Rampant host-switching shaped the termite gut microbiome

Thomas Bourguignon,[1,2,3,9,*] Nathan Lo,[1,9,10,*] Carsten Dietrich,[4,5] Jan Šobotník,[2] Sarah Sidek,[6] Yves Roisin,[7] Andreas Brune,[4] Theodore A. Evans[6,8]

[1]School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia

[2]Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Prague, Czech Republic

[3]Okinawa Institute of Science & Technology Graduate University, 1919–1 Tancha, Onna-son, Okinawa, 904–0495, Japan

[4]Department of Biogeochemistry, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

[5]Strategy and Innovation Technology Center, Siemens Healthcare GmbH, Erlangen, Germany

[6]Department of Biological Sciences, National University of Singapore, 117543 Singapore, Singapore

[7]Evolutionary Biology and Ecology, Université Libre de Bruxelles, Belgium

[8]School of Animal Biology, University of Western Australia, Perth WA 6009, Australia.

[9]These authors contributed equally

[10]Lead Contact

*Correspondence:

nathan.lo@sydney.edu.au

thomas.bourgui@gmail.com

SUMMARY

The gut microbiota of animals exert major effects on host biology [1]. Although horizontal transfer is generally considered the prevalent route for the acquisition of gut bacteria in mammals [2], some bacterial lineages have co-speciated with their hosts on timescales of several million years [3]. Termites harbor a complex gut microbiota, and their advanced social behavior provides the potential for long-term vertical symbiont transmission, and co-evolution of gut symbionts and host [4-6]. Despite clear evolutionary patterns in the gut microbiota of termites [7], a consensus on how microbial communities were assembled during termite diversification has yet to be reached. Although some studies have concluded that vertical transmission has played a major role [8, 9], others indicate that diet and gut microenvironment have been the primary determinants shaping microbial communities in termite guts [7, 10]. To address this issue, we examined the gut microbiota of 94 termite species, through 16S rRNA metabarcoding. We analyzed the phylogeny of 211 bacterial lineages obtained from termite guts, including their closest relatives from other environments, which were identified using BLAST. The results provided strong evidence for rampant horizontal transfer of gut bacteria between termite host lineages. While the majority of termite-derived phylotypes formed large monophyletic groups, indicating high levels of niche specialization, numerous other clades were interspersed with bacterial lineages from the guts of other animals. Our results indicate that 'mixed-mode' transmission, which combines colony-to-offspring vertical transmission with horizontal colony-to-colony transfer, has been the primary driving force shaping the gut microbiota of termites.

RESULTS AND DISCUSSION

The termite gut microbiome is among the most complex of any animal group. The hindguts of termites harbor upwards of 1000 species of bacteria and archaea, and, in all lower termites, a unique assemblage of flagellate protists [4, 5, 11]. This symbiosis has enabled termites to digest lignocellulose, to diversify their food source from the ancestral state of wood into leaf litter, grass, humus, and soil, and to achieve ecological dominance across tropical and subtropical regions of the globe [12]. The function of the termite gut microbiota, and how it was assembled over the ~150 million years of termite evolution [13], has interested biologists for over a century. Whether bacterial lineages present in termite guts have been acquired primarily through vertical inheritance (*i.e.* colony to offspring), or via horizontal acquisition from the environment, is considered to be a key unresolved question [14].

Recent metabarcoding studies have relied primarily on comparisons of community profiles between termite species to investigate the evolution of the microbiota [10, 14-15]. These studies have made only limited use of direct phylogenetic comparisons of individual termite-derived bacterial lineages with each other, and with phylotypes from other (non-termite) environments. Moreover, taxon sampling used in these studies (<20 species in each case) was highly biased towards termite species from particular geographic regions and diet groups. Consequently, several major lineages of termites have not yet been examined. To address these issues, we undertook the most extensive metabarcoding study of termite gut microbes to date, obtaining bacterial profiles from 94 termite species collected across four continents, including 77 species from the ecologically dominant higher termites (family Termitidae). This represents an increase in taxon sampling of more than 4-fold compared with previous studies, and provides unprecedented power to investigate the evolution of the termite microbiota. We used a novel approach involving phylogenetic comparison of each identified genus-level lineage with related environmental sequences derived from exhaustive BLAST searches.

