RESEARCH ARTICLE

Convergent Bacterial Microbiotas in the Fungal Agricultural Systems of Insects

Frank O. Aylward,[a,b]* Garret Suen,[a] Peter H. W. Biedermann,[c] Aaron S. Adams,[d] Jarrod J. Scott,[a,b,e] Stephanie A. Malfatti,[f] Tijana Glavina del Rio,[f] Susannah G. Tringe,[f] Michael Poulsen,[g] Kenneth F. Raffa,[d] Kier D. Klepzig,[h] Cameron R. Currie[a,b]

Department of Bacteriology, University of Wisconsin—Madison, Madison, Wisconsin, USA[a]; Department of Energy Great Lakes Bioenergy Research Center, University of Wisconsin—Madison, Madison, Wisconsin, USA[b]; Insect Symbiosis Research Group, Max Planck Institute for Chemical Ecology, Jena, Germany[c]; Department of Entomology, University of Wisconsin—Madison, Madison, Wisconsin, USA[d]; Smithsonian Tropical Research Institute, Balboa, Ancon, Panama[e]; Department of Energy Joint Genome Institute, Walnut Creek, California, USA[f]; Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark[g]; USDA Forest Service, Southern Research Station, Asheville, North Carolina, USA[h]

* Present address: Frank O. Aylward, Center for Microbial Oceanography Research and Education, University of Hawaii—Manoa, Honolulu, Hawaii, USA.

ABSTRACT The ability to cultivate food is an innovation that has produced some of the most successful ecological strategies on the planet. Although most well recognized in humans, where agriculture represents a defining feature of civilization, species of ants, beetles, and termites have also independently evolved symbioses with fungi that they cultivate for food. Despite occurring across divergent insect and fungal lineages, the fungivorous niches of these insects are remarkably similar, indicating convergent evolution toward this successful ecological strategy. Here, we characterize the microbiota of ants, beetles, and termites engaged in nutritional symbioses with fungi to define the bacterial groups associated with these prominent herbivores and forest pests. Using culture-independent techniques and the *in silico* reconstruction of 37 composite genomes of dominant community members, we demonstrate that different insect-fungal symbioses that collectively shape ecosystems worldwide have highly similar bacterial microbiotas comprised primarily of the genera *Enterobacter*, *Rahnella*, and *Pseudomonas*. Although these symbioses span three orders of insects and two phyla of fungi, we show that they are associated with bacteria sharing high whole-genome nucleotide identity. Due to the fine-scale correspondence of the bacterial microbiotas of insects engaged in fungal symbioses, our findings indicate that this represents an example of convergence of entire host-microbe complexes.

IMPORTANCE The cultivation of fungi for food is a behavior that has evolved independently in ants, beetles, and termites and has enabled many species of these insects to become ecologically important and widely distributed herbivores and forest pests. Although the primary fungal cultivars of these insects have been studied for decades, comparatively little is known of their bacterial microbiota. In this study, we show that diverse fungus-growing insects are associated with a common bacterial community composed of the same dominant members. Furthermore, by demonstrating that many of these bacteria have high whole-genome similarity across distantly related insect hosts that reside thousands of miles apart, we show that these bacteria are an important and underappreciated feature of diverse fungus-growing insects. Because of the similarities in the agricultural lifestyles of these insects, this is an example of convergence between both the life histories of the host insects and their symbiotic microbiota.

Received 8 October 2014 **Accepted** 29 October 2014 **Published** 18 November 2014

Citation Aylward FO, Suen G, Biedermann PHW, Adams AS, Scott JJ, Malfatti SA, Glavina del Rio T, Tringe SG, Poulsen M, Raffa KF, Klepzig KD, Currie CR. 2014. Convergent bacterial microbiotas in the fungal agricultural systems of insects. mBio 5(6):e02077-14. doi:10.1128/mBio.02077-14.

Editor Arturo Casadevall, Albert Einstein College of Medicine

Copyright © 2014 Aylward et al. This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license, which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Frank O. Aylward, faylward@hawaii.edu, or Cameron R. Currie, currie@bact.wisc.edu.

This article is a direct contribution from a Fellow of the American Academy of Microbiology.

Symbioses between metazoans and microbial communities are ubiquitous in nature and have contributed to many of the watershed events in the history of life on Earth (1–3). These symbiotic microbiota, which range from simple consortia of relatively few species to highly complex and dynamic communities, have been shown to benefit their hosts through defense against pathogens (4–7), degradation of recalcitrant dietary material (8–10), and biosynthesis of essential nutrients (11–13). The vast physiological potential of microbes plays an important role in the acquisition of novel ecological strategies in metazoans, and microbial symbionts have been argued to play an important role in host adaptation and speciation (14). Despite the importance of symbiotic microbiota for the evolution and physiology of their host, the forces that shape the structure and dynamics of these communities are not well understood.

Metazoans collectively associate with a vast phylogenetic diversity of microbes encompassing all three domains of life (15). Although these host-associated communities are also physiologically diverse, there often exists substantial functional redundancy between distantly related community members (15, 16). This diversity and functional redundancy of host-associated microbiota have implications for community structure. Specifically, patterns of convergence observed in the functional potential of host-associated communities of distantly related hosts have been pos-

