variations among four clades: Sphaerotermitinae, Foraminitermitinae, Macrotermitinae, and the clade composed of all other Termitidae. All other branches were kept unchanged. (C) Summary support coalescent-based species tree with posterior probabilities. The coalescent-based species tree summarizes branch posterior probabilities (PP) calculated from eight different ASTRAL analyses mapped onto tree topology reconstructed with ASTRAL using 462 manually-curated orthologous genes, with third codon positions included. Values in square brackets indicate topology conflicting with the represented coalescent-based species tree. See Data S1D for details on phylogenetic reconstruction parameters of all coalescent-based species trees. (D) Summary support coalescent-based species tree with quartet branch supports. The coalescent-based species tree summarizes quartet supports for the main topology (i.e. the topology shared by the highest proportion of gene trees) calculated from eight different ASTRAL analyses mapped onto tree topology reconstructed with ASTRAL using 462 manually-curated orthologous genes, with third codon positions included. Values in square brackets indicate topology conflicting with the represented coalescent-based species tree. See Data S1D for details on phylogenetic inference parameters of all coalescent-based species trees. (E) Time-calibrated maximum likelihood tree based on nucleotide data with third codon positions. The node bars represent the 95% confidence interval of age estimates. See Data S1E for overview of used fossil

calibrations. (F) Time-calibrated maximum likelihood tree based on protein data. The node bars represent the 95% confidence interval of age estimates. See Data S1E for overview of used fossil calibrations. (G) Time-calibrated maximum likelihood tree based on nucleotide data with third codon positions and first alternative set of fossil calibrations. The node bars represent the 95% confidence interval of age estimates. See Data S1F for overview of fossils used for calibration. (H) Time-calibrated maximum likelihood tree based on nucleotide data with third codon positions and second alternative set of fossil calibrations. The node bars represent the 95% confidence interval of age estimates. See Data S1G for overview of fossils used for calibration. (I) Time-calibrated maximum likelihood tree based on nucleotide data without third codon positions and first alternative set of fossil calibrations. The node bars represent the 95% confidence interval of age estimates. See Data S1F for overview of fossils used for calibration. (J) Time-calibrated maximum likelihood tree based on nucleotide data without third codon positions and second alternative set of fossil calibrations. The node bars represent the 95% confidence interval of age estimates. See Data S1G for overview of fossils used for calibration. (K) Ancestral diet reconstruction for the tree topology represented in Figure 1. Red: soil-feeder; black: wood-feeder (this category includes species feeding on grass, litter and epiphytes). (L) Ancestral diet reconstruction for the topology of tree14 represented in Data S2B. Red: soil-feeder; black: wood-feeder (this category includes species feeding on grass, litter and epiphytes).

REFERENCES

1. Krishna, K., Grimaldi, D.A., Krishna, V., and Engel, M.S. (2013). Treatise on the Isoptera of the World. Bull. Am. Museum Nat. Hist. *377*, 1–200. Available at: http://www.bioone.org/doi/abs/10.1206/377.1.

2. Ashton, L.A., Griffiths, H.M., Parr, C.L., Evans, T.A., Didham, R.K., Hasan, F., Teh, Y.A., Tin, H.S., Vairappan, C.S., and Eggleton, P. (2019). Termites mitigate the effects of drought in tropical rainforest. Science (80-.). *363*, 174–177. Available at: http://www.sciencemag.org/lookup/doi/10.1126/science.aau9565.

3. Sugimoto, A., Bignell, D.E., and MacDonald, J.A. (2000). Global Impact of Termites

on the Carbon Cycle and Atmospheric Trace Gases. In Termites: Evolution, Sociality, Symbioses, Ecology (Dordrecht: Springer Netherlands), pp. 409–435. Available at: http://link.springer.com/10.1007/978-94-017-3223-9_19.

4. Brune, A. (2014). Symbiotic digestion of lignocellulose in termite guts. Nat. Rev. Microbiol. *12*, 168–180. Available at: http://www.nature.com/doifinder/10.1038/nrmicro3182.

5. Ohkuma, M., and Brune, A. (2010). Diversity, Structure, and Evolution of the Termite Gut Microbial Community. In Biology of Termites: a Modern Synthesis (Dordrecht: Springer Netherlands), pp. 413–438. Available at: http://link.springer.com/10.1007/978-90-481-3977-4_15.