We obtained an average of 11,509 high-quality sequences from the V3-V4 region of the bacterial 16S rRNA (~450 bp) for each of the 94 samples (Table S1). We independently clustered the sequences of each library into

operational taxonomic units (OTUs) with a distance below 6%, from which we removed OTUs represented by less than five sequences (independent analyses using 3% OTUs did not significantly alter our results; data not shown). We then selected one reference sequence from each 6% OTU, and pooled them into a single dataset from which we produced groups of 12% sequence dissimilarity that we will refer to as genus-level bacterial lineages. We used 12% dissimilarity rather than the more commonly used 8% to avoid artificial splitting of large genus-level clusters that were abundantly represented in our dataset. For example, *Treponema* cluster I, *Endomicrobium*, and certain clades of Ruminococcaceae and Lachnospiraceae make up about 45% of the reads and have levels of dissimilarity that exceed 12%. From the 622 genus-level bacterial lineages, we selected 211 lineages that were represented by more than 10 OTUs (82.7% of the total reads) for downstream analyses (Table S2). For each group, we carried out BLAST analyses that specifically excluded matches with termite-derived sequences in order to target closely related environmental sequences in public databases. Phylogenetic comparisons of the termite-derived phylotypes with their closest relatives from other environments were then performed for each of the 211 groups using minimum evolution criteria in FastTree [16] (Data S1).

We classified the 211 trees generated in our analyses into three broad categories (see Figure S1). Category 1 represented trees in which ≥30% of termite-derived sequences formed a monophyletic group (Figures 1A-B; in some cases, multiple clades, each containing ≥30% of the termite derived sequences, were recovered within a single tree). Category 1 comprised 62% of all trees and made up 48.3% of the reads (Table S2). While several bacterial taxa of category 1 trees were encountered in all termites, others were restricted to particular host lineages (Table S2).

In many cases, termite-specific clades within category 1 trees had a sister group relationship to clades containing bacterial sequences derived from vertebrate or invertebrate guts (*e.g.* trees 21, 23, 76; Data S1). In other cases, termite-specific clades were sister to bacterial taxa from a variety of environments (soil, agricultural or industrial processes, marine environments). A large number of taxa from category 1 trees represented bacterial families

that are typically encountered in the intestinal tract of vertebrates as well as termite guts, indicating a general preference of these families for intestinal habitats (*e.g.* Lachnospiraceae, Ruminococcaceae, Porphyromonadaceae), or the termite-specific supercluster *Treponema* I (Spirochaetales), which comprises numerous genus-level lineages. For category 1 trees, we hypothesize that the last common ancestors of each termite-specific clade became specialized for termite gut environments, and eventually became widespread across a large number of termites through both colony-to-offspring vertical transmission in combination with horizontal colony-to-colony transfer. This category is consistent with a 'narrow' mixed mode of transmission [17].

Category 2 was defined in a similar way to category 1, with the exception that termite specific clades contained up to 10% non-termite-derived bacterial taxa nested within them. These clades contained primarily termite sequences, but were paraphyletic with respect to a small number of non-termite sequences (Figures 1C-D). This category comprised 10% of trees and 15.7% of the reads (Table S2). The nested taxa were, on many occasions, derived from the guts of other arthropods or animals, as exemplified by uncultured Lachnospiraceae (tree 87) and Ruminococceae (tree 116), which are common members of the mammalian gut microbiota. We hypothesise that taxa within clusters of termite-derived sequences are specialized for termite gut environments, with the exception of a relatively small proportion that have successfully colonised the guts of other organisms over evolutionary time. This category is consistent with a 'broad' mixed mode of transmission [17]. Here, we expect colony-to-offspring vertical transmission in combination with occasional horizontal transfers, not only between termites, but also between the guts of different animals, and potentially other environments. Although the topology of the trees suggest that category 2 taxa evolved within termites, and were subsequently transferred to other environments, it should be noted that bacteria derived from the guts of animals other than termites are likely to be underrepresented in our analyses. Further sampling of gut microbiomes, particularly those of other terrestrial arthropods, may reveal a higher level of horizontal transfer between different animal groups.

The remaining trees were assigned to Category 3, which comprised 28% of the trees and 18.7% of the reads in the dataset. Here, termite-derived sequences were interspersed with environmental sequences to a much greater degree than those in categories 1 and 2 (Figures 1E-F). The fact that many members of these groups are encountered in a variety of environments indicates that they are not transferred exclusively via a combination of vertical colony-to-offspring and colony-to-colony horizontal transmission. Nonetheless, several taxa in category 3 (as well as category 2) do belong to the core microbiota of termites, because they occur in high abundance in the majority of the termite lineages investigated (Table S2). For instance, uncultured members of Clostridiales (tree 138) or candidate division TM7 (tree 49) appear to be generally adapted to intestinal environments and may be easily exchanged even among unrelated host species.