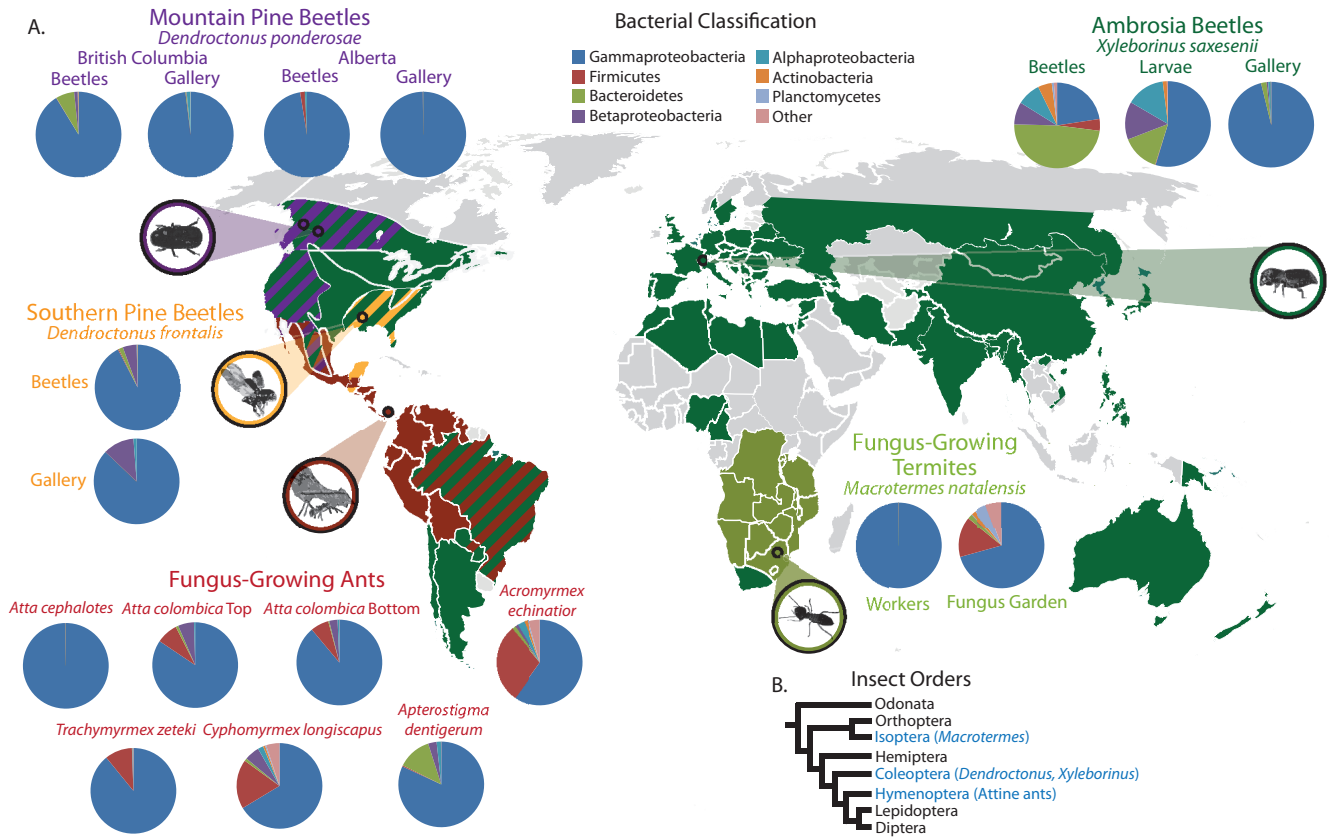

FIG 1 Distribution of insect-fungal symbioses and composition of their bacterial microbiota. (A) Map showing the global distribution of the insects analyzed here (colored regions on the map) and the locations from which samples were obtained in this study (circled). Pie charts show the phylogenetic composition of bacteria identified from 16S amplicon libraries sequenced from each sample, with colors corresponding to bacterial phylogenetic groups (in key). Metagenomes constructed from both the top and bottom strata of fungus gardens are shown for the leaf-cutter ant *Atta colombica*. Global insect distributions are based on previous estimates (see Materials and Methods). (B) Simplified phylogeny of select insect orders (based on that previously reported [79]). Orders that include insects with insect-fungal symbioses presented in this study are highlighted in blue.

tulated to be the product of selection for microbial groups that possess particular traits but are not always closely related (17, 18). For example, studies of various metazoan hosts have indicated that host niche is a primary determinant of the overall physiological capacity of microbiota that can lead to functional convergence irrespective of the evolutionary history of the host (19–21). Moreover, other recent studies have generally found that niche-specific factors are important in structuring host-associated microbiota (22–24). However, host phylogeny has also been linked to microbial community composition in studies of primates and insects (24–31), indicating that a number of factors are important for structuring the phylogenetic composition and functional capacity of host-associated microbiotas.

In this study, we characterized bacterial microbiotas associated with fungus-growing insects, which comprise distantly related ant, beetle, and termite lineages that have independently established symbioses with fungi (32). The microbial symbioses of these hosts are central to their life histories and have enabled them to become dominant herbivores and prevalent tree pests in widespread tropical and temperate ecosystems (32–34). These insects are able to shape ecosystems around the globe largely because their symbiotic microbiota act as "ancillary guts" that degrade recalcitrant plant biomass and convert it into nutrients more accessible to their hosts (32, 35–39), thereby allowing these insects to exploit ecological niches that would otherwise be unavailable. In some fungus-growing insects, the external digestive systems have led to the evolution of species that build elaborate colonies composed of millions of insects divided into castes with distinct tasks, behaviors, and morphologies, as demonstrated by leaf-cutter ants of the genus *Atta* and fungus-growing termites of the genus *Macrotermes* (32, 33).

Here, we analyzed fungus-growing ants (Tribe: Attini), ambrosia beetles (Tribe: Xyleborini), and termites (subfamily: Macrotermitinae), which all engage in obligate fungal agriculture (32), as well as mountain and southern pine beetles (genus *Dendroctonus*), which, although lacking many of the true agricultural characteristics of the other insects, also associate with mutualistic fungi that they consume for food (34, 40). Our samples included three insect orders and spanned a considerable portion of the global distribution of these insect-fungal symbioses (Fig. 1A and B; see Table S1 in the supplemental material). Due to the independent origins of these nutritional fungal symbioses across distantly related host lineages collected from across the globe, our analyses of their microbiotas provides a unique opportunity to assess the extent to which the similarity of their ecological niches has influenced the structure of their associated microbial communities. To this end,

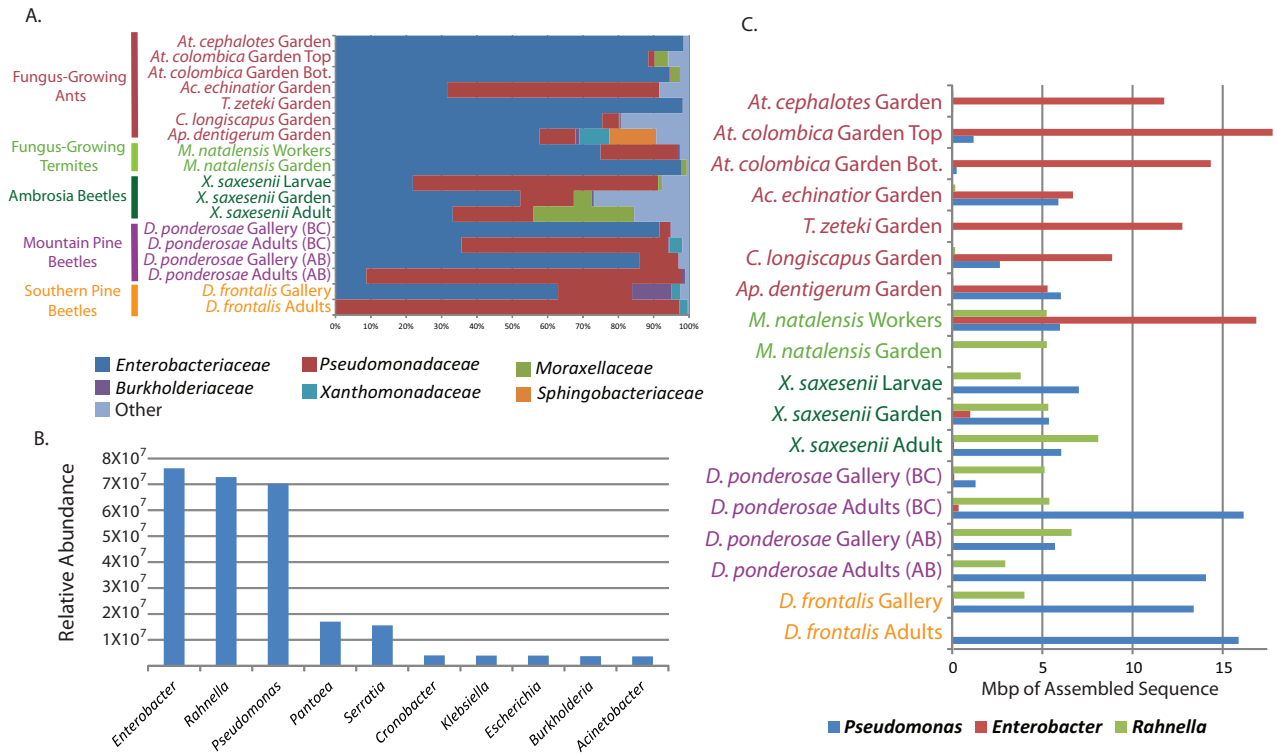

FIG 2 Phylogenetic binning comparisons of the 18 metagenomes analyzed in this study. (A) Family-level binning and coverage-weighted relative abundance comparison of the contigs in the metagenomes. The genera *Rahnella* and *Enterobacter* belong to the family *Enterobacteriaceae*, while the genus *Pseudomonas* belongs to the family *Pseudomonadaceae*. (B) Rank-abundance overview of the most abundant bacterial genera identified in the combined 18 metagenomes using coverage-weighted contig binning. Relative abundance estimates were obtained by multiplying the length of each contig by its coverage and summing the results for a given family- or genus-level bin. (C) Mbp of sequences binned to the genera *Rahnella*, *Enterobacter*, and *Pseudomonas* in the 18 metagenomes. Abbreviations: AB, Alberta; BC, British Columbia; Bot, Bottom.

we sought to provide a fine-scale comparison of the composition of bacterial communities associated with these insect fungal symbioses through sequencing of 18 16S amplicon libraries comprising a total of 136,400 quality-filtered sequences (minimum length of 200 bp) and 18 community metagenomes comprising a total of 6.8 Gbp of raw sequence data from which we reconstructed 37 composite genomes of dominant community members (see Tables S2 and S3).

