Received: 13 January 2018 | Revised: 27 April 2018 | Accepted: 14 May 2018

DOI: 10.1111/mec.14728

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Functional variation in the gut microbiome of wild *Drosophila* populations

Alyssa Bost[1] | Vincent G. Martinson[2] | Soeren Franzenburg[1] | Karen L. Adair[1] | Alice Albasi[1] | Martin T. Wells[3] | Angela E. Douglas[1,4]

[1]Department of Entomology, Cornell University, Ithaca, New York

[2]Department of Biology, University of Rochester, Rochester, New York

[3]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York

[4]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York

Correspondence
Angela E. Douglas, Cornell University, 5134 Comstock Hall, Ithaca, NY 14853.
Email: aes326@cornell.edu

Present addresses
Vincent G. Martinson, Department of Entomology, University of Georgia, Athena, GA 30602.
Soeren Franzenburg, Institute for Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany.
Alice Albasi, Department of Mathematics, University of Pisa, 56127 Pisa PI, Italy.

Funding information
National Science Foundation, Grant/Award Number: BIO 1241099

Abstract

Most of the evidence that the gut microbiome of animals is functionally variable, with consequences for the health and fitness of the animal host, is based on laboratory studies, often using inbred animals under tightly controlled conditions. It is largely unknown whether these microbiome effects would be evident in outbred animal populations under natural conditions. In this study, we quantified the functional traits of the gut microbiota (metagenome) and host (gut transcriptome) and the taxonomic composition of the gut microorganisms (16S rRNA gene sequence) in natural populations of three mycophagous *Drosophila* species. Variation in microbiome function and composition was driven principally by the period of sample collection, while host function varied mostly with *Drosophila* species, indicating that variation in microbiome traits is determined largely by environmental factors, and not host taxonomy. Despite this, significant correlations between microbiome and host functional traits were obtained. In particular, microbiome functions dominated by metabolism were positively associated with host functions relating to gut epithelial turnover. Much of the functional variation in the microbiome could be attributed to variation in abundance of Bacteroidetes, rather than the two other abundant groups, the γ-Proteobacteria or Lactobacillales. We conclude that functional variation in the interactions between animals and their gut microbiome can be detectable in natural populations, and, in mycophagous *Drosophila*, this variation relates primarily to metabolism and homeostasis of the gut epithelium.

KEYWORDS
Drosophila, gut microbiome, metagenome, microbial biodiversity, mycophagous drosophilids, transcriptome

1 | INTRODUCTION

The great majority of animals harbour gut microorganisms of variable composition (Brooks, Kohl, Brucker, van Opstal & Bordenstein, 2016; McFall-Ngai et al., 2013), and there is increasing evidence that this microbial variation can influence the phenotype of the host (Hamilton, Peng, Boulanger & Perlman, 2016; Kohl, Weiss, Cox, Dale & Dearing, 2014; Sampson et al., 2016; Surana & Kasper, 2017). However, most of the studies investigating the relationship between the gut microbiome and host traits have two important limitations. First, the taxonomic composition of the microbiome is, to a very large extent, used as a proxy for microbial function. Microbial function should provide more robust associations with host phenotype because the functional traits of the microbiome that influence host phenotype may be general to many taxa (resulting in functional redundancy among taxa) or vary at a very fine taxonomic scale, for

example, strain-level variation (Doolittle & Booth, 2017; Louca, Parfrey & Doebeli, 2016; Moya & Ferrer, 2016). Second, most of the research relating the gut microbiome and host phenotype has been conducted in the laboratory, usually using highly inbred animals colonized with standardized sets of microorganisms under tightly controlled environmental conditions (Macpherson & McCoy, 2015; Newell & Douglas, 2014; Nguyen, Vieira-Silva, Liston & Raes, 2015; Zhang et al., 2017). The experimental protocols are designed to minimize experimental noise and so maximize the opportunity to detect an effect. An unfortunate consequence is that these experiments cannot address whether the observed relationship between host traits and the gut microbiome is important in natural populations of animals, including humans. These caveats in our understanding can be resolved by research on the functional traits of the gut microbiome in natural populations of animals.