6. Bourguignon, T., Lo, N., Cameron, S.L., Šobotník, J., Hayashi, Y., Shigenobu, S., Watanabe, D., Roisin, Y., Miura, T., and Evans, T.A. (2015). The evolutionary history of termites as inferred from 66 mitochondrial genomes. Mol. Biol. Evol. *32*, 406–421. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25389205 [Accessed November 21, 2014].

7. Garnier-Sillam, E., Toutain, F., Villemin, G., and Renoux, J. (1989). Études préliminaires des meules originales du termite xylophage *Sphaerotermes sphaerothorax* (Sjostedt).

8. Aanen, D.K., and Eggleton, P. (2017). Symbiogenesis: Beyond the endosymbiosis theory? J. Theor. Biol. *434*, 99–103. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0022519317303612.

9. Holt, J.A., and Lepage, M. (2000). Termites and Soil Properties. In Termites: Evolution, Sociality, Symbioses, Ecology (Dordrecht: Springer Netherlands), pp. 389–407. Available at: http://link.springer.com/10.1007/978-94-017-3223-9_18.

10. Bignell, D.E. (2016). The Role of Symbionts in the Evolution of Termites and Their Rise to Ecological Dominance in the Tropics. In The Mechanistic Benefits of Microbial Symbionts. Advances in Environmental Microbiology, Christon J. Hurst, ed. (Springer, Cham), pp. 121–172. Available at: http://link.springer.com/10.1007/978-3-319-28068-4_6.

11. Eggleton, P. (2000). Global Patterns of Termite Diversity. In Termites: Evolution, Sociality, Symbioses, Ecology (Dordrecht: Springer Netherlands), pp. 25–51.

Available at: http://link.springer.com/10.1007/978-94-017-3223-9_2.

12. Eggleton, P., Bignell, D.E., Sands, W.A., Mawdsley, N.A., Lawton, J.H., Wood, T.G., and Bignell, N.C. (1996). The diversity, abundance and biomass of termites under differing levels of disturbance in the Mbalmayo Forest Reserve, southern Cameroon. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. *351*, 51–68. Available at: http://www.royalsocietypublishing.org/doi/10.1098/rstb.1996.0004.

13. Brune, A., and Dietrich, C. (2015). The gut microbiota of termites: digesting the diversity in the light of ecology and evolution. Annu. Rev. Microbiol. *69*, 145–66. Available at: http://www.annualreviews.org/doi/10.1146/annurev-micro-092412-155715.

14. Rouland-Lefèvre, C. (2000). Symbiosis With Fungi. In Termites: Evolution, Sociality, Symbioses, Ecology (Dordrecht: Springer Netherlands), pp. 289–306. Available at: http://link.springer.com/10.1007/978-94-017-3223-9_14.

15. Grassé, P.P. (1984). Termitologia vol.2 (Paris: Masson).

16. Donovan, S.E., Eggleton, P., and Bignell, D.E. (2001). Gut content analysis and a new feeding group classification of termites. Ecol. Entomol. *26*, 356–366. Available at: http://doi.wiley.com/10.1046/j.1365-2311.2001.00342.x.

17. Bourguignon, T., Šobotník, J., Lepoint, G., Martin, J.-M., Hardy, O.J., Dejean, A., and Roisin, Y. (2011). Feeding ecology and phylogenetic structure of a complex neotropical termite assemblage, revealed by nitrogen stable isotope ratios. Ecol. Entomol. *36*, 261–269. Available at: http://doi.wiley.com/10.1111/j.1365-2311.2011.01265.x.

18. Lo, N., Tokuda, G., Watanabe, H., Rose, H., Slaytor, M., Maekawa, K., Bandi, C., and Noda, H. (2000). Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. Curr. Biol. *10*, 801–804. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0960982200005613.

19. Inward, D.J.G., Beccaloni, G., and Eggleton, P. (2007). Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. Biol. Lett. *3*, 331–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17412673.

20. Lameere, A. (1910). Assemblée générale du 26 décembre 1909. Annls Soc. ent.

Belg. *53*, 507–517.

21. Cleveland, L.R., Hall, S.R., Sanders, E.P., and Collier, J. (1934). The wood-feeding roach, *Cryptocercus*, its protozoa and the symbiosis between protozoa and roach. Mem. Amer.Acad. Arts Sc. *17*, 185–342.