RESULTS AND DISCUSSION

The 16S libraries of all insect samples were dominated by sequences classified as belonging to the class *Gammaproteobacteria* (ranging from 54 to 99% of sequences per sample), except in the adult ambrosia beetles, where the phylum *Bacteroidetes* dominated (46% of sequences) and *Gammaproteobacteria* were the next most abundant group (23% of sequences) (Fig. 1A). The genera *Pseudomonas*, *Enterobacter*, and *Rahnella* were the most abundant in our genus-level analysis, and rank-abundance curves demonstrated that 2 or 3 groups dominated the composition of our 16S libraries (see Fig. S1 and S2 in the supplemental material). All libraries contained between 31 and 753 operational taxonomic units (OTUs; 95% identity cutoff), and the rarefaction analyses demonstrate sufficient sampling (Fig. S3). Relative abundance quantification of the contigs in the community metagenomes corroborated those of our 16S-based analysis in showing that the gammaproteobacterial families *Enterobacteriaceae* and *Pseudomonadaceae* dominated all of our samples (Fig. 2A). Comparison of the Clusters of Orthologous Groups (COG [41]) and Protein Families (Pfam [42]) profiles of these metagenomes to publicly available metagenomes revealed clustering into three groups: one containing all fungus-associated insect samples, one containing other host-associated communities, and one containing environmental (non-host-associated) communities (Fig. 3; see Table S4). In addition to the correspondence identified for the phylogenetic composition of the insect communities sampled here, this functional clustering demonstrates a general equivalence of the physiological potential of these microbiota that is distinct from other microbial communities for which metagenomes are available.

The genera *Pseudomonas*, *Enterobacter*, and *Rahnella* were particularly well represented in the community metagenomes, with the genus *Pseudomonas* abundant in most of the ant, beetle, and termite systems, the genus *Enterobacter* more common in the fungus-growing ant samples, and the genus *Rahnella* more abundant in the termite- and beetle-associated samples (Fig. 2B). Genus-level rank-abundance analysis also found these three genera to be overwhelmingly the most abundant in these communities (Fig. 2B and C). Mapping of all genes recovered from the *Enterobacteriaceae* and *Pseudomonas* bins onto a multilocus phylogeny of reference genomes (Fig. 4A) revealed highly similar phylogenetic profiles among the ambrosia beetle, termite, and pine beetle samples, with sequences most similar to *Rahnella aquatilis* and *Pseudomonas fluorescens* strains dominating the phylogenetic

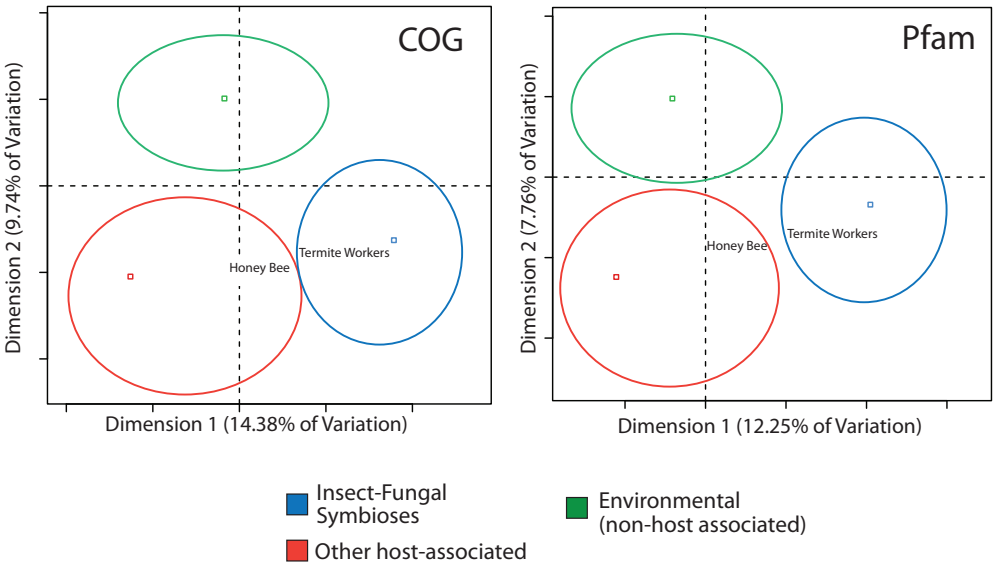

FIG 3 Principle component analyses (PCA) comparing the functional profiles of the metagenomes of insect-fungal symbioses to those of 57 publicly available metagenomes generated from environmental or gut-associated samples. Annotations were performed using both the Clusters of Orthologous Groups (COG) and Protein Families (Pfam) databases. A full list of metagenomes used can be found in Table S4. A metagenome constructed from the gut of the honey bee was the only metagenome found to cluster near the insect-fungal symbiosis samples (labeled in both panels and in both cases closest to the fungus-growing termite adult sample). Squares indicate the category averages.

profiles of these metagenomes. The fungus-growing ant samples were also dominated by sequences mapping to *Enterobacteriaceae* and *Pseudomonas* genomes, but the genes in these metagenomes mapped primarily to the *Enterobacter* and *Pseudomonas putida* clades in the phylogeny (Fig. 4A). This pattern held for all fungus-growing ants, including *Apterostigma dentigerum*, which cultivates a pterulaceous fungus distantly related to the Lepiotaceae family of fungi grown by other attine ants (43).

We reconstructed composite genomes for the dominant microbial community members to compare genomes directly across insect-fungal symbioses (see Materials and Methods). These included 12 *Enterobacter* genomes, 15 *Pseudomonas* genomes, and 10 *Rahnella* genomes, all estimated to be >40% complete (see Table S5 in the supplemental material). Calculation of the pairwise average nucleotide identities (ANI) of the reconstructed genomes revealed that the microbiota from different insect-fungal symbioses contained dominant bacterial constituents that were highly similar to each other at the whole-genome level (often >95% ANI) (Fig. 4B, and see Fig. S4), and multilocus phylogenetic analysis of conserved housekeeping genes confirmed that the majority of the reconstructed *Enterobacter*, *Rahnella*, and *Pseudomonas* genomes grouped together in well-supported clades (Fig. 5). The *Rahnella* and *Pseudomonas fluorescens* genomes reconstructed from the ambrosia beetle, fungus-growing termite, and pine beetle metagenomes were particularly similar across insect systems, with ANI values in some cases exceeding 98% (Fig. 4B). Composite genomes from the genus *Enterobacter* could be reconstructed from ant and termite samples and segregated into two distinct groups (ANI of 95.9 to 98.5% and 85.5 to 99.1% within the groups) (Fig. 4B; Fig. S4), while the reconstructed genomes from the *Pseudomonas putida* group were specific to fungus-growing ants (Fig. 4B and see Fig. S4).

Our results demonstrate that the microbiotas of diverse insect-fungal symbioses that collectively shape terrestrial ecosystems worldwide contain prevalent and highly similar bacterial constituents. Our comparison of the coding potential of these microbiotas demonstrates broad functional congruence. Moreover, our composite-genome reconstructions reveal that the *Enterobacter*, *Rahnella*, and *Pseudomonas* groups that dominate these communities exhibit high degrees of whole-genome similarity, with the ANI of reconstructed genomes often exceeding 95% (Fig. 4B). In other metazoans, similarities in host niche have been shown to drive convergence in the functional potential of associated microbiota but to influence the phylogenetic composition at broad taxonomic levels (e.g., phylum level) only weakly (19, 21). This pattern, where different hosts possess physiologically congruent microbiotas with phylogenetically distinct composition, is likely due to selection for particular physiological traits that are not necessarily linked to specific phylogenetic groups of microbes. Although the broad congruence of functional profiles across the communities described here is consistent with other systems in which distantly related hosts occupy similar ecological niches (19, 21), our finding of fine-scale phylogenetic convergence across hosts is unexpected, given that the insects analyzed collectively reside on three continents in both tropical and temperate ecosystems in which they are exposed to a vast diversity of microbes. This coupling of both functional and phylogenetic convergence in the microbiotas of insect-fungal symbioses, together with the dominance of 2 or 3 bacterial groups in each insect, suggests that highly specific mechanisms for maintaining host-bacterial or fungal-bacterial interactions are responsible for shaping diversity in these systems.

