

Research



Cite this article: Arora J *et al.* 2023 Evidence of cospeciation between termites and their gut bacteria on a geological time scale.

Proc. R. Soc. B **290**: 20230619.

<https://doi.org/10.1098/rspb.2023.0619>

Received: 14 March 2023

Accepted: 24 May 2023

Subject Category:

Evolution

Subject Areas:

evolution

Keywords:

cophylogeny, endosymbionts, isoptera, metagenomics, vertical inheritance

Author for correspondence:

Thomas Bourguignon

e-mail: thomas.bourguignon@oist.jp

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.6673724>.

Evidence of cospeciation between termites and their gut bacteria on a geological time scale

Jigyasa Arora¹, Aleš Buček^{1,2}, Simon Hellemans¹, Tereza Beránková², Johanna Romero Arias², Brian L. Fisher^{3,4}, Crystal Clitheroe¹, Andreas Brune⁵, Yukihiro Kinjo^{1,6}, Jan Šobotník^{2,7} and Thomas Bourguignon^{1,2}

¹Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Okinawa 904-0495, Japan

²Faculty of Tropical AgriScience, Czech University of Life Sciences, Kamýcká 129, Suchdol, 165 00, Prague 6, Czech Republic

³Madagascar Biodiversity Center, Parc Botanique et Zoologique de Tsimbazaza, Antananarivo 101, Madagascar

⁴California Academy of Sciences, San Francisco, CA, USA

⁵Research Group Insect Gut Microbiology and Symbiosis, Max Planck Institute for Terrestrial Microbiology, Marburg, 35043, Germany

⁶College of Economics and Environmental Policy, Okinawa International University, 2-6-1 Ginowan, Ginowan, 901-2701, Okinawa, Japan

⁷Biology Centre, Czech Academy of Sciences, Ceske Budejovice, 37011, Czech Republic

ID JA, 0000-0003-4338-2953; AB, 0000-0001-9479-3141; SH, 0000-0003-1266-9134; JRA, 0000-0002-0859-4662; BLF, 0000-0002-4653-3270; AB, 0000-0002-2667-4391; YK, 0000-0002-3325-0327; JŠ, 0000-0002-8581-637X; TB, 0000-0002-4035-8977

Termites host diverse communities of gut microbes, including many bacterial lineages only found in this habitat. The bacteria endemic to termite guts are transmitted via two routes: a vertical route from parent colonies to daughter colonies and a horizontal route between colonies sometimes belonging to different termite species. The relative importance of both transmission routes in shaping the gut microbiota of termites remains unknown. Using bacterial marker genes derived from the gut metagenomes of 197 termites and one *Cryptocercus* cockroach, we show that bacteria endemic to termite guts are mostly transferred vertically. We identified 18 lineages of gut bacteria showing cophylogenetic patterns with termites over tens of millions of years. Horizontal transfer rates estimated for 16 bacterial lineages were within the range of those estimated for 15 mitochondrial genes, suggesting that horizontal transfers are uncommon and vertical transfers are the dominant transmission route in these lineages. Some of these associations probably date back more than 150 million years and are an order of magnitude older than the cophylogenetic patterns between mammalian hosts and their gut bacteria. Our results suggest that termites have cospeciated with their gut bacteria since first appearing in the geological record.

1. Introduction

Symbiotic associations with bacteria are pervasive across the animal tree of life [1]. Some of these associations involve partners that have continuously and reciprocally adapted to each other over extended evolutionary time scales, a phenomenon referred to as coevolution [2]. The coevolution between bacteria and their hosts is sometimes coupled with mechanisms of vertical transmission leading to cospeciation and cophylogenetic patterns, whereby the phylogenetic trees of both symbiotic partners show congruence in terms of topology and timing [3,4]. For example, many insects, such as aphids, cockroaches or whiteflies (e.g. [5–7]), and many marine invertebrates, such as vesicomyid clams or

catenulid flatworms (e.g. [8,9]), harbour intracellular bacterial endosymbionts with phylogenetic trees closely matching that of their hosts. The congruence between host and symbiont phylogenetic trees reflects the strict vertical transmission of intracellular endosymbionts from mother to eggs or embryos [10,11], which sometimes takes place over several hundred million years [5].

In contrast with maternally inherited intracellular endosymbionts, there are no clear examples of gut bacteria being vertically transmitted on hundred-million-year time scales. Cophylogenetic patterns between gut bacteria and their hosts have been identified primarily in obligate symbioses with highly specialized modes of symbiont transmission, such as the nutritional symbiont *Candidatus* Ishikawaella capsulata of plataspid stinkbugs [12] and the pectinolytic *Candidatus* Stammera of cassidine leaf beetles [13]. The rarity of cophylogenetic patterns between gut bacteria and their host may be linked to the difficulty of establishing stable vertical transmission routes, especially in species with limited social interactions. In social insects, such as bees, ants and termites, nest-mates experience frequent social contacts and often exchange gut fluid through trophallaxis, a behaviour providing a stable route of gut bacterial transmission across generations [11]. In consequence, some social insects, such as the corbiculate bees, present cophylogenetic patterns with their gut bacteria [14,15], indicating that sociality may lead to the coevolution of gut bacteria with their host at geological time scales.

Cophylogenetic analyses have rarely been performed for animals with complex bacterial gut microbiota, perhaps because many studies have relied upon 16S rRNA sequences, a marker that diverges at about 1% per 50 Myr [16]. Because of its slow rate of evolution, the 16S rRNA gene does not provide the taxonomic resolution required to resolve cophylogenetic patterns. Studies of the gut microbiota of humans and great apes successfully used protein-coding sequences to identify cophylogenetic patterns between several bacterial lineages and their mammalian hosts, suggesting that both partners have coevolved over the past 15 Myr [17,18]. This coevolution may have occurred because of the limited ability of symbionts to survive outside their host as they lost genes involved in key metabolic functions and developed specific oxygen and temperature requirements [18–20]. Here, we used protein-coding marker genes obtained from termite gut metagenomes to analyse the cophylogenetic patterns between termites and their gut bacteria.

Termites host unique gut microbial communities composed of bacteria, archaea and cellulolytic flagellates [21]. The gut flagellates have cospeciated with their hosts since their acquisition by the common ancestor of termites and their sister group, the wood-feeding roach *Cryptocercus* [22–24]. Numerous bacterial lineages occur ubiquitously in all termite species investigated but have never been found outside of termite guts [25,26]. These endemic bacteria are believed to be acquired via two transmission routes: vertical and horizontal [26]. The vertical route involves both colony founders (the king and the queen) and nest-mates that provide each other gut fluid through trophallaxis, ensuring the transmission of gut bacteria among family members and, ultimately, from parent colonies to daughter colonies [27]. The horizontal route involves the transfer of bacteria between unrelated colonies sometimes belonging to different termite species, or the acquisition of environmental bacteria. The relative importance of the vertical and horizontal transmission routes in shaping the gut microbiota of termites remains unknown.