The purpose of this study was to quantify the pattern of functional variation in the gut microbiome of animals under natural conditions and then determine how microbiome function varies with both the taxonomic composition of the microbiome and host functional traits. The analysis was carried out on three mycophagous *Drosophila* species: *D. falleni*, *D. neotestacea* and *D. putrida*. These species co-occur on basidiomycete mushrooms, which they use as their food source and breeding site (Jaenike & James, 1991), and they harbour a taxonomically variable gut microbiota that is different from the microbiology of their food substrate (Martinson, Douglas & Jaenike, 2017). These mycophagous species are both phylogenetically distant and biologically different from *D. melanogaster* (Izumitani, Kusaka, Koshikawa, Toda & Katoh, 2016; Werner & Jaenike, 2017), which is the focus of most microbiome research on drosophilid flies (Broderick & Lemaitre, 2012). Whereas the mycophagous species studied here are native to natural habitats in NE America (Werner & Jaenike, 2017), *D. melanogaster* is a human commensal that originated in tropical Africa and feeds on rotting fruits, not mushrooms (Keller, 2007). The gut microbiota of the mycophagous drosophilids is both more diverse and taxonomically different from the microbiota in *D. melanogaster* (Wong, Chaston & Douglas, 2013).

In this study of the mycophagous *Drosophila* species, we adopted a molecular approach to quantify function, specifically parallel analysis of the metagenome and host transcriptome of guts dissected from field-collected flies, as indices of the functional capacity of the microbiome and host, respectively; and we also conducted 16S rRNA gene amplicon sequencing, to provide taxonomic context for the functional traits of the microbiome. Our set of 24 samples comprised pooled guts from either male or female flies of each of the three *Drosophila* collected on four time periods from one location. We discovered that variation in gut microbiome function was driven principally by environmental factors, rather than by the sex or species of the host. Despite this, we identified functional orthologs with correlated abundance in the metagenomes that were significantly correlated with the expression of certain host genes, and these correlations were particularly evident for metabolism-related traits. This analysis demonstrates correspondence between function of the host and its gut microbiome that relates to specific functional traits.

2 | MATERIALS AND METHODS

2.1 | The *Drosophila* samples

Adult *D. falleni*, *D. neotestacea* and *D. putrida* were collected from natural populations on *Russula*, *Amanita* and Boletaceae mushrooms in Mendon Ponds Park, Rochester, NY, USA (42°N, 78°W), with an insect net during four periods: August 8–19, 2013 (Collection-1), October 4–30, 2013 (Collection-2), July 27–30, 2014 (Collection-3), and August 22–27 2014 (Collection-4). On return to the laboratory, each fly was chill-anesthetized and identified to species and sex using the key of Werner and Jaenike (2017). Selected flies were rinsed in ice-cold 70% ethanol, then ice-cold sterile PBS and immersed in RNAlater (Thermo Fisher, Waltham, MA). The flies were immediately dissected under 20 × magnification on glass slides previously cleaned with RNase-AWAY (Thermo Fisher). The isolated guts (comprising the crop, midgut and hindgut) of 20–30 flies of the same sex and species were pooled in 0.5 ml TRIZOL reagent (Life Technologies, Carlsbad, CA) and stored at −80°C, to give a total of 24 samples (two sexes of three species for each of four collections).

2.2 | Nucleic acid extraction

All 24 samples of pooled fly guts and four negative controls comprising template-free samples were homogenized in 500 μl TRIzol with 500 μl 0.1-mm glass beads (Scientific Industries, Bohemia, NY) on an MP Fastprep24 at 4.5 M/S for 2 × 30 s. The procedure of Bost et al. (2017) was used for dual extraction of DNA and RNA. Briefly, total nucleic acids were extracted by phase separation with chloroform. RNA was then purified from the aqueous phase using the RNeasy Mini Kit (Qiagen, Hilden, Germany), including on-column DNase treatment, and eluted in 50 μl RNase-free water. DNA was obtained from the organic phase using back extraction buffer (4 M guanidine thiocyanate, 50 mM sodium citrate and 1 M Tris) following the TRI Reagent DNA/protein isolation protocol (Thermo Fisher, Waltham, MA) and then purified using the DNAeasy Blood & Tissue kit (Qiagen), following manufacturer's instructions. The purity of nucleotide extractions was checked by Nanodrop and the concentration determined by Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA) with RNA BR Assay Kit for the RNA samples and the dsDNA HS Assay Kit for DNA. The extracted DNA was amplified by multiple displacement amplification using the REPLI-g Mini Kit (Qiagen) for 14 hr, according to manufacturer's instructions.

