PNAS
www.pnas.org

Supplementary Information for

Diplogastrellus nematodes are sexually transmitted mutualists that alter the bacterial and fungal communities of their beetle host

Cristina C. Ledón-Rettig, Armin P. Moczek, and Erik J. Ragsdale

Cristina Ledón-Rettig
Email: crisledo@indiana.edu

This PDF file includes:

Supplementary Methods
Tables S1 to S3
Figs. S1 to S5
Captions for datasets S1 to S7
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Datasets S1 to S7

www.pnas.org/cgi/doi/10.1073/pnas.1809606115

Supplementary Materials and Methods

Identification and culture of nematodes. Genitalia from field-collected *O. taurus* beetles from North Carolina and Indiana were dissected and placed in M9 solution in watch glasses. Incidence of worms in the North Carolina population was recorded among males and female beetles (Table S1), as well as the approximate numbers in each; of those female beetles with genital nematodes, all were characterized as having 'few" (< 20), while all male beetles with genital nematodes were characterized as having "many" (> 50). Dauer nematodes were picked from the watch glasses either (i) into lysis buffer (10mM Tris-HCL; 50mM KCL; 2.5 mM $MgCl_2$; 0.45%NP40; 0.45% Tween20) with Proteinase K for lysis and molecular analysis or (ii) onto nematode growth medium (NGM) plates with OP50 bacterial lawns for culturing. Molecular sequencing of the long subunit (LSU) ribosomal RNA (rRNA) gene, using primers LSU_D2A; 5'-CAAGTACCGTGAGGGAAAGTTG-3' and LSU_D3B26R; 5'-TCGGAAGGAACCAGCTACTA-3' (1), and of the small subunit (SSU) rRNA gene, using primers SSU_18A; 5'-AAAGATTAAGCCATGCAT-3' and SSU_26R; 5'-CATTCTTGGCAAATGCTTTCG-3' (1), invariably identified all isolates from both locations (North Carolina and Indiana) as a single putative species of *Diplogastrellus.* Morphological assessment of cultured individuals identified this species as *D. monhysteroides* (Fig. 1). A bacterium carried with *D. monhysteroides* was isolated, cultured, and used as the bacterial lawn for all future *D. monhysteroides* cultures. To identify the isolate, we amplified and sequenced the variable V3 region of the 16S rRNA gene using primers to conserved regions of the 16S rRNA gene (2); based on this sequence this bacterium was most closely related to species of *Providencia* (Enterobacteriaceae).

Quantification of relative bacterial and fungal abundances. The same 12 samples used for microbial profiling were used for real-time, quantitative PCR (RT-qPCR). An additional 6 artificial brood balls were generated (3 control and 3 *D. monhysteroides*-treated) exclusively for RT-qPCR. We extracted DNA from these samples at the early time-point, *i.e.*, 7 days post-inoculation; although it was our intention to also collect late time-point samples from the same brood balls, all beetle individuals in the control treatment died by 21 days, such that all late time-point samples from these additional brood balls were necessarily excluded. We performed RT-qPCR on DNA from our samples in 10-µL volumes using the SYBR green Supermix (Applied

Biosystems, Foster City, CA, USA) on a Roche LightCycler 480 (Roche, Basel, Switzerland). Each well in a 384-well plate contained 1.5 µL water, 0.5 µL of a 10uM forward primer, 0.5 µL of a 10 µM reverse primer, 5 µL of SYBR Green Master Mix, and 2.5 µL of DNA template. For increased accuracy, two runs were performed where each run included all samples (*n* = 42) in triplicate. The RT-qPCR annealing temperature was 49 °C and the extension time was 45 sec. A fragment of the 16S ribosomal ribonucleic acid (rRNA) gene was amplified using the primers Eub338 (5'-ACTCCTACGGGAGGCAGCAG-3') and Eub518 (5'-ATTACCGCGGCTGCTGG-3') for bacteria and a fragment spanning the SSU internal transcribed spacer (ITS) region and the 5.8S rRNA gene for fungi using primers ITS1f (5'-TCCGTAGGTGAACCTGCGG-3') and 5.8S (5'-CGCTGCGTTCTTCATCG-3') (3).