In this study, we searched for evidence of cophylogeny between termites and their gut bacteria. We compared two phylogenetic trees of termites reconstructed using mitochondrial genomes and ultraconserved elements (UCEs), respectively, with phylogenetic trees of gut bacteria reconstructed using 10 independent universally occurring protein-coding marker genes [28]. Our study reveals that horizontal transfers between termite species may not be needed to explain the cophylogenetic patterns between termites and some of their endemic gut bacteria. Our results suggest that some bacterial lineages found in termite guts have been vertically transmitted over the past 150 Myr of termite evolution [29,30].

2. Material and methods

(a) Sample collection and metagenome analyses

We used the gut metagenomes of 141 termite samples and one sample of the cockroach *Cryptocercus kyebangensis* sequenced by Arora *et al.* [31] (electronic supplementary material, table S1). In addition, we sequenced 56 termite gut metagenomes for this study. All samples were preserved in RNA-later, stored for up to several weeks at room temperature, and subsequently stored at -80°C until DNA extraction. We extracted and sequenced DNA and assembled the metagenomes as described in Arora *et al.* [31].

Ten single-copy protein-coding marker genes were extracted from the assemblies using the mOTU software [28,32,33]. Genomes and metagenome-assembled genomes available in the Genome Taxonomy Database GTDB v. 95 [34] were downloaded, and the same 10 single-copy marker genes were extracted as described above.

The taxonomic annotation of the 10 marker genes extracted from termite gut metagenome assemblies was performed using the lowest common ancestor algorithm implemented in DIAMOND BLASTP [35] with $e\text{-value} \leq 1e-24$ as a threshold. The BLASTP search was performed against the GTDB database v.95 [34]. For downloaded genomes, we used the taxonomic annotation available from GTDB v. 95. The marker gene sequences from termite gut metagenomes and the GTDB database were analysed separately for every phylum. We reconstructed the phylogenetic tree of every phylum comprising more than 10 sequences.

(b) Reconstruction of marker gene phylogenetic trees

Sequences shorter than half the mean length of the marker gene were removed to improve the accuracy of phylogenetic reconstructions [36,37]. Protein sequences were aligned using MAFFT v. 7.305 with the *-auto* option [38]. Protein alignments were back-translated into their corresponding nucleotide alignments using PAL2NAL [39]. Aligned nucleotide sequences were converted into purines (R) and pyrimidines (Y) using BMGE v. 1.12 [40] to account for the variability of GC content observed across bacterial sequences. Maximum-likelihood (ML) phylogenetic trees were generated using these RY-recoded sequence alignments with IQ-TREE v. 1.6.12 [41]. We used the GTR2 + G + I model of binary state substitution. Node supports were assessed using the ultrafast bootstrap method [42] with the command *-bb 2000* for 2000 bootstrap replicates. The phylogenetic trees of every phylum were rooted using outgroup taxa selected from the bacterial tree of life [34,43]. The phylogenetic trees of archaeal and bacterial clades composed of sequences found exclusively in termite guts and represented by more than 10 termite species were extracted from the phylogenetic trees of each phylum. We refer to these trees, including sequences of termite gut bacteria exclusively, as termite-specific clades (TSCs). This procedure was followed for each marker gene.

(c) Phylogenetic reconstruction of termites using mitochondrial genomes

We reconstructed the phylogenetic tree of termites using mitochondrial genome sequences. Termite mitochondrial contigs longer than 5000 bp and more than 90% identical to the previously published whole mitochondrial genomes of termites [29,44–49] were identified using BLAST searches [50]. Complete or near-complete mitochondrial genomes were annotated using the MITOS web server [51]. We aligned the 13 protein-coding genes, two ribosomal RNA genes and 22 transfer RNA genes with MAFFT v. 7.305 [38]. All gene alignments were concatenated, and the third codon position of protein-coding genes was removed. The concatenated alignment was divided into four partitions: one for the first codon position of protein-coding genes, one for the second codon position of protein-coding genes, one for the combined transfer RNA genes and one for the combined ribosomal RNA genes. We reconstructed a Bayesian phylogenetic tree using BEAST v. 2.4.8 [52], following the approach described in Arora *et al.* [31]. We used an uncorrelated lognormal relaxed clock as a model of rate variation [53] and a birth–death process as a tree prior [54]. We used the nine fossil calibrations used by Arora *et al.* [31], which we implemented as exponential priors on node times. Sphaerotermitinae and Macrotermitinae were constrained to form a monophyletic group, as supported by phylogenetic trees based on transcriptomes and UCEs [30,55]. Similar constraints were applied to non-Styloptermitidae Neoisoptera, which were constrained to be monophyletic [31].

(d) Phylogenetic reconstruction of termites using ultraconserved elements

We extracted from each gut metagenome assembly termite UCEs, and their flanking 200 bp at both 5' and 3' ends, using PHYLUCE v. 1.6.6 [56] and LASTZ [57]. We used the termite-specific bait set targeting the 50 616 UCE loci described in Hellemans *et al.* [55] and followed the procedure described therein. The UCE dataset produced in this study (Contribution #3 to the Termite UCE Database available at: <https://github.com/oist/TER-UCE-DB/>) is available at the Dryad Digital Repository (<https://doi.org/10.5061/dryad.tmpg4f53w>). Loci were aligned with MAFFT [38], as implemented in `phyluce_align_seqcap_align`. Alignments were trimmed internally using `phyluce_align_get_gblocks_trimmed_alignments_from_untrimmed`, which implements Gblocks [58,59] with default parameters. UCE loci found in more than 57% of termite gut metagenomes were extracted with `phyluce_align_get_only_loci_with_min_taxa`. Of those, the 322 loci matching, at least partly, singly annotated exons from the draft genome of *Zootermopsis nevadensis* [60] were used for downstream analyses. The final supermatrix, composed of 322 UCE alignments, was obtained with `phyluce_align_format_nexus_files_for_raxml`. We carried out ML tree reconstruction on the supermatrix using IQ-TREE v. 1.6.12 with a GTR + G + I model of nucleotide substitution and 1000 ultrafast bootstrap replicates to assess branch supports [41,61].

(e) Matching termite-specific archaeal and bacterial clades across marker gene trees

The phylogenetic trees reconstructed with the marker gene coding for COG0552 (*ftsY*) were used as references. We attempted to link every TSC found in the phylogenetic trees reconstructed with COG0552 with their counterparts found in the phylogenetic trees reconstructed with the other nine marker genes. To do so, we searched the 198 gut metagenomes for contigs encompassing at least two of the 10 marker genes. The position of each marker gene sequence in their respective

phylogenetic trees was used to match TSCs across marker gene trees. We also used the 10 marker genes of the termite gut bacterial genomes found in the GTDB database. Of the 194425 genomes downloaded from the GTDB database, 37 were associated with termite guts.