Our findings indicate that bacteria are a common and perhaps even defining feature of widespread insect-fungal symbioses. This is supported by the high degrees of similarity in the dominant bacterial constituents of the ants, termites, and beetles, despite contrasts in the details of the insect-fungal symbioses. For example, these insects cultivate phylogenetically disparate fungi that

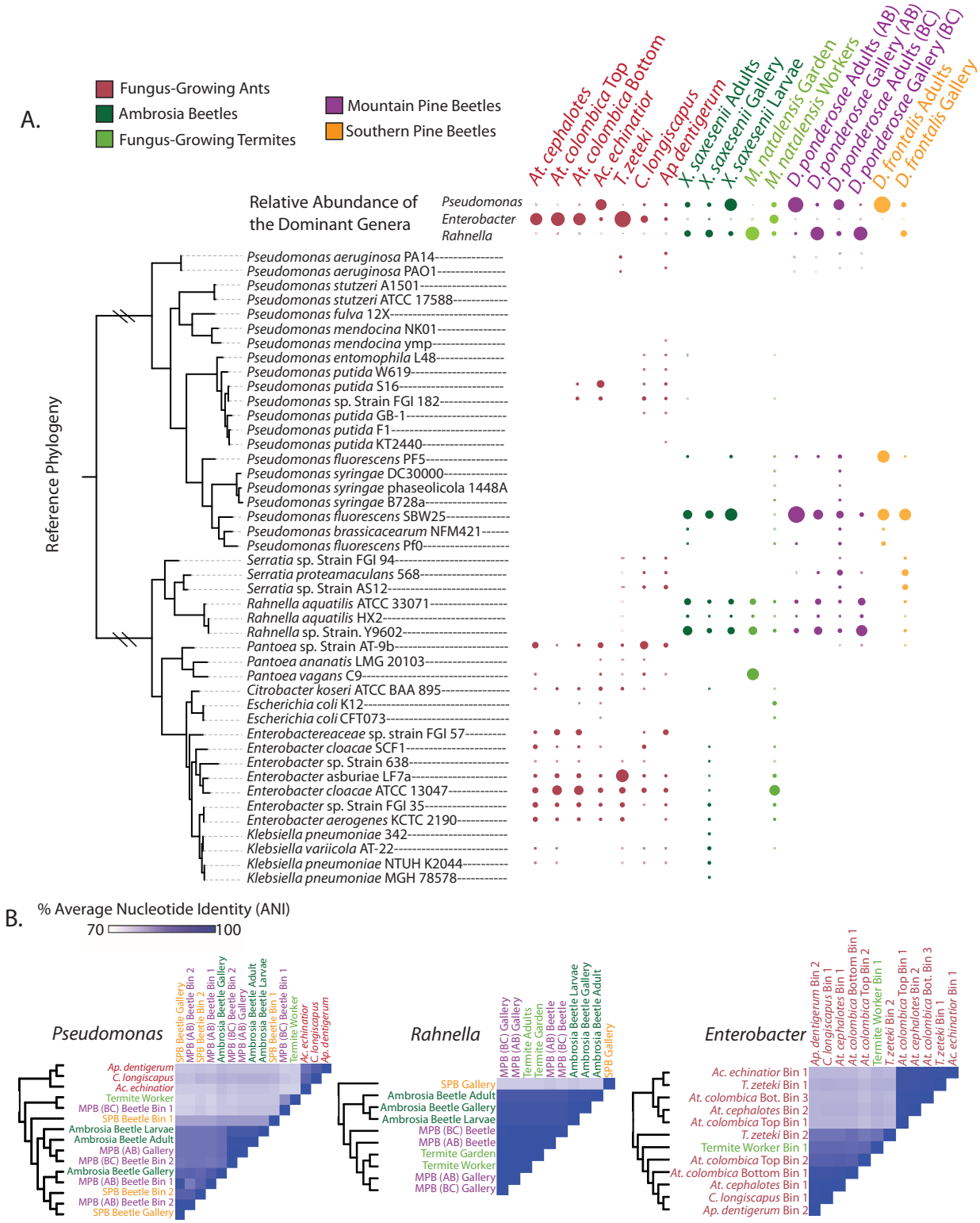

FIG 4 Comparisons of dominant groups represented in the metagenomes of insect-fungal symbioses. (A) Bubble chart showing the relative abundance of the most prevalent phylogenetic groups identified in the metagenomes. Relative abundances of the genera *Pseudomonas*, *Enterobacter*, and *Rahnella* were calculated using abundance-weighted coverage estimates of binned contigs. For the phylogenetic mapping analysis, all genes predicted from contigs classified to the family *Enterobacteriaceae* and genus *Pseudomonas* were mapped onto a maximum-likelihood phylogeny of representative sequenced genomes constructed using concatenated amino acid sequences from 9 highly conserved proteins (see Materials and Methods). Bootstrap support values have been omitted for clarity (a full phylogeny with support values can be found in Fig. S5). (B) Heatmaps showing the ANI values obtained from pairwise BLASTN comparisons of the composite *Enterobacter*, *Pseudomonas*, and *Rahnella* genomes reconstructed in this study. Dendrograms were constructed using a neighbor-joining algorithm with distance matrices constructed from pairwise ANI comparisons (see Materials and Methods).