22. Lo, N., Kitade, O., Miura, T., Constantino, R., and Matsumoto, T. (2004). Molecular phylogeny of the Rhinotermitidae. Insectes Soc. *51*, 365–371. Available at: http://link.springer.com/10.1007/s00040-004-0759-8.

23. Inward, D.J.G., Vogler, A.P., and Eggleton, P. (2007). A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology. Mol. Phylogenet. Evol. *44*, 953–967. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1055790307001662.

24. Legendre, F., Whiting, M.F., Bordereau, C., Cancello, E.M., Evans, T.A., and Grandcolas, P. (2008). The phylogeny of termites (Dictyoptera: Isoptera) based on mitochondrial and nuclear markers: Implications for the evolution of the worker and pseudergate castes, and foraging behaviors. Mol. Phylogenet. Evol. *48*, 615–627. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1055790308001929.

25. Legendre, F., Nel, A., Svenson, G.J., Robillard, T., Pellens, R., and Grandcolas, P. (2015). Phylogeny of Dictyoptera: dating the origin of cockroaches, praying mantises and termites with molecular data and controlled fossil evidence. PLoS One *10*, e0130127. Available at: https://dx.plos.org/10.1371/journal.pone.0130127.

26. Cameron, S.L., Lo, N., Bourguignon, T., Svenson, G.J., and Evans, T.A. (2012). A mitochondrial genome phylogeny of termites (Blattodea: Termitoidae): robust support for interfamilial relationships and molecular synapomorphies define major clades. Mol. Phylogenet. Evol. *65*, 163–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22683563 [Accessed September 26, 2014].

27. Bourguignon, T., Lo, N., Šobotník, J., Ho, S.Y.W., Iqbal, N., Coissac, E., Lee, M., Jendryka, M.M., Sillam-Dussès, D., Křížková, B., *et al.* (2017). Mitochondrial phylogenomics resolves the global spread of higher termites, ecosystem engineers of the tropics. Mol. Biol. Evol. *34*, 589–597. Available at: https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msw253.

28. Bourguignon, T., Lo, N., Šobotník, J., Sillam-Dussès, D., Roisin, Y., and Evans, T.A.

(2016). Oceanic dispersal, vicariance and human introduction shaped the modern distribution of the termites *Reticulitermes*, *Heterotermes* and *Coptotermes*. Proc. R. Soc. B Biol. Sci. *283*, 20160179. Available at: http://rspb.royalsocietypublishing.org/lookup/doi/10.1098/rspb.2016.0179.

29. Aanen, D.K., Eggleton, P., Rouland-Lefevre, C., Guldberg-Froslev, T., Rosendahl, S., and Boomsma, J.J. (2002). The evolution of fungus-growing termites and their mutualistic fungal symbionts. Proc. Natl. Acad. Sci. U. S. A. *99*, 14887–14892.

30. Wu, L.-W., Bourguignon, T., Šobotník, J., Wen, P., Liang, W.-R., and Li, H.-F. (2018). Phylogenetic position of the enigmatic termite family Stylotermitidae (Insecta : Blattodea). Invertebr. Syst. Available at: http://www.publish.csiro.au/?paper=IS17093.

31. Wang, M., Buček, A., Šobotník, J., Sillam-Dussès, D., Evans, T.A., Roisin, Y., Lo, N., and Bourguignon, T. (2019). Historical biogeography of the termite clade Rhinotermitinae (Blattodea: Isoptera). Mol. Phylogenet. Evol. *132*, 100–104. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1055790318303907.

32. Joly, S., McLenachan, P.A., and Lockhart, P.J. (2009). A statistical approach for distinguishing hybridization and incomplete lineage sorting. Am. Nat. *174*, E54–E70. Available at: https://www.journals.uchicago.edu/doi/10.1086/600082.

33. Buckley, T.R., Cordeiro, M., Marshall, D.C., and Simon, C. (2006). Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (*Maoricicada* Dugdale). Syst. Biol. *55*, 411–25. Available at: https://academic.oup.com/sysbio/article/55/3/411/1669040.

34. Meyer, B.S., Matschiner, M., and Salzburger, W. (2016). Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. Syst. Biol. *66*, 531–550. Available at: https://academic.oup.com/sysbio/article-lookup/doi/10.1093/sysbio/syw069.

35. Gatesy, J., and Springer, M.S. (2014). Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. *80*, 231–266. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1055790314002802.