(f) Cophylogenetic analyses

We used three approaches to test for cophylogeny between termites and TSCs. For the first approach, we used the R package PACo (Procrustean Approach to Cophylogeny) [62], which implements Procrustean superimposition to estimate the cophylogenetic signal between two phylogenies. The host and symbiont phylogenetic trees were converted into distance matrices using the `cophenetic()` function of the `vegan` R package [63]. The software was run using the `backtracking` method of randomization that conserves the overall degree of interactions between the two trees [64]. The second approach was the generalized Robinson–Foulds (RF) metric [65]. This method was implemented using the `ClusteringInfoDistance()` function of the `TreeDist` R package [65]. For the third approach, the host and symbiont phylogenetic trees were matched to find an optimal one-to-one map between branches using the method described by Nye *et al.* [66] and implemented in the `Nye-Similarity()` function of the `TreeDist` R package [65]. Because the two methods implemented in the `TreeDist` R package do not allow multiple symbiont tips in one host, each host tip was split into a number of tips of zero branch length equal to the number of archaeal and bacterial symbionts present in the metagenome corresponding to that given tip [67,68]. The strength of the cophylogenetic signal was computed from each cophylogenetic algorithm. Congruence between the host and symbiont trees was determined using 10 000 random permutations. We ran the cophylogenetic analyses on the phylogenetic trees reconstructed with mitochondrial genomes and UCEs.

We estimated the number of host transfer events for each TSC using the GeneRax software [69]. GeneRax is a ML-based method that reconciles the microbial gene tree with the host tree. It estimates rates of horizontal transfers within TSCs, the probability that a microbe is transferred from one host to a random host not ancestral to the donor host. We carried out each cophylogenetic analysis twice, once with the termite phylogenetic tree reconstructed with mitochondrial genomes and once with the tree reconstructed with UCEs. We compared the rates of transfers obtained for TSC trees with the rates of transfers calculated for the 13 protein-coding genes analysed without the third codon position and the two rRNA mitochondrial genes. Because mitochondrial genomes do not recombine, all mitochondrial genes have an identical evolutionary history and experienced no transfer. The positive rates of transfers found for mitochondrial gene trees reflect the uncertainties of phylogenetic reconstructions and provide a baseline for the estimated rates of horizontal transfer values in the absence of horizontal transfer. We consider that the evolutionary history of TSCs is predominantly explained by vertical transfers when their rates of horizontal transfers fall within the range of that found for mitochondrial genes.

3. Results and discussion

The sequences were derived from 197 termite gut metagenomes and one *Cryptocercus* metagenome combined with sequences from the GTDB database [34]. Our dataset comprises representatives of all termite families, spanning approximately 150 Myr of evolution, and the main lineages of Termitidae, which arose around 50 Ma [29,30]. It also includes samples of 30 species of *Microcerotermes*, a pantropical termitid genus that appeared around 20 Ma [45]. Therefore, our dataset

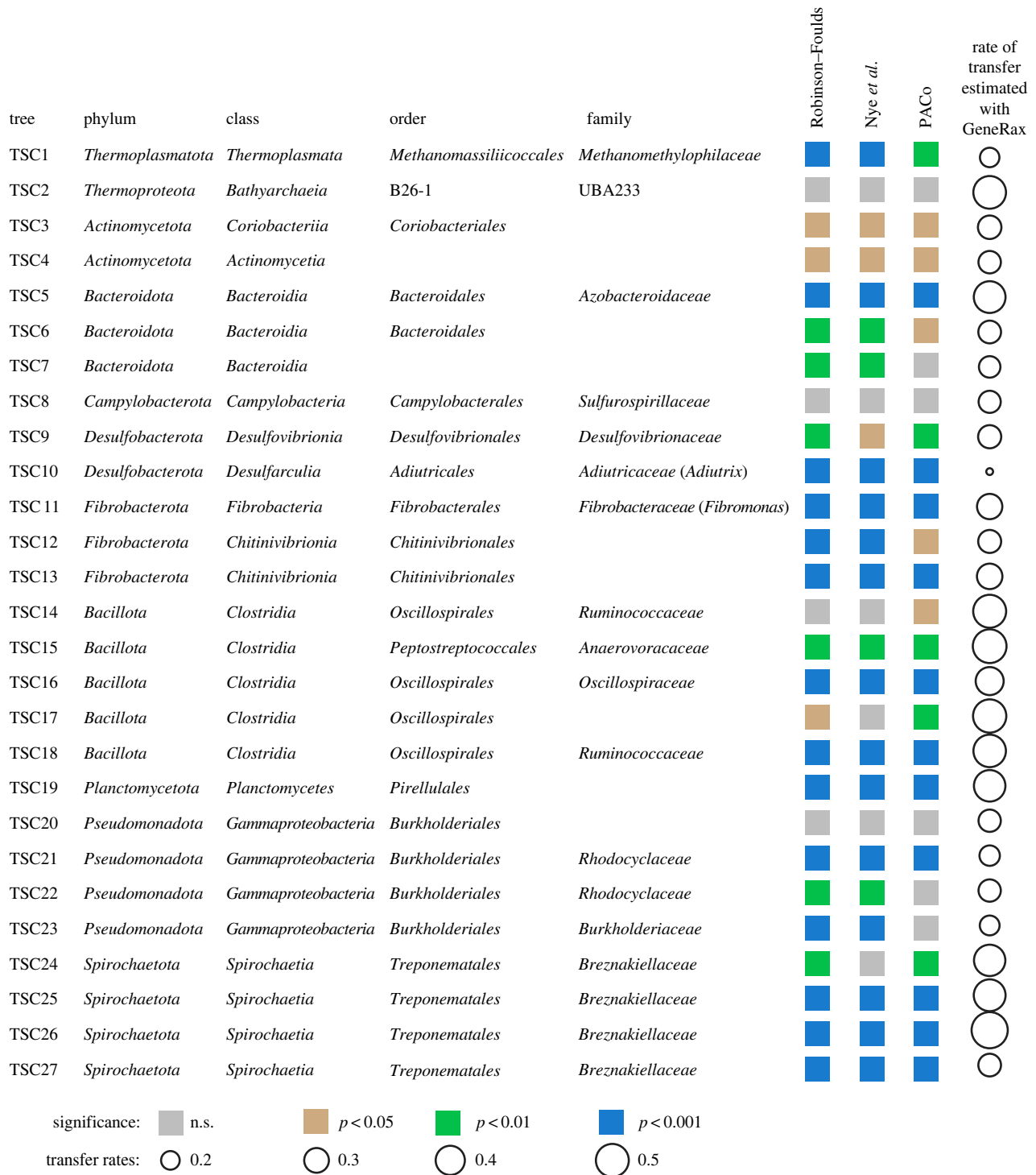


Figure 1. Results of the copylogenetic analyses performed on the marker gene COG0552 of 27 termite-specific archaeal and bacterial clades (TSCs). The copylogenetic analyses were performed with three different methods: PACo, the generalized RF metric, and the tree alignment algorithm described by Nye *et al.* [66]. The transfer rates were estimated using the ML method implemented in the GeneRax software.

captured both the intrageneric variations and ancient divergences of the termite hosts.