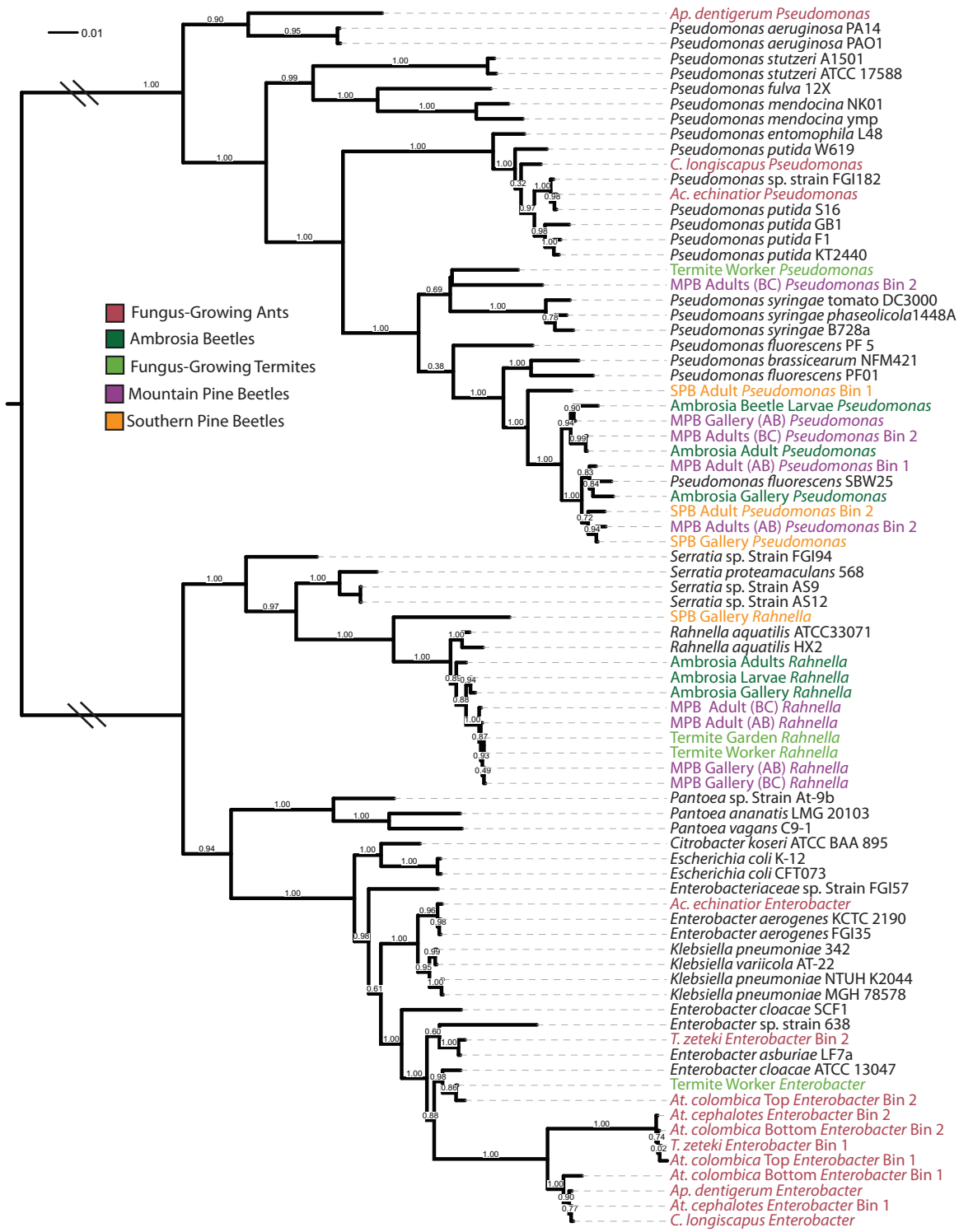

FIG 5 Maximum-likelihood multilocus phylogeny of reference *Enterobacter*, *Pseudomonas*, and *Rahnella* genomes together with the composite genomes reconstructed in this study (color coded according to host insects, as shown in the key). The phylogeny is based on concatenated amino acid sequences of 9 conserved proteins, and local support values were computed using the Shimodaira-Hasegawa test (see Materials and Methods for details).

collectively span two phyla (32), and it is even unclear to what extent mountain and southern pine beetles depend on their symbiotic fungi for nutrition (34). Due to the influence host-associated microbes have been shown to exert on metazoan behavior (44, 45), these findings raise the possibility that similar life history traits in distantly related insect lineages may in some ways be a consequence rather than the cause of the similar bacterial communities they harbor. Together with recent evidence that patterns of host speciation are recapitulated and potentially driven by symbiotic microbiota (14), our finding of convergent community assembly in the bacterial communities of insect-fungal symbioses underscores the fundamental importance of host-microbe interactions in shaping metazoan evolution.

MATERIALS AND METHODS

Insect distributions. The distributions of insect-fungal symbioses as depicted in Fig. 1 were compiled from previously reported estimates for attine ants (46, 47), *Macrotermes natalensis* (48), and *Xyleborinus saxesenii* (49). The distribution of *Dendroctonus frontalis* is based on estimates from the United States Department of Agriculture (USDA; http://www.fs.usda.gov/Internet/FSE_DOCUMENTS/fsbdev2_042840.pdf). The Invasive Species Compendium (http://www.cabi.org/isc/) was also used for *Xyleborinus saxesenii* and *Dendroctonus ponderosae*.

Sample collection, processing, and sequencing. Insect samples were collected between April and November of 2009. Details regarding the times and locations in which insect samples were acquired can be found in Table S1 in the supplemental material. Sampling of the leaf-cutter ant species *Atta cephalotes* and *Atta colombica* and the mountain pine beetle *Dendroctonus ponderosae* has been described previously (50, 51). All samples were placed on ice immediately after sampling and transported to the laboratory for processing. For all fungus garden, gallery, or whole-insect samples, the bacterial fraction was isolated by gently vortexing the samples in 1% phosphate-buffered saline (PBS) and 0.1% Tween before conducting a modified differential centrifugation procedure, as previously described (50). For the ambrosia beetles, fungus-growing termites, mountain pine beetles, and southern pine beetles, whole-insect or larval samples were also prepared for sequencing. For these samples, 90 to 300 whole insects were pooled for each sample. Once the bacterial fraction of each sample was isolated, DNA was extracted using the bacterial extraction protocol available in the Qiagen DNeasy plant maxikit (Qiagen Sciences, Germantown, MD), which has been shown to yield an accurate representation of community DNA (51). Community metagenomes were subsequently generated by using Roche 454 Titanium pyrosequencing (52), and assemblies were generated using Newbler version 2.1 with the default parameters. Details of the metagenomes and assembly statistics can be found in Table S2.

16S amplicon sequencing, processing, and analysis. Amplicon libraries spanning the V6-to-V8 (V6-8) region of the 16S ribosomal gene were constructed from DNA extracted from the 18 insect, fungus garden, and gallery samples using methods described previously (53). The libraries were processed using mothur (54) with procedures based on those outlined by Schloss et al. (55) and described on the mothur website (http://www.mothur.org/wiki/454_SOP) as of 1 July 2013. Briefly, flow data associated with the .sff files were extracted using the sffinfo command, and PyroNoise (56) as implemented by the shhh.flows command in mothur was used to trim sequences. Bar code and primer sequences were then removed, and all sequences of <200 bp were discarded. Sequences were then aligned to the SILVA 16S reference dataset (57) using the align.seqs command, and any not overlapping with the V6-8 region were removed (screen.seqs and filter.seqs commands). The pre.cluster command and UCHIME (58), as implemented by the chimera.uchime command, were used to remove chimeras. Sequences were classified through comparison with the NCBI 16S Ribosomal RNA Sequence Library (downloaded 2 January 2013) using BLASTN with the parameter -e^{-50}. Bacterial family- and phylum-level assessments were then made by comparing the top BLASTN hit with the UniProt taxonomy hierarchy (http://www.uniprot.org/taxonomy/) (downloaded 1 May 2013). Sequences not having a BLASTN hit were removed from subsequent analyses on the grounds that they comprised primarily misamplified fungal or insect 18S or chloroplast 16S sequences. Genus-level classifications were performed using only BLASTN hits with ≥97% nucleotide identity.

To generate operational taxonomic units (OTUs), quality-trimmed pyrotags were aligned using MAFFT version 5.662 E-INS-I (59), and distances were calculated using the dist.seqs command in mothur. OTUs were then generated using the cluster command in mothur. We focused our analyses on OTUs generated with a 95% identity cutoff, as the more traditional 97% cutoff has been shown to be more appropriate for full-length 16S sequences (60). To calculate the Shannon diversity index of the bacterial component of each 16S library, we first rarefied the data sets down to the number contained in the smallest library (248 bacterial sequences in the ambrosia beetle adult sample). The Shannon diversity index was then calculated for the 95% OTU cutoff for each of the 16S libraries using the summary.single command in mothur. Rarefaction curves for the confirmed bacterial sequences in each of the 16S libraries were also constructed using mothur (the rarefaction.single command). OTUs represented by a single sequence were not included in the rarefaction analysis on the grounds that these OTUs are more likely to be unidentified chimeras or misamplified sequences that may artificially inflate diversity estimates (61). Sequencing statistics for the 16S libraries can be found in Table S3 in the supplemental material, and rarefaction curves for the bacterial sequences are shown in Fig. S3.

Phylogenetic binning and relative abundance estimation. Phylogenetic bins from all metagenomes were generated using a combination of BLASTN (62) and PhymmBL (63). All contigs and singletons were first compared to a reference data set containing all completely sequenced bacterial and archaeal genomes available in the NCBI as of 1 January 2012. All contigs having BLASTN hits with E values of $<1e^{-10}$ were classified according to their best hit, while all other contigs were subsequently classified using PhymmBL. Contigs with no BLASTN matches that were classified by PhymmBL with confidence scores of <50 were considered "unclassified." A relative abundance value for each contig was calculated by multiplying the length of the contig by its percent coverage. The relative abundances of each phylogenetic group in the metagenomes were then calculated by summing the abundance value for each contig classified in a particular bacterial family or genus. To estimate the proportion of a particular phylogenetic group compared to the rest of the metagenome (as depicted for the genera *Enterobacter*, *Pseudomonas*, and *Rahnella* in Fig. 4A), the summed relative abundance of all contigs classified to that group was divided by the summed relative abundance of all contigs in that metagenome.

Phylogenetic mapping of community metagenome data. All contigs in the community metagenomes binned to the family *Enterobacteriaceae* or genus *Pseudomonas* were separated, and genes were predicted from these contigs by using Prodigal (64) with the metagenomic gene caller option. BLASTN was used to map genes and contigs onto the reference *Enterobacteriaceae* and *Pseudomonas* genomes listed in Fig. 4A (using nondefault parameters of -X 150, -q -1, -F F, and -e $1e^{-5}$), as these genomes represented bacterial families or genera found to be abundant in our phylogenetic binning analysis. Only best hits were retained, and the number of hits was catalogued for each genome. A phylogeny of the reference genomes was created (see below) and the number of hits to each reference were mapped using the interactive Tree of Life (iTOL [65]).