For the bacterial standard, a 203-bp, double-stranded DNA oligomer was synthesized (gBLOCK, Integrated DNA Technologies; Table S3) to span the 16S region of the *Escherichia coli* genome. For the fungal standard, a 333-bp, double-stranded DNA oligomer was synthesized (gBLOCK, Integrated DNA Technologies; Table S3) to span the 18S internal transcribed spacer (ITS) region and 5.8S rRNA gene of *Coprinopsis cinerea*, a fungal species commonly found in dung (4). For each taxon's standard, lyophilized material was used to prepare a 10 nM stock solution that was serially diluted to produce ten standards spanning the 10-1 x 10^8 copy number range; copy number was based on the weight of the lyophilized material, the average weight of a base pair (*i.e.*, 650 Daltons) and the length of the product. Plots of *Cp* versus $\log_{10}$(copy number) afforded linear calibration curves with amplification efficiencies between 91-95% and 72-73% for *E. coli* and *C. cinerea*, respectively, and R^2 values of 1.00 and 0.96-0.98 for these taxon standards, respectively. Based on melting-curve analyses, we found no evidence for primer dimers. We estimated fungal and bacterial abundance based on the estimated gene copy number from their respective standard curves and used these values to calculate fungal to bacterial ratios (F:B). Natural log-transformed ratios were used as the response variable in linear mixed models, which were executed using the R package nlme (5). Fixed factors were (i) dung type and (ii) worm treatment plus the interaction between (i) and (ii), and the random variable was the samples, which accounted for the shared variance between the two measures from each sample. Significance of the fixed factors was determined by comparing nested

models with likelihood ratio tests. Where necessary, multiple comparisons were tested using the R package lsmeans, which can accommodate mixed-effects models (6).

To address the possibility that nematodes might be differentially modifying the abundance of bacteria in previously frozen but not fresh dung, such that the correlated fitness effects might not be rendered under more natural conditions, we also assessed absolute abundances of bacteria as measured by RT-qPCR. We found that, while dung type was not a significant factor in determining bacterial copy number, nematode treatment was significant (χ^2= 4.2, df = 1, *P* = 0.04; Figure S1). This result suggests that, while the inclusion of *D. monhysteroides* increases bacterial load, it does equally across both dung types. Thus, the correlated effects on fitness observed in beetles raised on frozen dung are not likely due to an increase in bacteria that is unique to frozen dung. Likewise, we found that dung type was not a significant factor in determining fungal copy number, and instead we found that nematode treatment was significant (χ^2= 7.6, df = 1, *P* = 0.01; Figure S2). Importantly, treating brood balls with *D. monhysteroides* increased fungal load much more than it increased bacterial load (an approximately 1000-fold *vs*. 10-fold increase), likely explaining why the ratio of fungi to bacteria was higher in our *D. monhysteroides*-treated late time point samples.

Microbial profiling of brood balls. Before library construction, DNA was run on a DNA Tapestation (Agilent Technologies, Santa Clara, CA, USA) to ensure quality. Libraries were constructed using Bioo Scientific kits for the V4 hypervariable domain of the microbial 16S RNA gene to characterize bacterial communities and for the ITS1 and ITS2 hypervariable internal transcribed spacer regions of SSU rRNA to characterize fungal communities. Libraries were sequenced using paired-end sequencing on the Illumina MiSeq sequencing platform.

Statistical analyses of 16S and ITS sequences. Sequence information was used to assess aspects of brood ball microbial ecology, specifically (i) community identity and composition and (ii) specific taxa that differed among treatments. Following 16S rRNA-gene sequencing, reads were processed using the 2017.10 release of QIIME 2 (7). Reads were first demultiplexed into libraries and only reads with an exact match to a known library barcode were retained. From

within QIIME 2, DADA2 (8) was used to remove primer sequences from the reads, perform error correction, identify sequence variants, and remove chimeric sequences. This command was run with the following parameters "—p-n-reads-learn 4000000 –p-trim-left-f 31 –p-trim-left-r 33 –p-trunc-len-f 190 –p-trunc-len-r 150". Sequence variants were then classified by a naïve Bayes classifier within QIIME 2 using release 128 of the Silva SSU database (9). An abundance filter was applied, and sequence variants that did not compose at least 0.1% of at least one library were removed from all libraries. A multiple-sequence alignment was generated using MAFFT (10), and a phylogenetic tree inferred using FastTree (11), both as implemented from within QIIME 2 with default parameters. At this point, all remaining libraries were rarified to the count of the smallest library (65,484 reads).