We reconstructed separate ML phylogenies for each marker gene and each bacterial and archaeal phylum. Then, we searched each tree for TSC composed exclusively of sequences associated with termites and represented in at least 10 termite species. We found between 8 and 34 TSCs per marker gene and selected *ftsY* (COG0552), which was represented by 2299 sequences forming 27 TSCs, as a reference marker gene. The 27 TSCs of COG0552 belonged to nine bacterial and two archaeal phyla. We examined the copylogenetic signal between each TSC and its termite

host using the termite phylogenetic trees reconstructed with mitochondrial and UCE data and three different methods: PACo [62], the generalized RF metric [65], and the tree alignment algorithm described by Nye *et al.* [66]. The results of the analyses performed on the two termite phylogenetic trees were almost identical (electronic supplementary material, table S2), indicating that our analyses were robust to the type of data used to reconstruct the host phylogenetic tree. We discuss the results obtained with the termite mitochondrial phylogenetic tree for simplicity. Eighteen of 27 TSCs showed significant copylogenetic signals with termites with all three methods (figure 1). We then inferred TSCs

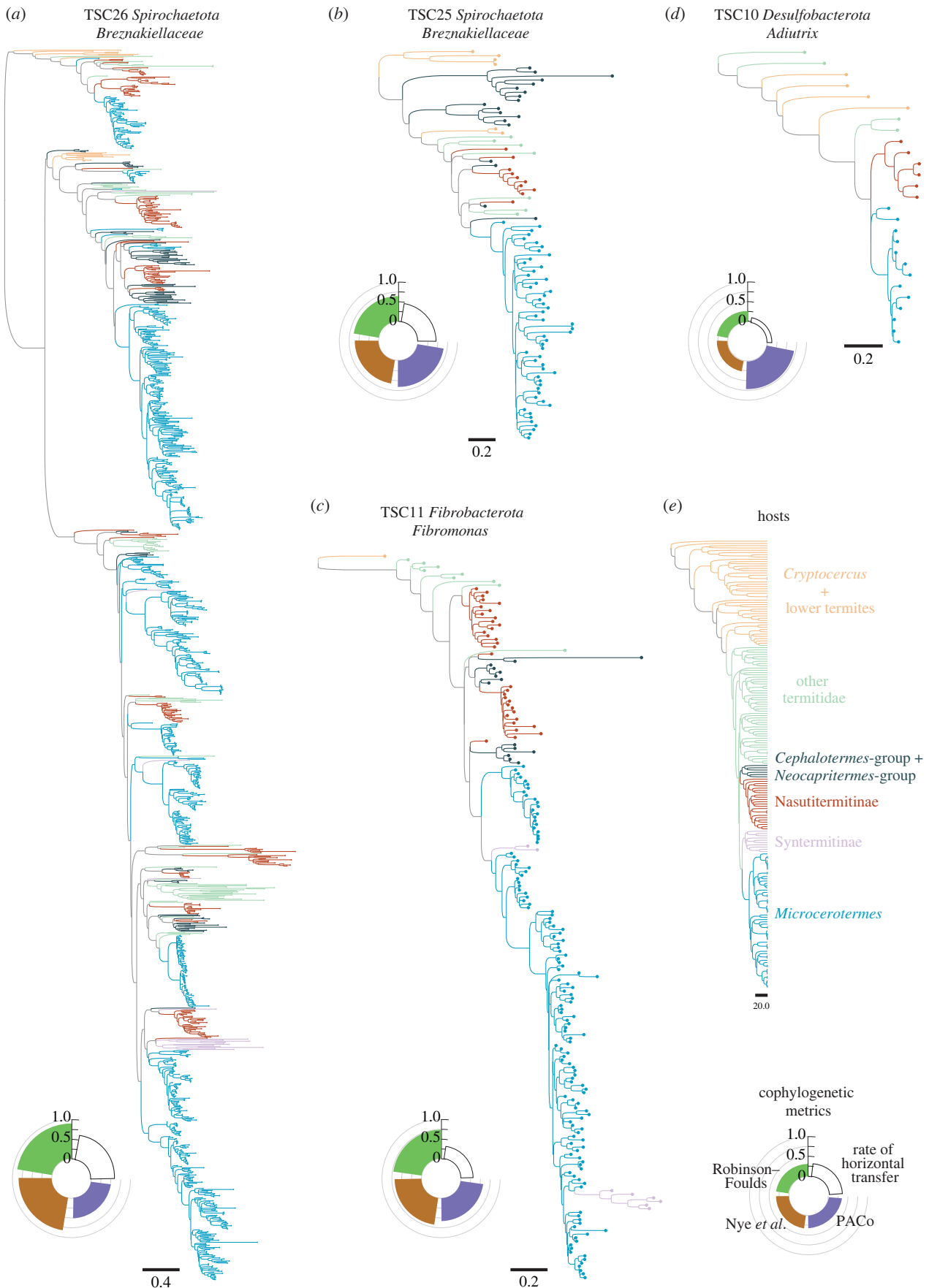


Figure 2. Selected phylogenetic trees of termite-specific bacterial clades (TSCs) showing strong copylogenetic signals with termites. The TSC phylogenetic trees were reconstructed with IQtree using the RY-recoded DNA sequence alignments from the marker gene COG0552. Phylogenetic trees of (a) the *Spirochaetota Breznakiellaceae* TSC26, (b) the *Spirochaetota Breznakiellaceae* TSC25, (c) the *Fibrobacterota Fibromonas* TSC11 and (d) the *Desulfobacterota Adiutrix* TSC10. (e) Phylogenetic tree of termites inferred from mitochondrial genomes. The diagrams below the phylogenetic trees indicate the results of the copylogenetic analyses and the estimation of the horizontal transfer rate. Scale bars represent substitutions per site.