We examined the level of congruence between host and bacterial relationships in category 1 and 2 trees. In no case did we find evidence for strict vertical inheritance of these lineages from colony to offspring. Instead, we found evidence for rampant horizontal transfer over evolutionary time between termite hosts for each of the bacterial lineages. This is manifested in the mixing of colours within each of the trees shown in Figure 1, where each subfamily of Termitidae (higher termites) is labeled with a different color, and all other families (lower termites) are labeled red. Nonetheless, we did identify a large number of cases in which host switching appears to be limited to taxa from one or more termite groups. In other words, some bacterial lineages appear to have become specialized for a particular termite subfamily, family, or multiple subfamilies or families, and have radiated significantly within this niche. For example, the *Treponema* tree (Figure 1B) shows a number of clades that are composed almost exclusively of phylotypes derived from either the subfamilies Apicotermitinae or Termitinae. The presence of family-specific clades within the termite-specific *Treponema* I supercluster is in agreement with a previous, comprehensive analysis of full-length 16S rRNA gene sequences obtained from the guts of 19 termite species [18].

To evaluate the robustness of our results, we reanalysed a subset of the 211 genus-level trees obtained with the 450-bp fragments using only full-length 16S rRNA sequences from termite guts and other environments

obtained from GenBank (Data S1). Although the number of sequences available for such analyses is much smaller, the results were consistent with those based on the short reads. For example, the corresponding trees of *Candidatus* Armantifilum (Figure 1A and Figure 2A) and *Endomicrobium* (Figure 1D and Figure 2B) show similar patterns and evidence the frequent switching of symbionts between distantly related termite taxa (Figures 1G and 2C show relationships among hosts).

Horizontal transfer of bacteria among termite species could occur either via aggressive encounters, during which the weaker contender is often eaten [19-21], or indirectly through soil or feeding substrates (*e.g.* via uptake of heterospecific faecal matter). That 62% of trees (category 1) contained large clades of termite-derived bacteria suggests that these taxa are incompatible with other environments that have been surveyed to date. Category 2 and 3 taxa appear to have higher levels of compatibility with alternative environments, and in the case of other animal guts are likely to have been transferred through contact with soil, or through feeding.

The gut microbiota of lower termites contains many bacterial lineages that are specifically associated with the surface, the cytoplasm, or the nucleus of their symbiotic flagellates (e.g., [22-24]). Phylogenetic analyses have documented co-speciation between flagellates and their bacterial symbionts and flagellates [8, 25, 26], but co-cladogenesis between bacterial symbionts and termites remains an exception [8] because of the occasional horizontal transfer of flagellates between termites of different families. This is illustrated by the case of *Endomicrobium*, which were acquired more than once from ancestral free-living lineages of gut bacteria [27], and whose flagellate hosts (together with their endosymbionts) have been transferred horizontally between lower termites of different families [28, 29].

Our results are in agreement with observations in numerous earlier, clone-library based studies of termite gut bacteria, which often showed clustering of termite-derived bacterial lineages from distantly related host taxa [4, 6]. Moreover, the relationship between termites and their gut bacteria is somewhat reminiscent of that between fungus-growing termites and the basidiomycete fungus cultivated in their fungal gardens. Symbiotic *Termitomyces* strains, which occur exclusively in symbiosis with termites of

the family Macrotermitinae, are not specialized on a particular termite species, and most lineages have retained the capacity to switch among multiple hosts [30].

The majority of the bacterial lineages identified in this study are subject to 'narrow' mixed-mode transmission [17]. They show a strong host specificity for termites, but co-cladogenesis – if present at all – is likely limited to closely related host lineages (see Figure S1). Prominent examples are termite-specific lineages in the *Fibrobacteres* and the candidate division TG3 (Trees 40-41, 56-57), which have been implicated in fibre digestion in wood-feeding higher termites [14, 31, 32]. We also found a number of bacterial lineages that had a more general affinity for animal guts, such as members of the *Clostridiales* family Ruminococcaceae, which made up 16.5% of the reads we analysed and are considered to contribute to cellulose and hemicellulose digestion in their intestinal habitats [33, 34]. Our results provide support for the theory of ecological fitting [35], which posits that traits developed by a symbiont during its evolutionary history may be co-opted for a new purpose in a different host. We predict that some groups of bacteria present in termites might be much more widespread among the guts of other organisms than currently appreciated.