Following ITS sequencing, reads were demultiplexed using the 2017.10 release of QIIME 2 requiring an exact barcode match. Because the forward and reverse reads do not overlap for this amplicon, DADA2 was used to trim forward and reverse reads to 200 and 100 bp, respectively. For both forward and reverse reads, the expected number of errors was calculated from the quality scores and any read pair with more than two expected errors in either the forward or reverse read was removed from further analysis. Cutadapt (12) was used to trim primer sequences from the sequence variants identified by DADA2. Sequences that did not contain these primer sequences were discarded. VSEARCH (13) was used to perform *de novo* chimera detection and removal. The remaining ITS sequence variants were clustered into operational taxonomic units (OTUs) at 97% similarity using AbundantOTU+ (14). The software package ITSx (15) was then used to filter non-ITS and non-fungal OTUs from our data set. We then classified our OTUs using the Naïve Bayes Classifier in QIIME 2 against the UNITE database version 7.2 (16). To confirm that our OTUs represented fungal taxa, we performed BLAST searches of our OTUs against the NCBI nt database (Dataset S1).

Following this initial processing for bacterial and fungal sequences, OTU tables containing read counts, taxonomy information (where possible), and sample metadata were exported from QIIME and imported into R (17) for visualization and statistical assessment using phyloseq (18). To exclude bacterial sequences observed at very low frequencies, any OTUs that did not represent at least 0.1% of the total number of sequences in at least one sample were

removed from all samples, resulting in 6.9% of the data being removed. To exclude fungal sequences observed at very low frequencies, any OTUs that did not represent at least 1% of the total number of sequences in at least one sample were removed, resulting in 1.4% of the data being removed.

To quantify host-associated bacterial community identity and composition, we calculated pairwise Unifrac distances (19); specifically, we assessed both abundance (weighted analyses) and presence/absence information (unweighted analyses). By considering both measures, we could assess the relative importance of abundance and phylogenetic lineage on the variation in community composition between treatments (20). To quantify host-associated fungal community identity and composition, we used Bray-Curtis dissimilarity measures. One sample from the early control group was an extreme outlier and was removed, and this removal had no qualitative impact on the results. We assessed these metrics (Unifrac distances and Bray-Curtis dissimilarities for bacterial and fungal data, respectively) with PERMANOVAs using the adonis function in the R package vegan (21).

Supplementary tables

Table S1. Incidence of *D. monhysteroides* in NC field-collected beetles.

	Total collected	Number with worms	Worm Incidence
Females	26	3	12%
Males	15	5	33%

Table S2. Primers.

Primer	Application	Sequence (5'-3')	Reference
LSU_D2A	species identification	CAAGTACCGTGAGGGAAAGTTG	Blaxter et al. 1998
LSU_D3B26R	species identification	TCGGAAGGAACCAGCTACTA	Blaxter et al. 1998
SSU_18A	species identification	AAAGATTAAGCCATGCAT	Blaxter et al. 1998
SSU_26R	species identification	CATTCTTGGCAAATGCTTTCG	Blaxter et al. 1998
Eub338	qPCR, bacterial estimate	ACTCCTACGGGAGGCAGCAG	Fierer et al. 2005
Eub518	qPCR, bacterial estimate	ATTACCGCGGCTGCTGG	Fierer et al. 2005
5.8s	qPCR, fungal estimate	TCCGTAGGTGAACCTGCGG	Fierer et al. 2005
ITS1f	qPCR, fungal estimate	TCCGTAGGTGAACCTGCGG	Fierer et al. 2005

Table S3. gBlocks.