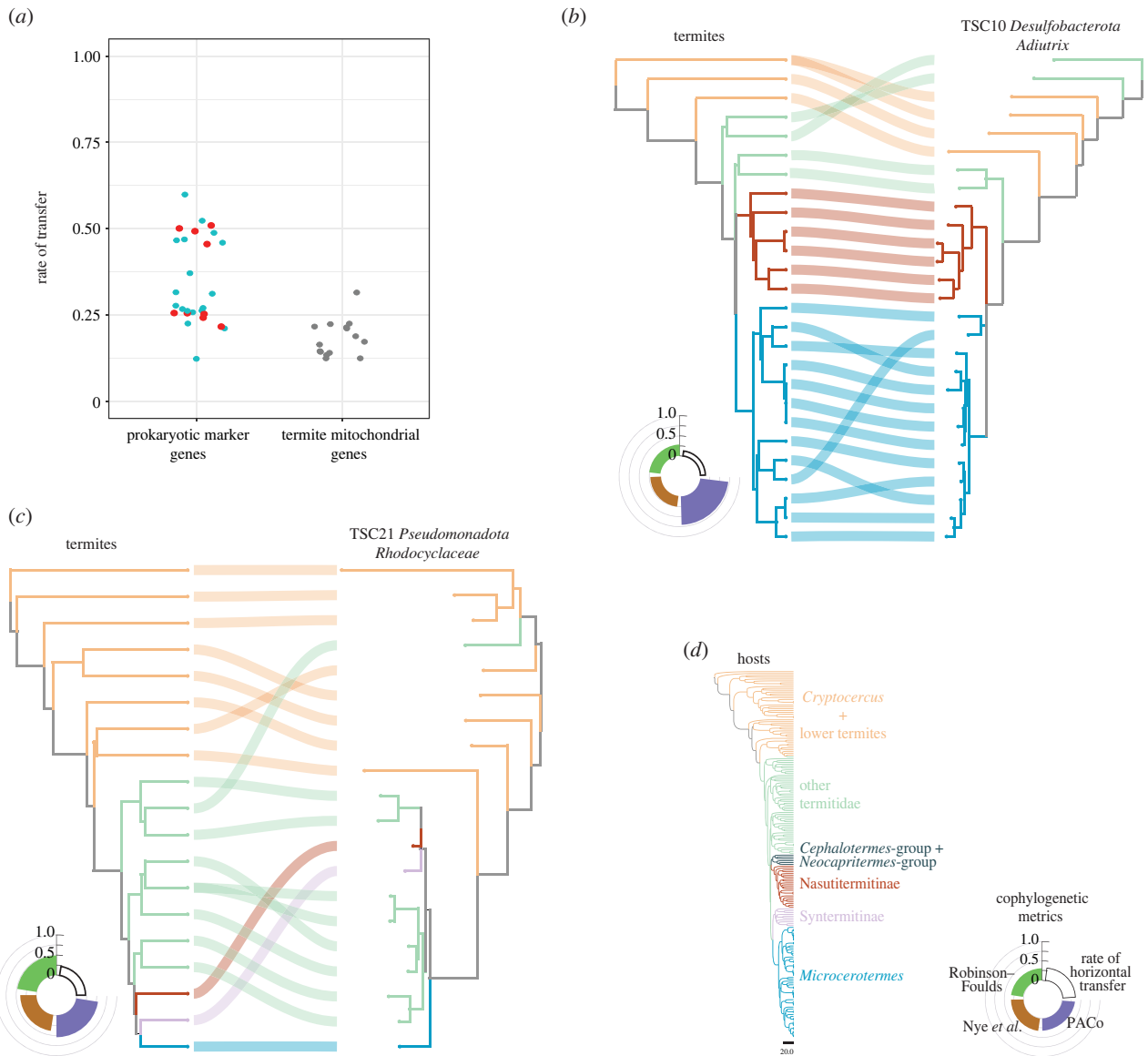


Figure 3. Rate of transfer and phylogenetic trees of some termite-specific bacterial clades (TSCs) showing strong cophylogenetic signals with termites. The TSC phylogenetic trees were reconstructed with IQtree using the RY-recoded DNA sequence alignments from the marker gene COG0552. (a) Rates of horizontal transfer of 27 TSCs were estimated using the ML method implemented in the GeneRax software. Red dots represent bacterial clades showing no significant cophylogenetic signals, and cyan dots represent bacterial clades showing significant cophylogenetic signals. Tanglegrams between termites and (b) the *Desulfobacterota* *Adiutrix* TSC10 and (c) the *Pseudomonadota* *Rhodocyclaceae* TSC21. (d) Phylogenetic tree of termites inferred from mitochondrial genomes. The diagrams below the phylogenetic trees indicate the results of the cophylogenetic analyses and the estimation of the horizontal transfer rate. The scale bar represents substitutions per site.

from the other nine marker genes and identified their correspondence to the COG0552 marker gene-based TSCs based on the physical linkage of marker genes on contigs (see Material and methods for additional details). Cophylogenetic analyses on corresponding TSCs from all marker genes yielded similar results (electronic supplementary material, table S2), indicating that the choice of reference marker gene did not influence the outcome of our analyses.

The TSCs with the strongest cophylogenetic signals included key components of the gut microbiota of termites. For example, the families *Ruminococcaceae* (phylum *Bacillota*, formerly *Firmicutes*) and *Breznakiellaceae* (phylum *Spirochaetota*), respectively, made up 16.5% and 20.0% of the 16S rRNA gene sequences found in a survey of 94 termite species [26]. *Breznakiellaceae* generally have a fermentative metabolism and include strains capable of reductive acetogenesis [70,71]. They have been isolated from the guts of cockroaches, suggesting that they were already present in the ancestor of termites and

their cockroach sister group, *Cryptocercidae* [71,72]. Therefore, TSCs with essential functions and a long history of association with termites show cophylogenetic signals.

In principle, the observed cophylogenetic signals between TSCs and their termite hosts could be caused by two different mechanisms: (i) vertical transmission of gut bacteria from parent colonies to daughter colonies, which is caused by the transmission of gut bacteria among family members and results in the coevolution of symbionts and hosts; (ii) limited horizontal transfers of gut bacteria among the diverging termite species due to geographical barriers, which would not require vertical transfers and results in allopatric speciation [3,4]. If vertical transfer were responsible for the cophylogenetic signals, it should give rise to bacterial lineages associated exclusively with specific termite clades and not shared with other sympatric termites. We indeed found such termite clade-specific lineages (TCSL) within many TSCs (figure 2). For example, we found several TCSLs belonging to the family

Breznakiellaceae, the genus *Fibromonas* (phylum *Fibrobacterota*) and the genus *Adiutrix* (phylum *Desulfobacterota*) that were associated exclusively with the densely sampled genus *Microcerotermes* (figure 2a–d). These TCSs were absent from the guts of other termites, including many species that are sympatric with *Microcerotermes*, demonstrating that some TCSs are endemic to the gut of specific termite genera, as previously hypothesized based on smaller datasets [73]. They were found in the guts of *Microcerotermes* species collected across four continents and six biogeographic realms, indicating that *Microcerotermes* dispersed worldwide with their specific gut bacteria. We also found TCSs associated with termite clades sampled less intensively. For example, a group of Nasutitermitinae that shares a common ancestor approximately 25 Mya and has been sampled across multiple continents hosted several TCSs belonging to the family *Breznakiellaceae* and the genus *Adiutrix* (figure 2a,b,d). These examples of the absence of horizontal transfer of bacteria between sympatric termites belonging to different clades indicate that allopatry is not required to maintain the association between termite clades and their symbiotic bacteria. Therefore, even if allopatric speciation of termites and TCSs likely occurred, TCSs are transmitted vertically from parent colonies to daughter colonies and possibly horizontally among related termite species forming a clade.