Composite genome reconstruction and completeness estimates. Composite genomes were reconstructed from the assembled community metagenomes through manual analysis of BLASTN-based homology searches against reference genomes, PhymmBL binning of contigs, and coverage estimates of contigs. First, all contigs >800 bp were binned at the genus level based on their binning assignment, and plots of contig length versus coverage were then generated for each genus in each metagenome. Contigs having similar coverage were placed in the same bin, and the best

BLASTN hit of each contig was again cross-referenced to ensure the contigs in each bin were mapping to the same genome (or genomes of closely related bacteria: for example, different strains of *Pseudomonas putida*). Estimates of the completeness of the composite genomes were generated by extrapolating from the number of single-copy housekeeping genes present in the assembled contig bins. We used a list of 182 core proteins previously used for this purpose (53), although 20 of these proteins were excluded on the grounds that they were not present in complete genomes of bacteria closely related to those identified in the metagenomes, namely, other sequenced *Pseudomonas*, *Enterobacter*, and *Rahnella* genomes. COG models for the remaining 162 core proteins were used for their identification, and the presence of these proteins in each composite genome was ascertained using reverse position specific (RPS)-BLAST (66) (E value, $<1e^{-5}$) of the predicted proteins in each bin. Only bins with >2 Mb of total sequence that were predicted to be >40% complete (see below) were considered for subsequent analyses (a list of the final composite genomes used can be found in Table S5 in the supplemental material).

Phylogenetic analysis of complete and composite genomes. Phylogenetic trees of select *Enterobacteriaceae* and *Pseudomonas* genomes were constructed from a concatenated amino acid alignment of translations of the highly conserved single-copy housekeeping genes *recA*, *fusA*, *recG*, *rpoB*, *rplB*, *lepA*, *ileS*, *pyrG*, and *leuS*, which have previously been shown to be useful phylogenetic markers (67). Hidden Markov models (HMMs) were created from curated alignments of these proteins available from the Ribosomal Database Project (68), using the hmmbuild command in HMMER3 (69). Proteins encoded in the genomes of interest were predicted using Prodigal (64) and compared to the HMMs using the hmmersearch command in HMMER3, and only best matches were retained. Partial-length proteins were manually annotated before being included in the final protein set. Proteins from the same genome were concatenated using custom PERL scripts, and missing proteins were replaced with the character "X," as this would allow proteins to be included in the final concatenated alignment even if they were not present in all genomes analyzed. Alignments were created using MAFFT version 5.662 E-INS-I (59) and trimmed using the program Trimal version 1.2 (70) with the parameter -automated1. Maximum-likelihood phylogenetic trees were constructed using FastTree (71), and support values for the nodes were calculated using the Shimodaira-Hasegawa (SH) test (72). Trees were visualized using iTOL.

Composite genome comparisons and BLASTN mapping. Composite genomes were compared to each other and to complete genomes in the same genera using both average nucleotide identity (ANI) and average amino acid identity (AAI) analyses. While ANI is typically used for comparing closely related bacteria, AAI is useful for more divergent comparisons. Thus, we used ANI to ascertain the degrees of similarity between our composite genomes belonging to the same genus, which were typically very similar, while AAI was used for comparisons of our composite genomes and representative genomes spanning a broader phylogenetic diversity (i.e., across the genus *Pseudomonas*). Overall, the results of our AAI, ANI, and multilocus phylogenetic analyses provided equivalent results in terms of the overall similarity of the composite genomes analyzed (Fig. 4B, and see Fig. S4 and S5 in the supplemental material). ANI and AAI values were generated for pairs of complete or composite genomes by averaging the percent identities of reciprocal best BLASTN or BLASTP hits of each gene/protein using the parameters -X 150, -q -1, -F F, and -e $1e^{-5}$. These BLASTN parameters have previously been shown to provide accurate ANI and AAI values (73). Dendrograms representing the ANI- or AAI-based similarity between different genomes were generated by converting ANI or AAI percent identity values into distances and using the Neighbor algorithm in PHYLIP (74) to calculate corresponding newick trees with the Mobyle web interface (75).

Comparison of the functional profiles of metagenomes. The 18 metagenomes generated here were compared to 24 gut-associated metagenomes and 33 non-host-associated metagenomes that are publicly available in the Integrated Microbial Genomes/Metagenomes database (IMG/M) (76) or the Metagenome Rapid Annotation using Subsystem Technology (MG-RAST) database (77) (details are in Table S4 in the supplemental material). Predicted proteins obtained from either the IMG/M or MG-RAST databases were compared to the Clusters of Orthologous Groups (COG) (41) and Protein Families (Pfam) (42) databases using RPS-BLAST (66) (E value, $<1e^{-5}$). Best RPS-BLAST hits were compiled in a matrix and normalized by the total number of COG or Pfam hits. COG or Pfam families representing <0.1% of total annotated proteins were excluded from subsequent analyses. The normalized matrices were then used to for principle component analyses (PCA) using the module FactoMineR (http://factominer.free.fr/) in the R statistical programming environment (http://www.R-project.org/) (78).

Accession numbers. All metagenomic data generated in this study can be found in the Integrated Microbial Genomes/Metagenomes Database (IMG/M [76]) under accession numbers 2029527003 to 2029527007, 2030936005, 2032320008, 2032320009, 2035918000, 2035918003, 2043231000, 2044078006, 2044078007, 2065487013, 2065487014, 2084038008, 2084038018, and 2228664020; further details on the metagenomic data are presented in Table S2 in the supplemental material. All 16S pyrotag data sets have been deposited in the NCBI Sequence Read Archive (SRA) under accession numbers SRP006785 and SRA047411.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.02077-14/-/DCSupplemental.

Figure S1, EPS file, 1.3 MB.
Figure S2, EPS file, 0.7 MB.
Figure S3, EPS file, 1.3 MB.
Figure S4, PDF file, 0.6 MB.
Figure S5, EPS file, 1.2 MB.
Table S1, DOCX file, 0.02 MB.
Table S2, DOCX file, 0.01 MB.
Table S3, DOCX file, 0.01 MB.
Table S4, DOCX file, 0.02 MB.
Table S5, DOCX file, 0.02 MB.

ACKNOWLEDGMENTS

We thank the staff of the DOE Joint Genome Institute and the Smithsonian Tropical Research Institute for their expertise and support in the collection and processing of all samples. We thank Gabriel J. Starrett for assistance with figures.

This work was supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231, the USDA NRI (2008-02438), National Science Foundation grants DEB-0747002, MCB-0702025, and MCB-0731822 to C.R.C., and the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494). P.H.W.B. was funded by an SNSF postdoctoral research grant (PBBEP3_141472), and M.P. was supported by a STENO grant from The Danish Council for Independent Research—Natural Sciences.