gBlock	Application	Sequence (5'-3')	reference genome
LSU_D2A	bacterial standard	CAGACTCCTACGGGAGGCAGCAG TGGGGAATATTGCACAATGGGCGC AAGCCTGATGCAGCCATGCCGCG TGTATGAAGAAGGCCTTCGGGTTG TAAAGTACTTTCAGCGGGGAGGAA GGGAGTAAAGTTAATACCTTTGCT CATTGACGTTACCCGCAGAAGAAG CACCGGCTAACTCCGTGCCAGCA GCCGCGGTAATACG	*Escherichia coli*
LSU_D3B26R	fungal standard	GTTTCCGTAGGTGAACCTGCGGAA GGATCATTATTGAATAAATCTGAC GTGGTTGTCGCTGGCTCCCCCGG GAGCATGTGCACGCCCGTCACCTT TATTTCTCCACCTGTGCACACACT GTAGGCCTGGATACCTCTCGTCGC AAGGCGGATGCGTGGCTTGCTGTC GCTTTCGAAAGAAGGCCGGCTTGC CATGAATTTCCAGGTCTATGATTTC TTACACACCCCAAACTGAATGTTA TGGAATGTCATCTCAAGGCCTTGT GCCTATAAACCTATACAACTTTCA GCAACGGATCTCTTGGCTCTCGCA TCGATGAAGAACGCAGCGAAA	*Coprinopsis cinerea*

Supplementary figures

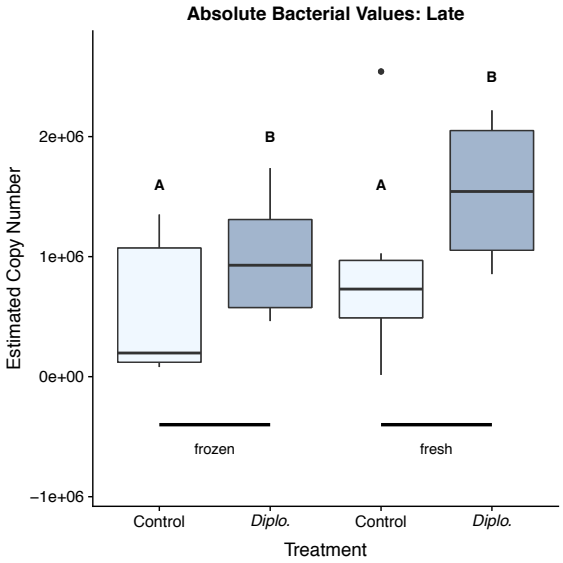

Fig. S1. The inclusion of *D. monhysteroides* in brood balls increases bacterial abundance equally across dung types. A comparison between frozen and fresh dung revealed that the same nematode-mediated mechanisms governing alterations to bacterial load are likely operating across samples generated from either previously frozen or fresh dung. Abbr: *Diplo.*, *D. monhysteroides*.

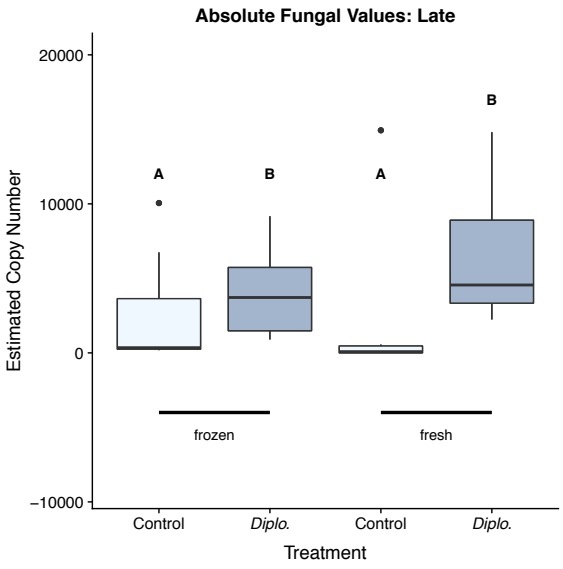

Fig. S2. The inclusion of *D. monhysteroides* in brood balls increases fungal abundance equally across dung types. A comparison between frozen and fresh dung revealed that the same nematode-mediated mechanisms governing alterations to fungal load are likely operating across samples generated from either previously frozen or fresh dung. Abbr: *Diplo.*, *D. monhysteroides*.