We next estimated the number of host transfer events for each TSC using the ML method implemented in the GeneRax software [69]. The estimated rates of transfer varied between 0.12 and 0.60 for TSCs showing cophylogenetic signals with termites (figure 3a). Note that the rates of transfer estimated with the UCE-based termite phylogenetic tree were almost identical, varying between 0.13 and 0.61 (electronic supplementary material, table S2). Notably, 16 TSCs had rates of transfer falling between 0.11 and 0.32, the range of rates of transfer estimated for each of the 13 protein-coding and two rRNA mitochondrial genes used in this study to build the phylogenetic tree of termites (figure 3a). Mitochondrial genes are expected to experience no transfer and have an identical evolutionary history, providing a baseline for estimated rates of transfer values obtained for genes expected to experience no horizontal transfer. While these results do not prove the absence of horizontal transfers, they suggest that the cophylogenetic patterns observed between some TSCs and termites may not involve any horizontal transfers. Cophylogenetic patterns would be obfuscated by bacterial extinction (or insufficient sequencing depth, from which it cannot be distinguished) and speciation taking place within non-speciating termite hosts [4].

Several TSCs, less speciose than *Breznakiellaceae* and *Fibromonas*, depicted patterns of cophylogeny across large parts of the termite phylogenetic tree (figure 3). For example, the phylogenetic tree of the genus *Adiutrix* found in the termite sister group Cryptoceridae, three families of termites, and across

Termitidae, was highly congruent with the phylogenetic tree of termites (figure 3b). The phylogenetic tree of the family *Rhodocyclaceae* (phylum *Pseudomonadota*, formerly *Proteobacteria*) (figure 3c) is another example of a clade showing significant cophylogenetic signal with termites. We interpret these cophylogenetic patterns between termites and some of their gut bacterial symbionts as evidence of coevolution with vertical transmission taking place over several tens of millions of years.

4. Conclusion

We identified the oldest known cophylogenetic patterns between animals and their gut bacteria. They involve multiple bacterial lineages and their termite hosts and span tens of millions of years—some may even trace back to the first appearance of termites around 150 Ma. These findings substantiate previous claims of coevolution between termites and their gut microbiota [74] and provide concrete evidence that proctodeal trophallaxis, a social behaviour in which nest-mates exchange droplets of hindgut contents [27], indeed serves as a stable vertical transmission route over geological time scales.

Data accessibility. Raw sequence data generated in this study are available in two MGRAST projects (<https://www.mg-rast.org/mgmain.html?mgpage=project&project=mgp101108> and <https://www.mg-rast.org/mgmain.html?mgpage=metazen2&project=mgp84199>) (see electronic supplementary material, table S1 for individual IDs). The single-copy marker gene sequences extracted in this study and the 13 phylogenetic trees of prokaryotic phyla reconstructed with COG0552 marker genes (Newick format) are available on Figshare (<https://figshare.com/account/home#/projects/134324>). The mitochondrial genomes sequenced in this study are available on GenBank (see electronic supplementary material, table S1). The UCE sequences are available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.tmpg4f53w> [75].

Authors' contributions. J.A.: formal analysis, investigation and writing—original draft; A.B.: formal analysis, investigation, methodology and writing—review and editing; S.H.: formal analysis, investigation, methodology and writing—review and editing; T.Be.: investigation and visualization; J.R.A.: investigation and visualization; B.L.F.: resources; C.C.: investigation; A.B.: writing—review and editing; Y.K.: investigation and methodology; J.Š.: resources; T.Bo.: conceptualization, supervision and writing—original draft.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Fundings. This work was supported by the Czech Science Foundation (project No. 20–20548S), the project IGA No. 20223112 from the Faculty of Tropical AgriSciences of the Czech University of Life Sciences Prague, and the Japan Society for the Promotion of Science through a DC2 graduate student fellowship awarded to J.A.

Acknowledgements. We thank the Data Analysis Section of the Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan, for assistance with the OIST computing cluster.

References

- McFall-Ngai M *et al.* 2013 Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl Acad. Sci. USA* **110**, 3229–3236. (doi:10.1073/pnas.1218525110)
- Janzen DH. 1980 When is it coevolution? *Evolution* **34**, 611–612. (doi:10.2307/2408229)
- de Vienne DM, Refrégier G, López-Villavicencio M, Tellier A, Hood ME, Giraud T. 2013 Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytol.* **198**, 347–385. (doi:10.1111/nph.12150)
- Groussin M, Mazel F, Alm EJ. 2020 Co-evolution and co-speciation of host-gut bacteria systems.