AUTHOR CONTRIBUTIONS

TB and NL conceptualized the experiments and approach, with significant input from TAE. JS, YR and TB collected the samples. SS and TB performed the lab experiments. CD, TB, AB, and NL analyzed the data. TB and NL wrote the paper, with significant input from AB.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Sommer, F., and Backhed, F. (2013). The gut microbiota: masters of host development and physiology. Nat. Rev. Micro. *11*, 227-238.
2. Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., et al. (2008). Evolution of mammals and their gut microbes. Science *320*, 1647.
3. Moeller, A.H., Caro-Quintero, A., Mjungu, D., Georgiev, A.V., Lonsdorf, E.V., Muller, M.N., Pusey, A.E., Peeters, M., Hahn, B.H., and Ochman, H. (2016). Cospeciation of gut microbiota with hominids. Science *353*, 380.
4. Brune, A., and Dietrich, C. (2015). The gut microbiota of termites: digesting the diversity in the light of ecology and evolution. Ann. Rev. Microbiol. *69*, 145-166.
5. Hongoh, Y. (2011). Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. Cell. Mol. Life Sci. *68*, 1311-1325.
6. Ohkuma, M., and Brune, A. (2010). Diversity, structure, and evolution of the termite gut microbial community. In Biology of Termites: A Modern Synthesis, D.E. Bignell, Y. Roisin and N. Lo, eds. (Dordrecht: Springer), pp. 413-438.
7. Dietrich, C., Köhler, T., and Brune, A. (2014). The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. Appl. Environ. Microbiol. *80*, 2261-2269.
8. Noda, S., Kitade, O., Inoue, T., Kawai, M., Kanuka, M., Hiroshima, K., Hongoh, Y., Constantino, R., Uys, V., Zhong, J., et al. (2007). Cospeciation in the triplex symbiosis of termite gut protists (*Pseudotrichonympha* spp.), their hosts, and their bacterial endosymbionts. Mol. Ecol. *16*, 1257-1266.
9. Abdul Rahman, N., Parks, D.H., Willner, D.L., Engelbrektson, A.L., Goffredi, S.K., Warnecke, F., Scheffrahn, R.H., and Hugenholtz, P. (2015). A molecular survey of Australian and North American termite genera indicates that vertical inheritance is the primary force shaping termite gut microbiomes. Microbiome *3*, 1-16.
10. Mikaelyan, A., Meuser, K., and Brune, A. (2017). Microenvironmental heterogeneity of gut compartments drives bacterial community

structure in wood- and humus-feeding higher termites. FEMS Microbiol. Ecol. *93*, fiw210.

11. Brune, A. (2014). Symbiotic digestion of lignocellulose in termite guts. Nature Rev. Microbiol. *12*, 168-180.

12. Bignell, D.E., and Eggleton, P. (2000). Termites in ecosystems. In Termites: evolution, sociality, symbiosis, ecology, T. Abe, D.E. Bignell and M. Higashi, eds. (Dordrecht: Kluwer Academic Publishers), pp. 363-388.

13. Bourguignon, T., Lo, N., Cameron, S.L., Sobotnik, 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.

14. Mikaelyan, A., Dietrich, C., Köhler, T., Poulsen, M., Sillam-Dussès, D., and Brune, A. (2015). Diet is the primary determinant of bacterial community structure in the guts of higher termites. Mol. Ecol. *24*, 5284-5295.

15. Otani, S., Mikaelyan, A., Nobre, T., Hansen, L.R., Koné, N.A., Sørensen, S.J., Aanen, D.K., Boomsma, J.J., Brune. A., Poulsen, M. (2014). Identifying the core microbial community in the gut of fungus-growing termites. Mol. Ecol. *23*, 4631-4644.

16. Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum-evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. *26*, 1641-1650.

17. Ebert, D. (2013). The epidemiology and evolution of symbionts with mixed-mode transmission. Ann. Rev. Ecol. Evol. Syst. *44*, 623-643.

18. Mikaelyan, A., Köhler, T., Lampert, N., Rohland, J., Boga, H., Meuser, K., and Brune, A. (2015). Classifying the bacterial gut microbiota of termites and cockroaches: A curated phylogenetic reference database (DictDb). Syst. Appl. Micro. *38*, 472-482.

19. Shelton, T.G., and Grace, J.K. (1996). Review of agonistic behaviors in the Isoptera. Sociobiology *28*, 155-176.

20. Dropkin, V.H. (1946). The use of mixed colonies of termites in the study of host-symbiont relations. J. Parasitol. *32*, 247-251.

21. Thorne, B.L., and Haverty, M.I. (1991). A review of intracolony, intraspecific, and interspecific agonism in termites. Sociobiology *19*, 115-145.

22. Sato, T.K., H., Fujita, K., Noda, S., Kihara, K., Yamada, A., Ohkuma, M., and Hongoh, Y. (2014). Intranuclear verrucomicrobial symbionts and evidence of lateral gene transfer to the host protist in the termite gut. ISME J *8*, 1008-1019.