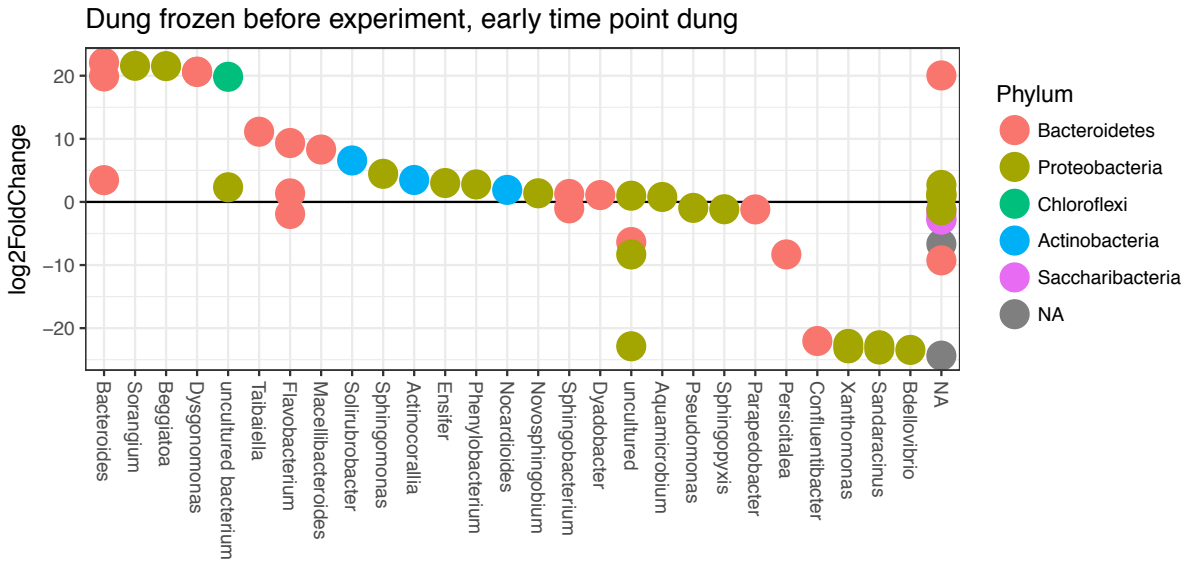

Fig. S3. Differentially abundant bacterial taxa at early development of larvae raised on previously frozen dung. To determine what bacterial taxa were significantly abundant between control and *D. monhysteroides* brood balls, negative binomial modeling was conducted using DESeq2 (22). The $\log_2$(fold change) of each taxon's abundance is plotted against the genus to which that taxon belongs: positive values indicate those higher in *D. monhysteroides*-treated brood balls; negative values, those higher in control brood balls.