- Cell Host Microbe* **28**, 12–22. (doi:10.1016/j.chom.2020.06.013)
5. Moran NA, McCutcheon JP, Nakabachi A. 2008 Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* **42**, 165–190. (doi:10.1146/annurev.genet.41.110306.130119)
 6. Joussetin E, Desdevises Y, Coeur d'acier A. 2009 Fine-scale cospeciation between *Brachycaudus* and *Buchnera aphidicola*: bacterial genome helps define species and evolutionary relationships in aphids. *Proc. R. Soc. B* **276**, 187–196. (doi:10.1098/rspb.2008.0679)
 7. Kinjo Y, Lo N, Martín PV, Tokuda G, Pigolotti S, Bourguignon T. 2021 Enhanced mutation rate, relaxed selection, and the 'domino effect' are associated with gene loss in *Blattabacterium*, a cockroach endosymbiont. *Mol. Biol. Evol.* **38**, 3820–3831. (doi:10.1093/molbev/msab159)
 8. Peek AS, Feldman RA, Lutz RA, Vrijenhoek RC. 1998 Cospeciation of chemoautotrophic bacteria and deep sea clams. *Proc. Natl Acad. Sci. USA* **95**, 9962–9966. (doi:10.1073/pnas.95.17.9962)
 9. Gruber-Vodicka HR *et al.* 2011 *Paracatenula*, an ancient symbiosis between thiotrophic *Alphaproteobacteria* and catenulid flatworms. *Proc. Natl Acad. Sci. USA* **108**, 12 078–12 083. (doi:10.1073/pnas.1105347108)
 10. Bright M, Bulgheresi S. 2010 A complex journey: transmission of microbial symbionts. *Nat. Rev. Microbiol.* **8**, 218–230. (doi:10.1038/nrmicro2262)
 11. Onchuru TO, Javier Martinez A, Ingham CS, Kaltenpoth M. 2018 Transmission of mutualistic bacteria in social and gregarious insects. *Curr. Opin. Insect Sci.* **28**, 50–58. (doi:10.1016/j.cois.2018.05.002)
 12. Hosokawa T, Kikuchi Y, Nikoh N, Shimada M, Fukatsu T. 2006 Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. *PLoS Biol.* **4**, e337. (doi:10.1371/journal.pbio.0040337)
 13. Salem H *et al.* 2020 Symbiont digestive range reflects host plant breadth in herbivorous beetles. *Curr. Biol.* **30**, 2875–2886. (doi:10.1016/j.cub.2020.05.043)
 14. Koch H, Abrol DP, Li J, Schmid-Hempel P. 2013 Diversity and evolutionary patterns of bacterial gut associates of corbiculate bees. *Mol. Ecol.* **22**, 2028–2044. (doi:10.1111/mec.12209)
 15. Kwong WK, Medina LA, Koch H, Sing KW, Soh EJ, Ascher JS, Jaffé R, Moran NA. 2017 Dynamic microbiome evolution in social bees. *Sci. Adv.* **3**, e1600513. (doi:10.1126/sciadv.1600513)
 16. Ochman H, Elwyn S, Moran NA. 1999 Calibrating bacterial evolution. *Proc. Natl Acad. Sci. USA* **96**, 12 638–12 643. (doi:10.1073/pnas.96.22.12638)
 17. Moeller AH *et al.* 2016 Cospeciation of gut microbiota with hominids. *Science* **353**, 380–382. (doi:10.1126/science.aaf3951)
 18. Suzuki TA *et al.* 2022 Codiversification of gut microbiota with humans. *Science* **377**, 1328–1332. (doi:10.1126/science.abm7759)
 19. Browne HP *et al.* 2021 Host adaptation in gut Firmicutes is associated with sporulation loss and altered transmission cycle. *Genome Biol.* **22**, 204. (doi:10.1186/s13059-021-02428-6)
 20. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. 2019 New insights from uncultivated genomes of the global human gut microbiome. *Nature* **568**, 505–510. (doi:10.1038/s41586-019-1058-x)
 21. Brune A. 2014 Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* **12**, 168–180. (doi:10.1038/nrmicro3182)
 22. Noda S *et al.* 2007 Cospeciation in the triplex symbiosis of termite gut protists (*Pseudotriconympha* spp.), their hosts, and their bacterial endosymbionts. *Mol. Ecol.* **16**, 1257–1266. (doi:10.1111/j.1365-294X.2006.03219.x)
 23. Ohkuma M, Noda H, Hongoh Y, Nalepa CA, Inoue T. 2009 Inheritance and diversification of symbiotic trichonymphid flagellates from a common ancestor of termites and the cockroach *Cryptocercus*. *Proc. R. Soc. B* **276**, 239–245. (doi:10.1098/rspb.2008.1094)
 24. Ohkuma M, Brune A. 2011 Diversity, structure, and evolution of the termite gut microbial community. In *Biology of termites: a modern synthesis* (eds DE Bignell, Y Roisin, N Lo). Berlin, Germany: Springer.
 25. Mikaelyan A, Dietrich C, Köhler T, Poulsen M, Sillam-Dussès D, Brune A. 2015 Diet is the primary determinant of bacterial community structure in the guts of higher termites. *Mol. Ecol.* **24**, 5284–5295. (doi:10.1111/mec.13376)
 26. Bourguignon T, Lo N, Dietrich C, Šobotník J, Sidek S, Roisin Y, Brune A, Evans TA. 2018 Rampant host switching shaped the termite gut microbiome. *Curr. Biol.* **28**, 649–654. (doi:10.1016/j.cub.2018.01.035)
 27. Nalepa CA, Bignell DE, Bandi C. 2001 Detritivory, coprophagy, and the evolution of digestive mutualisms in Dictyoptera. *Insectes Soc.* **48**, 194–201. (doi:10.1007/PL00001767)
 28. Sunagawa S *et al.* 2013 Metagenomic species profiling using universal phylogenetic marker genes. *Nat. Methods* **10**, 1196–1199. (doi:10.1038/nmeth.2693)
 29. Bourguignon T *et al.* 2015 The evolutionary history of termites as inferred from 66 mitochondrial genomes. *Mol. Biol. Evol.* **32**, 406–421. (doi:10.1093/molbev/msu308)
 30. Bucek A *et al.* 2019 Evolution of termite symbiosis informed by transcriptome-based phylogenies. *Curr. Biol.* **29**, 3728–3734. (doi:10.1016/j.cub.2019.08.076)
 31. Arora J *et al.* 2022 The functional evolution of termite gut microbiota. *Microbiome* **10**, 78. (doi:10.1186/s40168-022-01258-3)
 32. Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM. 2007 Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**, 1449–1452. (doi:10.1126/science.1147112)
 33. Milanese A *et al.* 2019 Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat. Commun.* **10**, 1014. (doi:10.1038/s41467-019-08844-4)
 34. Parks DH, Chuvpochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. 2020 A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol.* **38**, 1079–1086. (doi:10.1038/s41587-020-0501-8)
 35. Buchfink B, Xie C, Huson DH. 2015 Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60. (doi:10.1038/nmeth.3176)
 36. Wiens JJ, Chippindale PT, Hillis DM. 2003 When are phylogenetic analyses misled by convergence? A case study in Texas cave salamanders. *Syst. Biol.* **52**, 501–514. (doi:10.1080/10635150309320)
 37. von Mering C, Hugenholtz P, Raes J, Tringe SG, Doerks T, Jensen LJ, Ward N, Bork P. 2007 Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* **315**, 1126–1130. (doi:10.1126/science.1133420)
 38. Katoh K, Standley DM. 2013 MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
 39. Suyama M, Torrents D, Bork P. 2006 PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612. (doi:10.1093/nar/gkl315)
 40. Criscuolo A, Gribaldo S. 2010 BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210. (doi:10.1186/1471-2148-10-210)
 41. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2014 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)
 42. Minh BQ, Nguyen MAT, von Haeseler A. 2013 Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195. (doi:10.1093/molbev/mst024)
 43. Parks DH *et al.* 2017 Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533–1542. (doi:10.1038/s41564-017-0012-7)
 44. Bourguignon T, Lo N, Šobotník J, Sillam-Dussès D, Roisin Y, Evans TA. 2016 Oceanic dispersal, vicariance and human introduction shaped the modern distribution of the termites *Reticulitermes*, *Heterotermes* and *Captotermes*. *Proc. R. Soc. B* **283**, 20160179. (doi:10.1098/rspb.2016.0179)
 45. Bourguignon T *et al.* 2017 Mitochondrial phylogenomics resolves the global spread of higher termites, ecosystem engineers of the tropics. *Mol. Biol. Evol.* **34**, 589–597. (doi:10.1093/molbev/msw253)
 46. Bucek A *et al.* 2022 Molecular phylogeny reveals the past transoceanic voyages of drywood termites (Isoptera, Kalotermitidae). *Mol. Biol. Evol.* **39**, msac093. (doi:10.1093/molbev/msac093)
 47. Wang M, Bucek A, Šobotník J, Sillam-Dussès D, Evans TA, Roisin Y, Lo N, Bourguignon T. 2019 Historical biogeography of the termite clade