23. Yuki, M., Kuwahara, H., Shintani, M., Izawa, K., T., S., Starns, D., Hongoh, Y., and M, O. (2015). Dominant ectosymbiotic bacteria of cellulolytic protists in the termite gut also have the potential to digest lignocellulose. Environ. Microbiol. *17*, 4942-4953.

24. Ikeda-Ohtsubo, W., Strassert, J.F.H., Köhler, T., Mikaelyan, A., Gregor, I., McHardy, A.C., Tringe, S.G., Hugenholtz, P., Radek, R., and Brune, A. (2016). 'Candidatus Adiutrix intracellularis', an endosymbiont of termite gut flagellates, is the first representative of a deep-branching clade of Deltaproteobacteria and a putative homoacetogen. Environ. Microbiol. *18*, 2548-2564.

25.   Ikeda-Ohtsubo, W., and Brune, A. (2009). Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and *Candidatus* Endomicrobium trichonymphae. Mol. Ecol. *18*, 332-342.

26.   Desai, M.S., Strassert, J.F.H., Meuser, K., Hertel, H., Ikeda-Ohtsubo, W., Radek, R., and Brune, A. (2010). Strict cospeciation of devescovinid flagellates and Bacteroidales ectosymbionts in the gut of dry-wood termites (Kalotermitidae). Environ. Microbiol. *12*, 2120-2112.

27.   Mikaelyan, A., Thompson, C.L., Meuser, K., Zheng, H., Rani, P., Plarre, R., and Brune, A. (2017). High-resolution phylogenetic analysis of Endomicrobia reveals multiple acquisitions of endosymbiotic lineages by termite gut flagellates. Environ. Microbiol. Rep., 477-583.

28.   Kitade, O. (2004). Comparison of symbiotic flagellate faunae between termites and a wood-feeding cockroach of the genus *Cryptocercus*. Microbes Env. *19*, 215-220.

29.   Radek, R., Meuser, K., Strassert, J.F.H., Arslan, O., Teßmer, A., Sobotník, J., Sillam-Dussès, D., Nink, R.A., and Brune, A. (2017). Exclusive gut flagellates of Serritermitidae suggest a major transfaunation event in lower termites: description of *Heliconympha glossotermitis* gen. nov. spec. nov. . J. Eukaryot. Microbiol., in press, doi: 10.1111/jeu.12441

30.   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.

31.   Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., Cayouette, M., McHardy, A.C., Djordjevic, G., Aboushadi, N., et al. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature *450*, 560-565.

32.   Mikaelyan, A., Strassert, J.F.H., Tokuda, G., and Brune, A. (2014). The fiber-associated cellulolytic bacterial community in the hindgut of wood-feeding higher termites (*Nasutitermes* spp.). Environ. Microbiol. *16*, 2711-2722.

33.   He, S., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R.H., Kyrpides, N.C., Warnecke, F., Tringe, S.G., and Hugenholtz, P. (2013). Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. PLOS One *8*, e61126.

34.   Ben David, Y., Dassa, B., Borovok, I., Lamed, R., Koropatkin, N.M., Martens, E.C., White, B.A., Bernalier-Donadille, A., Duncan, S.H., Flint, H.J., et al. (2015). Ruminococcal cellulosome systems from rumen to human. Environ. Microbiol. *17*, 3407-3426.

35.   Brune, A. (2016). Co-evolution of marine worms and their chemoautotrophic bacterial symbionts: unexpected host switches explained by ecological fitting? Mol. Ecol. *25*, 2964-2966.

36.   Bourguignon, T., Lo, N., Šobotník, J., Ho, S.Y.W., Iqbal, N., Coissac, E., Lee, M., Jendryka, 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.

37.   Köhler, T., Dietrich, C., Scheffrahn, R.H., and Brune, A. (2012). High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). Appl. Environ. Microbiol. *78*, 4691-4701.

38.   Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics *27*, 2194-2200.

39.   Edgar, G.C. (2013). Highly accurate OTU sequences from microbial amplicon reads. Nat. Methods *10*, 996-998.

40.   Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335-336.

41.   Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059-3066.

42.   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.

43.   Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. *61*, 539-542.

44.   Rambaut, A., and Drummond, A.J. (2007). Tracer. Available at: http://www.beast.bio.ed.ac.uk/Tracer.

45.   Rambaut, A. (2016). FigTree v1.4.3. [http://tree.bio.ed.ac.uk/software/figtree/].

46.   Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R. a, Oakley, B.B., Parks, D.H., Robinson, C.J., *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. *75*, 7537-7541.