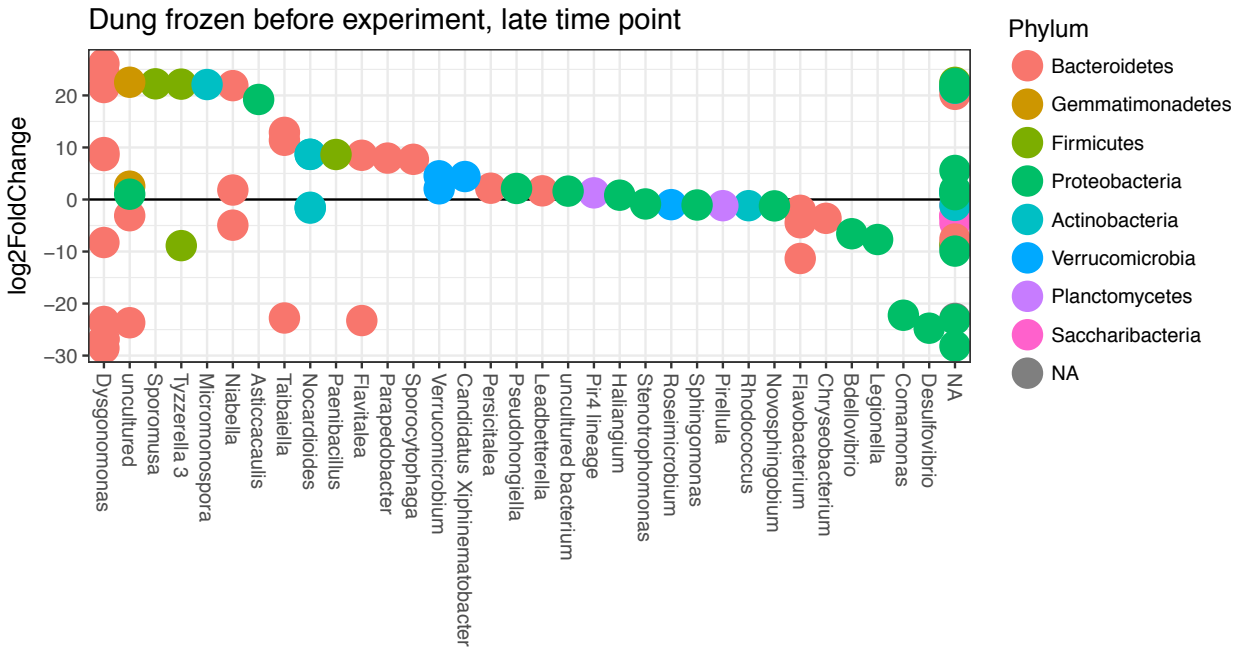

Fig. S4. Differentially abundant bacterial taxa at late development of larvae raised on previously frozen dung. To determine what bacterial taxa were significantly abundant between control and *D. monhysteroides* brood balls, negative binomial modeling was conducted using DESeq2 (22). The $\log_2$(fold change) of each taxon's abundance is plotted against the genus to which that taxon belongs: positive values indicate those higher in *D. monhysteroides*-treated brood balls; negative values, those higher in control brood balls.