- Rhinotermitinae (Blattodea: Isoptera). *Mol. Phylogenet. Evol.* **132**, 100–104. (doi:10.1016/j.ympev.2018.11.005)
48. Wang M *et al.* 2022 Phylogeny, biogeography and classification of Teletisoptera (Blattaria: Isoptera). *Syst. Entomol.* **47**, 581–590. (doi:10.1111/syen.12548)
49. Wang M *et al.* 2023 Neoisoptera repeatedly colonised Madagascar after the Middle Miocene climatic optimum. *Ecography* **2023**, e06463. (doi:10.1111/ecog.06463)
50. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410. (doi:10.1016/S0022-2836(05)80360-2)
51. Bernt M *et al.* 2013 MITOS: improved *de novo* metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* **69**, 313–319. (doi:10.1016/j.ympev.2012.08.023)
52. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018 Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016. (doi:10.1093/ve/vey016)
53. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006 Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88. (doi:10.1371/journal.pbio.0040088)
54. Gernhard T. 2008 The conditioned reconstructed process. *J. Theor. Biol.* **253**, 769–778. (doi:10.1016/j.jtbi.2008.04.005)
55. Hellemans S, Wang M, Hasegawa N, Šobotník J, Scheffrahn RH, Bourguignon T. 2022 Using ultraconserved elements to reconstruct the termite tree of life. *Mol. Phylogenet. Evol.* **173**, 107520. (doi:10.1016/j.ympev.2022.107520)
56. Faircloth BC. 2016 PHYLUCe is a software package for the analysis of conserved genomic loci. *Bioinformatics* **32**, 786–788. (doi:10.1093/bioinformatics/btv646)
57. Harris RS. 2007 Improved pairwise alignment of genomic DNA. Ph.D. dissertation, The Pennsylvania State University.
58. Castresana J. 2000 Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552. (doi:10.1093/oxfordjournals.molbev.a026334)
59. Talavera G, Castresana J. 2007 Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577. (doi:10.1080/10635150701472164)
60. Terrapon N *et al.* 2014 Molecular traces of alternative social organization in a termite genome. *Nat. Commun.* **5**, 3636. (doi:10.1038/ncomms4636)
61. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018 UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522. (doi:10.1093/molbev/msx281)
62. Balbuena JA, Míguez-Lozano R, Blasco-Costa I. 2013 PACo: a novel procrustes application to cophylogenetic analysis. *PLoS ONE* **8**, e61048. (doi:10.1371/journal.pone.0061048)
63. Oksanen J *et al.* 2015 Vegan: community ecology package. R package version 2.2-0. See <http://CRAN.Rproject.org/package=vegan>.
64. Hutchinson MC, Cagua EF, Balbuena JA, Stouffer DB, Poisot T. 2017 PACo: implementing Procrustean Approach to Cophylogeny in R. *Methods Ecol. Evol.* **8**, 932–940. (doi:10.1111/2041-210X.12736)
65. Smith MR. 2020 Information theoretic generalized Robinson–Foulds metrics for comparing phylogenetic trees. *Bioinformatics* **36**, 5007–5013. (doi:10.1093/bioinformatics/btaa614)
66. Nye TMW, Liò P, Gilks WR. 2006 A novel algorithm and web-based tool for comparing two alternative phylogenetic trees. *Bioinformatics* **22**, 117–119. (doi:10.1093/bioinformatics/bti720)
67. Perez-Lamarque B, Morlon H. 2019 Characterizing symbiont inheritance during host–microbiota evolution: application to the great apes gut microbiota. *Mol. Ecol. Resour.* **19**, 1659–1671. (doi:10.1111/1755-0998.13063)
68. Satler JD, Herre EA, Jandér KC, Eaton DA, Machado CA, Heath TA, Nason JD. 2019 Inferring processes of coevolutionary diversification in a community of Panamanian strangler figs and associated pollinating wasps. *Evolution* **73**, 2295–2311. (doi:10.1111/evo.13809)
69. Morel B, Kozlov AM, Stamatakis A, Szöllösi GJ. 2020 GeneRax: a tool for species-tree-aware maximum likelihood-based gene family tree inference under gene duplication, transfer, and loss. *Mol. Biol. Evol.* **37**, 2763–2774. (doi:10.1093/molbev/msaa141)
70. Leadbetter JR, Schmidt TM, Graber JR, Breznak JA. 1999 Acetogenesis from H₂ plus CO₂ by spirochetes from termite guts. *Science* **283**, 686–689. (doi:10.1126/science.283.5402.686)
71. Song Y, Hervé V, Radek R, Pfeiffer F, Zheng H, Brune A. 2021 Characterization and phylogenomic analysis of *Breznakiella homolactica* gen. nov. sp. nov. indicate that termite gut treponemes evolved from non-acetogenic spirochetes in cockroaches. *Environ. Microbiol.* **23**, 4228–4245. (doi:10.1111/1462-2920.15600)
72. Brune A, Song Y, Oren A, Paster BJ. 2022 A new family for ‘termite gut treponemes’: description of *Breznakiellaceae* fam. nov., *Gracilinema caldarium* gen. nov., comb. nov., *Leadbettera azotonutricia* gen. nov., comb. nov., *Helmutkoenigia isoptericoles* gen. nov., comb. nov., and *Zuelzera stenostrepta* gen. nov., comb. nov., and proposal of *Rectinemataceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* **72**, 005439. (doi:10.1099/ijsem.0.005439)
73. Hongoh Y, Deevong P, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Vongkaluang C, Noparatnaraporn N, Kudo T. 2005 Intra- and interspecific comparisons of bacterial diversity and community structure support coevolution of gut microbiota and termite host. *Appl. Environ. Microbiol.* **71**, 6590–6599. (doi:10.1128/AEM.71.11.6590-6599.2005)
74. Brune A, Dietrich C. 2015 The gut microbiota of termites: digesting the diversity in the light of ecology and evolution. *Annu. Rev. Microbiol.* **69**, 145–166. (doi:10.1146/annurev-micro-092412-155715)
75. Arora J *et al.* 2023 Data from: Evidence of cospeciation between termites and their gut bacteria on a geological time scale. Dryad Digital Repository. (doi:10.5061/dryad.tmpg4f53w)
76. Arora J *et al.* 2023 Evidence of cospeciation between termites and their gut bacteria on a geological time scale. Figshare. (doi:10.6084/m9.figshare.c.6673724)