47.   Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. *73*, 5261-5267.
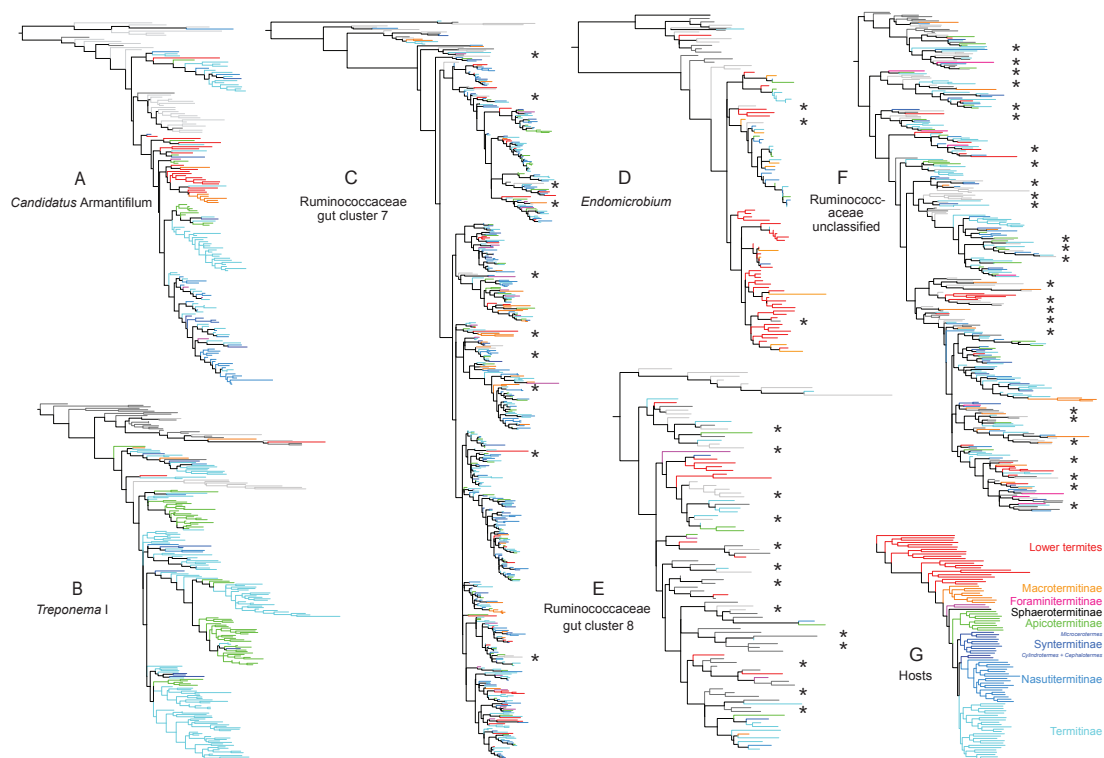
Figure 1. Selected phylogenetic trees showing relationships between termite derived sequences and related environmental sequences recovered using BLAST, based on 450-bp of 16S rRNA. Trees were inferred using FastTree.

(A, B) Category 1 trees of *Candidatus* Armantifilum (see Tree 17, Data S1 for additional detail) and *Treponema* I (see Tree 197 for additional detail) respectively.

(C, D) Category 2 trees of one Ruminococcaceae clade (see Tree 116 for additional detail) and *Endomicrobium* (see Tree 55 for additional detail) respectively.

(E, F) Category 3 trees of two Ruminococcaceae clades (See Tree 118 and Tree 138 respectively for additional detail).

(G) Relationships among the host taxa examined in this study, based on full mitochondrial genomes [13, 36].

Asterisks highlight environmental sequences nested within or among termite derived sequences in category 2 and 3 trees. Note that a uniform color of all branches within a clade does not indicate strictly vertical transfer of the respective taxa. Close inspection of host relationships revealed significant