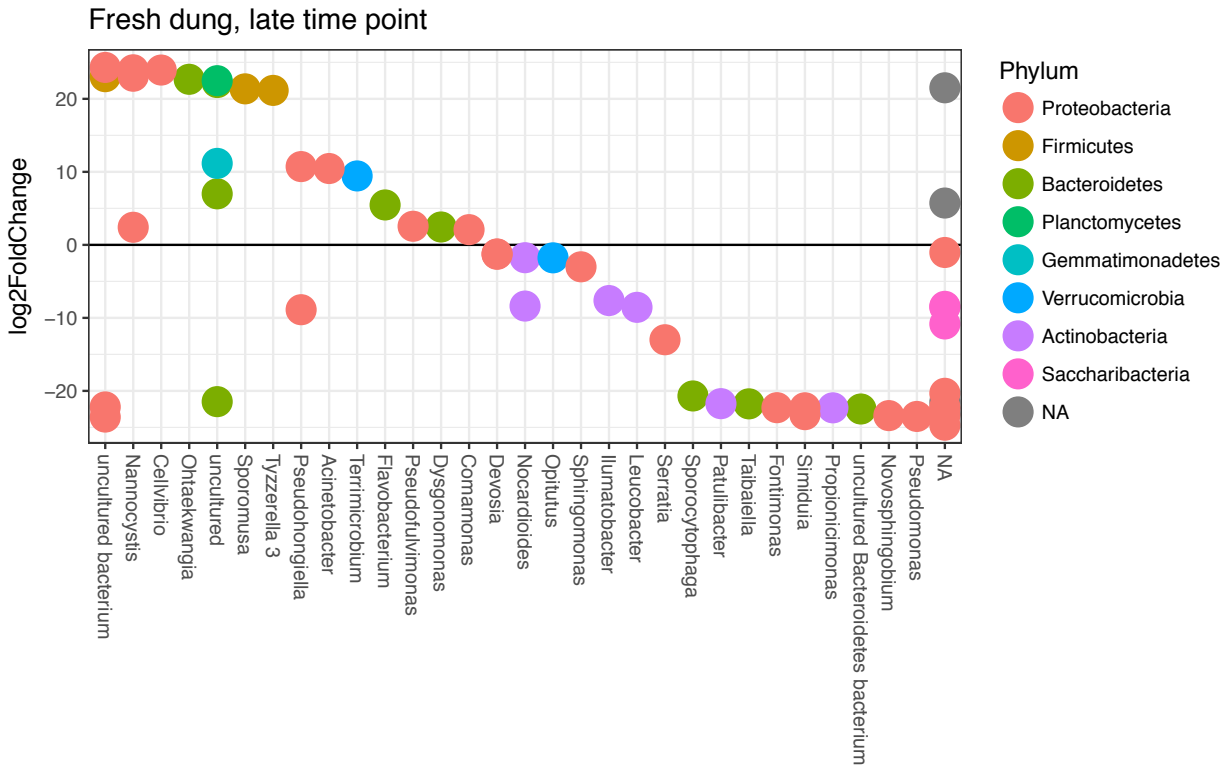

Fig. S5. Differentially abundant bacterial taxa at late development of larvae raised on fresh dung. To determine what bacterial taxa were significantly abundant between control and *D. monhysteroides* brood balls, negative binomial modeling was conducted using DESeq2 (22). The $\log_2$(fold change) of each taxon's abundance is plotted against the genus to which that taxon belongs: positive values indicate those higher in *D. monhysteroides*-treated brood balls; negative values, those higher in control brood balls.

Supplementary Datasets

Dataset S1. Fungal taxa as classified by both the UNITE fungal database and the more general NCBI nt database.

Dataset S2. Candidate bacterial taxa differentially regulated by the presence of *D. monhysteroides* in frozen dung 7 days post inoculation.

Dataset S3. Candidate bacterial taxa differentially regulated by the presence of *D. monhysteroides* in frozen dung 21 days post inoculation.

Dataset S4. Candidate bacterial taxa differentially regulated by the presence of *D. monhysteroides* in fresh dung 21 days post inoculation.

Dataset S5. Candidate fungal taxa differentially regulated by the presence of *D. monhysteroides* in frozen dung 7 days post inoculation.

Dataset S6. Candidate fungal taxa differentially regulated by the presence of *D. monhysteroides* in frozen dung 21 days post inoculation.

Dataset S7. Candidate fungal taxa differentially regulated by the presence of *D. monhysteroides* in fresh dung 21 days post inoculation.

Supplementary references

1. Blaxter ML, et al. (1998) A molecular evolutionary framework for the phylum Nematoda. *Nature* 392(6671):71–75.
2. Muyzer G, De Waal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* 59(3):695–700.
3. Fierer N, Jackson J (2005) Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. *Appl Environ Microbiol* 71(7):4117.
4. Kües U (2000) Life history and developmental processes in the basidiomycete Coprinus cinereus. *Microbiol Mol Biol Rev* 64(2):316–353.
5. Pinheiro J, et al. (2017) nlme: linear and nonlinear mixed effects models. *R* 3(version):1–336.
6. Lenth R, Hervé M (2015) lsmeans: least-squares means. *R* (version 2.4).
7. Caporaso JG, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5):335–336.
8. Callahan BJ, et al. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13(7):581–583.
9. Quast C, et al. (2013) The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596.
10. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30(4):772–780.
11. Price MN, Dehal PS, Arkin AP (2010) FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3):9490.
12. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17(1):10.
13. Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584.
14. Ye Y (2010) Identification and quantification of abundant species from pyrosequences of 16S rRNA by consensus alignment. *Proceedings of the IEEE International Conference on*

Bioinformatics and Biomedicine, pp 153–157.

15. Bengtsson-Palme J, et al. (2013) Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol Evol* 4(10):914–919.
16. Kõljalg U, et al. (2013) Towards a unified paradigm for sequence-based identification of fungi. *Mol Ecol* 22(21):5271–5277.
17. R Core Team (2017) R: a language and environment for statistical computing. *R Found Stat Comput Vienna, Austria*.
18. McMurdie PJ, Holmes S (2013) Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8(4). doi:10.1371/journal.pone.0061217.
19. Lozupone C, Hamady M, Knight R (2006) UniFrac – an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* 7:371.
20. Lozupone CA, Hamady M, Kelley ST, Knight R (2007) Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* 73(5):1576–85.
21. Oksanen J, et al. (2017) vegan: community ecology package. *R* (version 2.4-4).
22. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.