36. Song, S., Liu, L., Edwards, S. V., and Wu, S. (2012). Resolving conflict in eutherian

mammal phylogeny using phylogenomics and the multispecies coalescent model. Proc. Natl. Acad. Sci. *109*, 14942–14947. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1211733109.

37. Evangelista, D.A., Wipfler, B., Béthoux, O., Donath, A., Fujita, M., Kohli, M.K., Legendre, F., Liu, S., Machida, R., Misof, B., *et al.* (2019). An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea). Proc. R. Soc. B Biol. Sci. *286*, 20182076. Available at: http://www.royalsocietypublishing.org/doi/10.1098/rspb.2018.2076.

38. Hunt, B.G., Wyder, S., Elango, N., Werren, J.H., Zdobnov, E.M., Yi, S. V, and Goodisman, M.A.D. (2010). Sociality is linked to rates of protein evolution in a highly social insect. Mol. Biol. Evol. *27*, 497–500. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20110264.

39. Grimaldi, D.A., Engel, M.S., and Krishna, K. (2008). The species of Isoptera (Insecta) from the early Cretaceous Crato Formation: a revision. Am. Museum Novit. *3626*, 1. Available at: http://www.bioone.org/perlserv/?request=get-abstract&doi=10.1206%2F616.1.

40. Krishna, K., and Grimaldi, D.A. (2003). The first Cretaceous Rhinotermitidae (Isoptera): a new species, genus, and subfamily in Burmese amber. Am. Museum Novit. *95*, 1–10.

41. Engel, M., Grimaldi, D., Nascimbene, P., and Singh, H. (2011). The termites of Early Eocene Cambay amber, with the earliest record of the Termitidae (Isoptera). Zookeys *148*, 105–123. Available at: http://zookeys.pensoft.net/articles.php?id=3004.

42. Engel, M.S., Grimaldi, D.A., and Krishna, K. (2009). Termites (Isoptera): Their phylogeny, classification, and rise to ecological dominance. Am. Museum Novit. *3650*, 1–27. Available at: http://www.bioone.org/doi/abs/10.1206/651.1.

43. Engel, M.S., Barden, P., Riccio, M.L., and Grimaldi, D.A. (2016). Morphologically Specialized Termite Castes and Advanced Sociality in the Early Cretaceous. Curr. Biol. *26*, 522–530. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0960982216000427.

44. Pagel, M. (1994). Detecting correlated evolution on phylogenies: a general method

for the comparative analysis of discrete characters. Proc. R. Soc. London. Ser. B Biol. Sci. *255*, 37–45. Available at: http://www.royalsocietypublishing.org/doi/10.1098/rspb.1994.0006.

45. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. Available at: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btu170.

46. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., *et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

47. Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. Available at: http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btv351.

48. Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., *et al.* (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience *1*, 18. Available at: https://academic.oup.com/gigascience/article-lookup/doi/10.1186/2047-217X-1-18.

49. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., *et al.* (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. *19*, 455–77. Available at: http://www.liebertpub.com/doi/10.1089/cmb.2012.0021.

50. Altenhoff, A.M., Škunca, N., Glover, N., Train, C.-M., Sueki, A., Piližota, I., Gori, K., Tomiczek, B., Müller, S., Redestig, H., *et al.* (2015). The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. Nucleic Acids Res. *43*, D240–D249. Available at: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gku1158.

51. Kriventseva, E. V., Tegenfeldt, F., Petty, T.J., Waterhouse, R.M., Simao, F.A., Pozdnyakov, I.A., Ioannidis, P., and Zdobnov, E.M. (2015). OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. Nucleic Acids

Res. *43*, D250–D256. Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku1220.

52. Petersen, M., Meusemann, K., Donath, A., Dowling, D., Liu, S., Peters, R.S., Podsiadlowski, L., Vasilikopoulos, A., Zhou, X., Misof, B., *et al.* (2017). Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. BMC Bioinformatics *18*, 111. Available at: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1529-8.

53. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. Available at: https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mst010.

54. Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. *34*, W609–W612. Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkl315.

55. Kück, P., and Meusemann, K. (2010). FASconCAT: Convenient handling of data matrices. Mol. Phylogenet. Evol. *56*, 1115–1118. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1055790310001909.

56. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274. Available at: https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msu300.