REFERENCES

1. **Moran NA.** 2006. Symbiosis. Curr. Biol. **16:**R866–R871. http://dx.doi.org/10.1016/j.cub.2006.09.019.
2. **Klepzig KD, Adams AS, Handelsman J, Raffa KF.** 2009. Symbioses: a key driver of insect physiological processes, ecological interactions, evolutionary diversification, and impacts on humans. Environ. Entomol. **38:**67–77. http://dx.doi.org/10.1603/022.038.0109.
3. **Douglas AE.** 2010. The symbiotic habit. Princeton University Press, Princeton, NJ.
4. **Ichinohe T, Pang IK, Kumamoto Y, Peaper DR, Ho JH, Murray TS, Iwasaki A.** 2011. Microbiota regulates immune defense against respiratory tract influenza A virus infection. Proc. Natl. Acad. Sci. U. S. A. **108:** 5354–5359. http://dx.doi.org/10.1073/pnas.1019378108.
5. **Harris RN, Brucker RM, Walke JB, Becker MH, Schwantes CR, Flaherty DC, Lam BA, Woodhams DC, Briggs CJ, Vredenburg VT, Minbiole KP.** 2009. Skin microbes on frogs prevent morbidity and mortality caused by a lethal skin fungus. ISME J. **3:**818–824. http://dx.doi.org/10.1038/ismej.2009.27.
6. **Becker MH, Harris RN.** 2010. Cutaneous bacteria of the redback sala-

mander prevent morbidity associated with a lethal disease. PLoS One **5:**e10957. http://dx.doi.org/10.1371/journal.pone.0010957.
7. **Dillon RJ, Vennard CT, Buckling A, Charnley AK.** 2005. Diversity of locust gut bacteria protects against pathogen invasion. Ecol. Lett. **8:**1291–1298. http://dx.doi.org/10.1111/j.1461-0248.2005.00828.x.
8. **Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernández M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P, Leadbetter JR.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature **450:**560–565. http://dx.doi.org/10.1038/nature06269.
9. **Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM.** 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science **331:** 463–467. http://dx.doi.org/10.1126/science.1200387.
10. **Distel DL, Roberts SJ.** 1997. Bacterial endosymbionts in the gills of the deep-sea wood-boring bivalves *Xylophaga atlantica* and *Xylophaga washingtona*. Biol. Bull. **192:**253–261. http://dx.doi.org/10.2307/1542719.
11. **Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI.** 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature **444:**1027–1031. http://dx.doi.org/10.1038/nature05414.
12. **Douglas AE.** 2009. The microbial dimension in insect nutritional ecology. Funct. Ecol. **23:**38–47. http://dx.doi.org/10.1111/j.1365-2435.2008.01442.x.
13. **Nasir H, Noda H.** 2003. Yeast-like symbiotes as a sterol source in anobiid beetles (Coleoptera, Anobiidae): possible metabolic pathways from fungal sterols to 7-dehydrocholesterol. Arch. Insect Biochem. Physiol. **52:** 175–182. http://dx.doi.org/10.1002/arch.10079.
14. **Brucker RM, Bordenstein SR.** 2013. The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus *Nasonia*. Science **341:** 667–669. http://dx.doi.org/10.1126/science.1240659.
15. **Robinson CJ, Bohannan BJ, Young VB.** 2010. From structure to function: the ecology of host-associated microbial communities. Microbiol. Mol. Biol. Rev. **74:**453–476. http://dx.doi.org/10.1128/MMBR.00014-10.
16. **Ley RE, Peterson DA, Gordon JI.** 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell **124:** 837–848. http://dx.doi.org/10.1016/j.cell.2006.02.017.
17. **Costello EK, Stagaman K, Dethlefsen L, Bohannan BJ, Relman DA.** 2012. The application of ecological theory toward an understanding of the human microbiome. Science **336:**1255–1262. http://dx.doi.org/10.1126/science.1224203.
18. **Boon E, Meehan CJ, Whidden C, Wong DH, Langille MG, Beiko RG.** 2014. Interactions in the microbiome: communities of organisms and communities of genes. FEMS Microbiol. Rev. **38:**90–118. http://dx.doi.org/10.1111/1574-6976.12035.
19. **Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, Henrissat B, Knight R, Gordon JI.** 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science **332:**970–974. http://dx.doi.org/10.1126/science.1198719.
20. **Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI.** 2008. Evolution of mammals and their gut microbes. Science **320:**1647–1651. http://dx.doi.org/10.1126/science.1155725.
21. **Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, Thomas T.** 2012. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. Proc. Natl. Acad. Sci. U. S. A. **109:**E1878–E1887. http://dx.doi.org/10.1073/pnas.1203287109.
22. **Mason CJ, Raffa KF.** 2014. Acquisition and structuring of midgut bacterial communities in gypsy moth (Lepidoptera: Erebidae) larvae. Environ. Entomol. **43:**595–604. http://dx.doi.org/10.1603/EN14031.
23. **Kaltenpoth M, Steiger S.** 2014. Unearthing carrion beetles' microbiome: characterization of bacterial and fungal hindgut communities across the Silphidae. Mol. Ecol. **23:**1251–1267. http://dx.doi.org/10.1111/mec.12469.
24. **Yun JH, Roh SW, Whon TW, Jung MJ, Kim MS, Park DS, Yoon C, Nam YD, Kim YJ, Choi JH, Kim JY, Shin NR, Kim SH, Lee WJ, Bae JW.** 2014. Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. Appl. Environ. Microbiol. **80:**5254–5264. http://dx.doi.org/10.1128/AEM.01226-14.
25. **Ochman H, Worobey M, Kuo CH, Ndjango JB, Peeters M, Hahn BH, Hugenholtz P.** 2010. Evolutionary relationships of wild hominids recapitulated by gut microbial communities. PLoS Biol. **8:**e1000546. http://dx.doi.org/10.1371/journal.pbio.1000546.
26. **Walter J, Britton RA, Roos S.** 2011. Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm. Proc. Natl. Acad. Sci. U. S. A. **108** (Suppl 1):4645–4652. http://dx.doi.org/10.1073/pnas.1000099107.
27. **Engel P, Moran NA.** 2013. Functional and evolutionary insights into the simple yet specific gut microbiota of the honey bee from metagenomic analysis. Gut Microbes **4:**60–65. http://dx.doi.org/10.4161/gmic.22517.
28. **Colman DR, Toolson EC, Takacs-Vesbach CD.** 2012. Do diet and taxonomy influence insect gut bacterial communities? Mol. Ecol. **21:** 5124–5137. http://dx.doi.org/10.1111/j.1365-294X.2012.05752.x.
29. **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. http://dx.doi.org/10.1128/AEM.71.11.6590-6599.2005.
30. **Kautz S, Rubin BE, Russell JA, Moreau CS.** 2013. Surveying the microbiome of ants: comparing 454 pyrosequencing with traditional methods to uncover bacterial diversity. Appl. Environ. Microbiol. **79:**525–534. http://dx.doi.org/10.1128/AEM.03107-12.
31. **Sanders JG, Powell S, Kronauer DJ, Vasconcelos HL, Frederickson ME, Pierce NE.** 2014. Stability and phylogenetic correlation in gut microbiota: lessons from ants and apes. Mol. Ecol. **23:**1268–1283. http://dx.doi.org/10.1111/mec.12611.
32. **Mueller U, Gerardo N, Aanen D, Six D, Schultz T.** 2005. The evolution of agriculture in insects. Annu. Rev. Ecol. Evol. Sys. **36:**563–595. http://dx.doi.org/10.1146/annurev.ecolsys.36.102003.152626.
33. **Holldobler B, Wilson EO.** 2008. The superorganism: the beauty, elegance, and strangeness of insect societies. W. W. Norton, New York, NY.
34. **Paine TD, Raffa KF, Harrington TC.** 1997. Interactions among scolytid bark beetles, their associated fungi, and live host conifers. Annu. Rev. Entomol. **42:**179–206. http://dx.doi.org/10.1146/annurev.ento.42.1.179.
35. **Pinto-Tomás AA, Anderson MA, Suen G, Stevenson DM, Chu FS, Cleland WW, Weimer PJ, Currie CR.** 2009. Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. Science **326:**1120–1123. http://dx.doi.org/10.1126/science.1173036.
36. **Suen G, Scott JJ, Aylward FO, Adams SM, Tringe SG, Pinto-Tomás AA, Foster CE, Pauly M, Weimer PJ, Barry KW, Goodwin LA, Bouffard P, Li L, Osterberger J, Harkins TT, Slater SC, Donohue TJ, Currie CR.** 2010. An insect herbivore microbiome with high plant biomass-degrading capacity. PLoS Genet. **6:**e1001129. http://dx.doi.org/10.1371/journal.pgen.1001129.
37. Reference deleted.
38. **Aylward FO, Currie CR, Suen G.** 2012. The evolutionary innovation of nutritional symbioses in leaf-cutter ants. Insects **3:**41–61. http://dx.doi.org/10.3390/insects3010041.
39. **Aylward FO, Burnum-Johnson KE, Tringe SG, Teiling C, Tremmel DM, Moeller J, Scott JJ, Barry KW, Piehowski PD, Nicora CD, Malfatti S, Monroe ME, Purvine SO, Goodwin LA, Smith RD, Weinstock GM, Gerardo NM, Suen G, Lipton MS, Currie CR.** 2013. *Leucoagaricus gongylophorus* produces diverse enzymes for the degradation of recalcitrant plant polymers in leaf-cutter ant fungus gardens. Appl. Environ. Microbiol. **79:**3770–3778. http://dx.doi.org/10.1128/AEM.03833-12.
40. **Six DL, Klepzig KD.** 2004. *Dendroctonus* bark beetles as model systems for the study of symbioses. Symbiosis **37:**207–232.
41. **Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV.** 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. **29:**22–28. http://dx.doi.org/10.1093/nar/29.4.e22.
42. **Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammier EL, Bateman A.** 2008. The Pfam protein families database. Nucleic Acids Res. **36**(Database issue): D281–D288. http://dx.doi.org/10.1093/nar/gkm960.
43. **Munkacsi AB, Pan JJ, Villesen P, Mueller UG, Blackwell M, McLaughlin