amounts of transfer between hosts and a lack of co-cladogenesis (see Data S1). Taxon names and support values for each of these trees, as well as the other 205 trees generated in this study, are shown in Data S1.
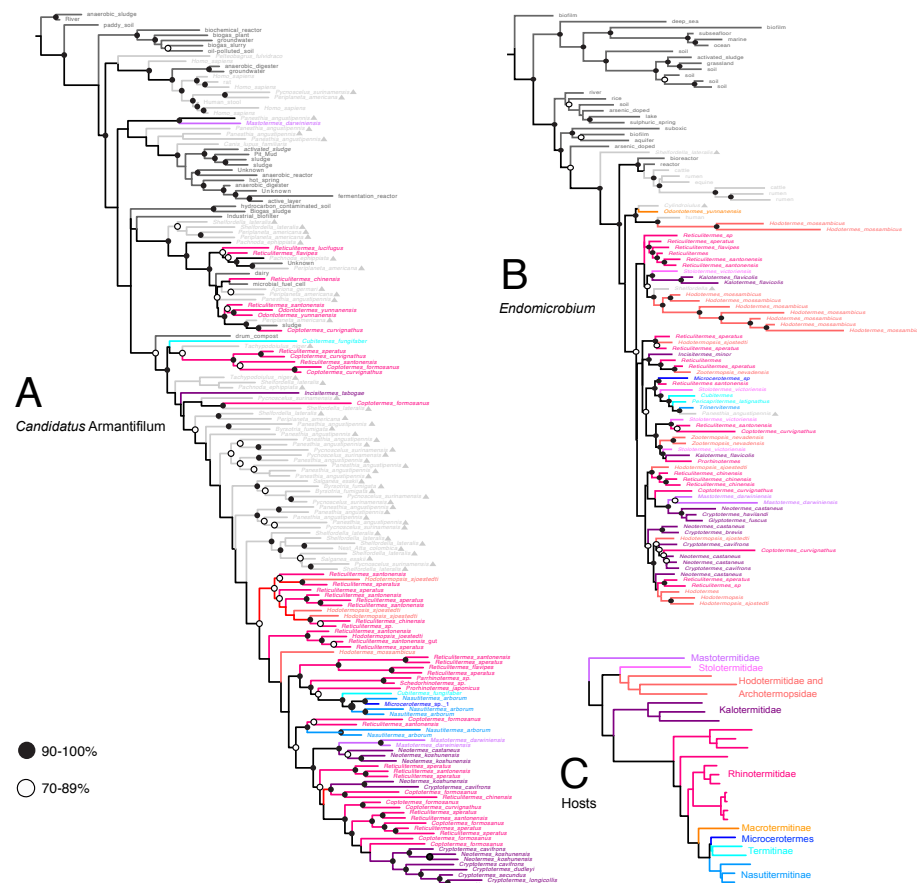
Figure 2. Phylogenetic relationships among termite-derived sequences and related environmental sequences, based on full-length 16S rRNA sequences and inferred using MrBayes. Percentage values show posterior probabilities. All sequences were obtained from GenBank.

(A) Relationships between representatives of *Candidatus* Armantifilum and related sequences. The corresponding tree based on short reads is shown in Figure 1A.

(B) Relationships between representatives of *Endomicrobium* and related sequences. The corresponding tree based on short reads is shown in Figure 1D.

(C) Relationships among the host taxa examined in these bacterial trees, based on full mitochondrial genomes [13, 36].

Taxa marked with light grey triangles are derived from other arthropods.

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests may be directed to and will be fulfilled by the Lead Contact, Nathan Lo (nathan.lo@sydney.edu.au).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We studied the gut bacterial communities of 94 samples, each belonging to distinct species of termites, and representative of global termite diversity (Table S1). All samples were collected in the field, preserved in RNA-later® and stored at -80°C until DNA extraction.

METHOD DETAILS

Whole genomic DNA was extracted from dissected digestive tracts of five to ten workers using the NucleoSpin® Soil kit of Macherey-Nagel according to manufacturer protocol. We used the primers 343Fmod (TACGGGWGGCWGCA) and 784Rmod (GGGTMTCTAATCCBKTT) to PCR amplify a fragment of 16S rRNA gene [37]. We conducted PCR amplifications using GoTaq® with the same conditions described in [37], that is initial denaturation (3 min at 95°C), 26 cycles of amplification (20 s at 95°C, 20 s at 48°C, and 30 s at 72°C), and a terminal extension (3 min at 72°C). Multiplexing and subsequent paired-end sequencing with Illumina MiSeq were carried out through a commercial service (BGI Tech. Solutions Co., China).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data filtering

We selected combined reads longer than 350 base pairs and removed others from subsequent analyses. We identified chimeras using UCHIME [38], implemented in USEARCH v7.0 [39] against the DictDb bacterial reference database [18], using a score threshold of 0.5 and a minimum divergence of 1.5. We independently sorted reads for each library into operational taxonomic units (OTUs) (6% sequence dissimilarity) using UPARSE [39], implemented in USEARCH v7.0 [39]. We excluded OTUs represented by less than five sequences from subsequent analyses. For each OTU, the most abundantly represented sequence was selected as reference. All OTU

reference sequences were then clustered into genus-level bacteria lineages, defined as groups of 12% sequence dissimilarity. 12% dissimilarity is generally higher than the commonly accepted threshold for genera (8%) of bacteria, but it was still too low to group the most abundant termite gut microbes, such as *Endomicrobium*, *Candidatus* Arthromitus and the *Treponema* I supercluster, into single groups. Downstream analyses were performed on all genus-level bacterial lineages that comprised more than 10 OTUs.