57. Chernomor, O., von Haeseler, A., and Minh, B.Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. Syst. Biol. *65*, 997–1008. Available at: https://academic.oup.com/sysbio/article-lookup/doi/10.1093/sysbio/syw037.

58. Lanfear, R., Calcott, B., Kainer, D., Mayer, C., and Stamatakis, A. (2014). Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. *14*, 82. Available at: http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-14-82.

59. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018).

UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522. Available at: https://academic.oup.com/mbe/article/35/2/518/4565479.

60. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589. Available at: http://www.nature.com/doifinder/10.1038/nmeth.4285.

61. Stöver, B.C., and Müller, K.F. (2010). TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics *11*, 7. Available at: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-7.

62. Zhang, C., Sayyari, E., and Mirarab, S. (2017). ASTRAL-III: increased scalability and impacts of contracting low support branches. In, pp. 53–75. Available at: http://link.springer.com/10.1007/978-3-319-67979-2_4.

63. Junier, T., and Zdobnov, E.M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics *26*, 1669–1670. Available at: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btq243.

64. Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. Syst. Biol. *51*, 492–508. Available at: http://academic.oup.com/sysbio/article/51/3/492/1616895.

65. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591. Available at: https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msm088.

66. Reis, M. dos, and Yang, Z. (2018). Bayesian molecular clock dating using genome-scale datasets. In Evolutionary Genomics: Statistical and computational methods, M. Anisimova, ed. (Humana Press).

67. Alroy, J. (2016). Fossilworks: gateway to the paleobiology database. Available at: http://fossilworks.org.

68. Vršanský, P. (2002). Origin and the early evolution of mantises. Amba Proj. *6*, 1–16.

69. Grimaldi, D. (2003). A revision of Cretaceous mantises and their relationships, including new taxa (Insecta: Dictyoptera: Mantodea). Am. Museum Novit. *3412*, 1–47. Available at: http://www.bioone.org/doi/abs/10.1206/0003-

0082%282003%29412%3C0001%3AAROCMA%3E2.0.CO%3B2.

70. Krishna, K., and Grimaldi, D. (2009). Diverse Rhinotermitidae and Termitidae (Isoptera) in Dominican amber. Am. Museum Novit. *3640*, 1–48. Available at: https://bioone.org/journals/american-museum-novitates/volume-2009/issue-3640/633.1/Diverse-Rhinotermitidae-and-Termitidae-Isoptera-in-Dominican-Amber/10.1206/633.1.full.

71. Krishna, K. (1996). New fossil species of termites of the subfamily Nasutitermitinae from Dominican and Mexican amber (Isoptera, Termitidae). Am. Museum Novit. *3176*, 1–13.

72. Bourguignon, T., Šobotník, J., Dahlsjö, C.A.L., and Roisin, Y. (2016). The soldierless Apicotermitinae: insights into a poorly known and ecologically dominant tropical taxon. Insectes Soc. *63*, 39–50. Available at: http://link.springer.com/10.1007/s00040-015-0446-y.

73. Schlemmermeyer, T., and Cancello, E.M. (2000). New fossil termite species: Dolichorhinotermes dominicanus from Dominican amber. Pap. Avulsos Zool. *41*, 303–311.

74. Engel, M.S., and Krishna, K. (2007). New Dolichorhinotermes from Ecuador and in Mexican amber (Isoptera: Rhinotermitidae). Am. Museum Novit. *2811*, 1–8.

75. Engel, M.S., Grimaldi, D.A., and Krishna, K. (2007). A synopsis of Baltic amber termites (Isoptera). Stuttgarter Beitrage zur Naturkd *372*, 1–20.

76. Emerson, A.E. (1971). Tertiary fossil species of the Rhinotermitidae (Isoptera), phylogeny of genera, and reciprocal phylogeny of associated Flagellata (Protozoa) and the Staphylinidae (Coleoptera). Bull. Am. Museum Nat. Hist. *146*, 243–304.

77. Charpentier, T. (1843). Über einige fossile Insecten aus Radoboj in Croatien. Novorum Actorum Acad. Caesareae Leopoldino-Carolinae Naturae Curiosorum *20*.

78. Snyder, T.E. (1949). Catalog of the termites (Isoptera) of the world. Smithson. Misc. Collect. *112*, 1–490.

79. Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217–223. Available at: http://doi.wiley.com/10.1111/j.2041-210X.2011.00169.x.