DJ. 2004. Convergent coevolution in the domestication of coral mushrooms by fungus-growing ants. Proc. Biol. Sci. **271:**1777–1782. http://dx.doi.org/10.1098/rspb.2004.2759.
44. **Sharon G, Segal D, Ringo JM, Hefetz A, Zilber-Rosenberg I, Rosenberg E.** 2010. Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. **107:**20051–20056. http://dx.doi.org/10.1073/pnas.1009906107.
45. **Diaz Heijtz R, Wang S, Anuar F, Qian Y, Björkholm B, Samuelsson A, Hibberd ML, Forssberg H, Pettersson S.** 2011. Normal gut microbiota modulates brain development and behavior. Proc. Natl. Acad. Sci. U. S. A. **108:**3047–3052. http://dx.doi.org/10.1073/pnas.1010529108.
46. **Weber NA.** 1966. Fungus-growing ants. Science **153:**587–604. http://dx.doi.org/10.1126/science.153.3736.587.
47. **Weber NA.** 1972. Gardening ants: the Attines (Memoirs of the Amer Phil Society), vol **92**. American Philosophical Society, Philadelphia, PA.
48. **Ruelle JE, Carton WGH, Sheasby JL.** 1975. National survey of the Isoptera of Southern Africa. 8. The genus Macrotermes Holmgren (Termitidae: Macrotermitinae). Cimbebasia **A3:**74–94.
49. **Wood SL,** Bright D. 1992. A catalog of Scolytidae and Platypodidae (Coleoptera), part 2. Taxonomic index, vol A. Great Basin Naturalist Memoirs, no 13. Brigham Young University, Provo, UT.
50. **Adams AS, Aylward FO, Adams SM, Erbilgin N, Aukema BH, Currie CR, Suen G, Raffa KF.** 2013. Mountain pine beetles colonizing historical and naive host trees are associated with a bacterial community highly enriched in genes contributing to terpene metabolism. Appl. Environ. Microbiol. **79:**3468–3475. http://dx.doi.org/10.1128/AEM.00068-13.
51. **Aylward FO, Burnum KE, Scott JJ, Suen G, Tringe SG, Adams SM, Barry KW, Nicora CD, Piehowski PD, Purvine SO, Starrett GJ, Goodwin LA, Smith RD, Lipton MS, Currie CR.** 2012. Metagenomic and metaproteomic insights into bacterial communities in leaf-cutter ant fungus gardens. ISME J. **6:**1688–1701. http://dx.doi.org/10.1038/ismej.2012.10.
52. **Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM, Rothberg JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376–380. http://dx.doi.org/10.1038/nature03959.
53. **García Martín H, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC, Yeates C, He S, Salamov AA, Szeto E, Dalin E, Putnam NH, Shapiro HJ, Pangilinan JL, Rigoutsos I, Kyrpides NC, Blackall LL, McMahon KD, Hugenholtz P.** 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nat. Biotechnol. **24:**1263–1269. http://dx.doi.org/10.1038/nbt1247.
54. **Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF.** 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. **75:**7537–7541. http://dx.doi.org/10.1128/AEM.01541-09.
55. **Schloss PD, Gevers D, Westcott SL.** 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One **6:**e27310. http://dx.doi.org/10.1371/journal.pone.0027310.
56. **Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF, Sloan WT.** 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. Nat. Methods **6:**639–641. http://dx.doi.org/10.1038/nmeth.1361.
57. **Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO.** 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. **41:**D590–D596. http://dx.doi.org/10.1093/nar/gks1219.
58. **Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R.** 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27:**2194–2200. http://dx.doi.org/10.1093/bioinformatics/btr381.
59. **Katoh K, Kuma K, Toh H, Miyata T.** 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. **33:**511–518. http://dx.doi.org/10.1093/nar/gki198.
60. **Schloss PD.** 2010. The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. PLoS Comput. Biol. **6:**e1000844. http://dx.doi.org/10.1371/journal.pcbi.1000844.
61. **Huse SM, Welch DM, Morrison HG, Sogin ML.** 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environ. Microbiol. **12:**1889–1898. http://dx.doi.org/10.1111/j.1462-2920.2010.02193.x.
62. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410. http://dx.doi.org/10.1006/jmbi.1990.9999.
63. **Brady A, Salzberg SL.** 2009. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. Nat. Methods **6:**673–676. http://dx.doi.org/10.1038/nmeth.1358.
64. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119. http://dx.doi.org/10.1186/1471-2105-11-119.
65. **Letunic I, Bork P.** 2007. Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics **23:**127–128. http://dx.doi.org/10.1093/bioinformatics/btl529.
66. **Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and psi-blast: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.
67. **Santos SR, Ochman H.** 2004. Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. Environ. Microbiol. **6:**754–759. http://dx.doi.org/10.1111/j.1462-2920.2004.00617.x.
68. **Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM, Tiedje JM.** 2007. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. Nucleic Acids Res. **35:**D169–D172. http://dx.doi.org/10.1093/nar/gkl889.
69. **Eddy SR.** 2009. A new generation of homology search tools based on probabilistic inference. Genome Inform. **23:**205–211.
70. **Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T.** 2009. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25:**1972–1973. http://dx.doi.org/10.1093/bioinformatics/btp348.
71. **Price MN, Dehal PS, Arkin AP.** 2010. FastTree2—approximately maximum-likelihood trees for large alignments. PLoS One **5:**e9490. http://dx.doi.org/10.1371/journal.pone.0009490.
72. **Anisimova M, Gascuel O.** 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst. Biol. **55:**539–552. http://dx.doi.org/10.1080/10635150600755453.
73. **Konstantinidis KT, Tiedje JM.** 2005. Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U. S. A. **102:**2567–2572. http://dx.doi.org/10.1073/pnas.0409727102.
74. **Felsenstein J.** 1989. Notices: PHYLIP—phylogeny inference package (version 3.2). Cladistics **5:**163–166. http://dx.doi.org/10.1111/j.1096-0031.1989.tb00562.x.
75. **Néron B, Ménager H, Maufrais C, Joly N, Maupetit J, Letort S, Carrere S, Tuffery P, Letondal C.** 2009. Mobyle: a new full web bioinformatics framework. Bioinformatics **25:**3005–3011. http://dx.doi.org/10.1093/bioinformatics/btp493.
76. **Markowitz VM, Ivanova NN, Szeto E, Palaniappan K, Chu K, Dalevi D, Chen IM, Grechkin Y, Dubchak I, Anderson I, Lykidis A, Mavromatis K, Hugenholtz P, Kyrpides NC.** 2008. IMG/M: a data management and analysis system for metagenomes. Nucleic Acids Res. **36**(Suppl 1):D534–D538. http://dx.doi.org/10.1093/nar/gkm869.
77. **Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F.** 2010. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. Cold Spring Harb. Protoc. 2010:pdb.prot5368. http://dx.doi.org/10.1101/pdb.prot5368.
78. **R Core Team.** 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
79. **Trautwein MD, Wiegmann BM, Beutel R, Kjer KM, Yeates DK.** 2012. Advances in insect phylogeny at the dawn of the postgenomic era. Annu. Rev. Entomol. **57:**449–468. http://dx.doi.org/10.1146/annurev-ento-120710-100538.