Identification of genus-level bacterial lineages
We identified the taxonomic affiliation of each genus-level bacterial lineage using the naïve Bayesian classifier from MOTHUR [46, 47] implemented in QIIME [40]. The DictDb database (version 3.0) was used as a reference for taxonomic assignment [18].In some cases, the classification was further refined using the current SILVA reference database (http: www.arb-silva.de).

Related species searches through BLAST
For each genus-level termite-derived bacterial lineage we searched for closely related, non-termite derived sequences available on GenBank. BLAST (blastn) searches were performed using each OTU from every genus-level bacterial lineage, with "Max target sequences" set on 10, and the option "Entrez Query" specifying "NOT termite". BLAST-obtained sequences for each genus-level lineage were then clustered in groups of 6% dissimilarity, and the most abundantly represented sequence was selected as a reference for each group. This method of sequence selection was similar to that used for selecting termite-derived sequences for analysis. BLAST reference sequences were classified into seven categories based on the information provided on GenBank: agricultural, industrial, mammal gut and feces, marine and aquatic, terrestrial arthropod gut, other animal-derived sequences, and Unclassified. All BLAST-derived reference sequences were then subject to phylogenetic analysis together with the termite-borne sequences of each genus-level bacterial lineage.

Phylogenetic analyses

Sequences of all genus-level bacterial lineages (including those from BLAST analyses) were aligned independently with MAFFT v7.300b using the option "adjustdirectionaccurately", and otherwise default settings [41, 42]. Phylogenetic trees were reconstructed with FastTree [16] implemented in QIIME under default settings [40]. All trees were visualized using FigTree v1.4.3 [45].

Categorization of phylogenetic trees
We defined three categories and assigned each phylogenetic tree to one of these three categories. Category 1 comprised trees in which ≥30% of termite-derived sequences formed a monophyletic group (in some cases, multiple clades, each containing ≥30% of the termite derived sequences, were recovered within a single tree). Category 2 was defined in a similar way to category 1, with the exception that termite specific clades contained up to 10% non-termite derived bacterial taxa nested within them. Category 3 contained all other trees. Typically, category 1 trees comprised termite-specific bacterial clades, category 2 trees comprised bacterial clades with termite affinities, but with evidence for transfer of bacteria between termites and other environments, and category 3 trees comprised bacterial clades with broad affinities, including termites.

Trees derived from full 16S rRNA sequences derived from GenBank
The phylogenetic trees generated in this study were based on sequences of about 450 pairs of bases. To test whether our findings held with longer sequences, we carried out phylogenetic reconstructions based on the full 16S rRNA gene using GenBank-derived sequences. We selected ten bacterial lineages for which full-length 16S rRNA termite-derived sequences were generated in previous studies (Data S1). For each of these bacterial lineages, we randomly selected one sequence that we used for a BLAST search to recover other full length 16S rRNA sequences from termite gut bacteria and related environmental sequences. BLAST searches were carried out with the options "Max target sequences" set on 500 and otherwise default settings. All sequences obtained that way were clustered in groups of 6% similarities, from which reference sequences were selected for analyses, as described above.

Phylogenetic analyses were performed in a Bayesian framework using MrBayes version 3.2.1 [43]. Posterior distributions were estimated using Markov chain Monte Carlo (MCMC) sampling with four chains (three hot and one cold). Samples were drawn every 1000 steps over a total of MCMC 2×10^6 steps. Each analysis was repeated twice. The final tree was obtained using a combination of the two replicated analyses, and the first 5×10^5 steps were discarded, based on inspection of the trace files using Tracer v1.5 [44]. Example trees (*Candidatus* Armantifilum and *Endomicrobium*) for these analyses are shown in Figure 2.

DATA AND SOFTWARE AVAILABILITY

The accession numbers for the 16S rRNA amplicon libraries generated in this study are freely available in GenBank: PRJNA422502.

Data S1. Phylogenetic trees of the 211 genus-level bacterial lineages detected across 94 termite species and their closest relatives outside of termite guts. Related to Figures 1-2.

List of bacterial lineages for which we analysed the full length 16S rRNA sequences retrieved from GenBank are indicated (phylogenetic analyses were carried out with MrBayes (see STAR methods).

Table S1. List of termite species included in this study and sample-associated information. Related to STAR Methods.

Table S2. List of genus-level bacterial lineages detected in the respective termite species. Only lineages represented by more than 10 OTUs were included. Related to Figure 1.