2.3 | 16S rRNA gene amplicon sequencing

16S rRNA gene amplicon libraries were prepared using a dual-indexing strategy modified from Kozich, Westcott, Baxter, Highlander and Schloss (2013) targeting the V3–V4 region with primers 341F (5′-CCTACGGGAGGCAGCAG-3′) and 806R (5′-GGACTACHVGGGTWT

CTAAT). Using 2 μl DNA per sample, indexed amplicons were generated with Phusion Hot Start II mix (Thermo Fisher) following manufacturer's instructions with the PCR cycle: 95°C for 2 min, 30 amplification cycles of 95°C for 20 s, 55°C for 15 s, and 72°C for 5 min, followed by a final extension of 10 min at 72°C. Products were cleaned using AMPure XP beads (Agencourt), quantified by Qubit, pooled to equimolar concentrations and then quality-checked by Bioanalyzer. The samples were multiplexed and subjected to paired-end sequencing (2 × 250 bp) on the Illumina MiSeq platform.

The sequences were processed using Qiime 1.9.1 (Caporaso et al., 2010). The paired-end reads were joined with fastq (*ea-utils*: http://code.google.com/p/ea-utils), de-multiplexed and quality-filtered with default parameters. USEARCH6.1 (Edgar, 2010) was used to identify potentially chimeric sequences, which were removed. Sequences were assigned to operational taxonomic units (OTUs) with 97% sequence similarity using open-reference OTU picking (Rideout et al., 2014). OTUs represented by <100 reads across all samples and OTUs assigned to endosymbionts (genus *Wolbachia* and order Entomoplasmatales (likely the genus *Spiroplasma*)) were removed from the data sets (Supporting Information Data S1a), resulting in 4,119,013 16S reads (Supporting Information Data S1b). Prior to analysis, these samples were evenly subsampled to a depth of 102,000 reads (Supporting Information Data S1c), which yielded saturation of OTUs in all samples (Supporting Information Figure S1). These values were then normalized to relative abundances, and OTUs detected in ≥18 samples were included for clustering and correlation analyses.

2.4 | Metagenome analyses

The DNA was diluted in water to 20–40 ng/μl and fragmented with a S220 Sonicator (Covaris, Woburn, MA). Fragment ends were then repaired (End Repair Mix LC of Enzymatics, Beverly, MA), dA-tailed with Klenow (Enzymatics) and ligated to a sequencing adapter using T4 Ligase and 2× Rapid Ligation Buffer (Enzymatics) following manufacturer's protocol. AMPure XP beads (Agencourt, Brea, CA) were used after each step above for purification, according to manufacturer's instructions. The processed fragments were sized to 700–750 bp either by AMPure XP beads (Collections 1 and 2) or by BluePippin (Sage Science, Beverly, MA) (Collections 3 and 4), tagged by PCR amplification with indexed primers (Supporting Information Table S1), using Phusion Hot Start II (Thermo Fisher) and the following temperature cycle: 94°C for 2 min, 14 amplification cycles of 98°C for 10 s, 65°C for 30 s and 72°C for 30 s, followed by a final extension of 5 min at 72°C. Products were cleaned using AMPure XP beads (Agencourt), and then, the DNA concentration was determined by Qubit and quality by Bioanalyzer (Agilent, Santa Clara, CA). The multiplexed samples were subjected to paired-end sequencing (2 × 250 bp) on the Illumina HiSeq platform.

Sequencing reads were trimmed and split into paired and unpaired categories using Trimmomatic v 0.32 (Bolger, Lohse & Usadel, 2014) with the following sliding window cut-off specifications: LEADING:20, TRAILING:20, SLIDINGWINDOW:4:20, MINLEN:50. Vector sequence was removed by mapping to a library of common contaminating sources (UniVec: https://www.ncbi.nlm.nih.gov/tools/vecscreen/univec/) and the Illumina sequencing standard PhiX, using BOWTIE2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). The cleaned reads were paired using FLASH v1.2.11 (https://ccb.jhu.edu/software/FLASH/), and assembled into contigs using VELVET v1.2.09 (Zerbino & Birney, 2008), and merged using MEGAMERGE (https://github.com/LANL-Bioinformatics/MeGAMerge). The resultant assemblies were uploaded to METAGENEMARK (http://exon.gatech.edu/meta_gmhmmp.cgi) for gene prediction, creating representative protein sequence clusters of genes with 90% amino acid sequence similarity using the uclust algorithm. The protein sequences were annotated using IMG/MER (https://img.jgi.doe.gov), and this annotation identified KOs for a mean of 22% of contigs in each sample. Coverage from contigs with KOs assigned to bacteria was summed, and sequences from the endosymbionts *Wolbachia* and *Spiroplasma* (which do not reside in the gut lumen) were removed. The KOs represented by sequences in the negative controls were not removed from the experimental samples because the data indicated no systematic contamination (3.5% of KOs in experimental samples were present in two or more negative controls, with none in all four negative controls). Bacterial functions were annotated by KEGG Orthology Identifiers (KO numbers) and organized into a functional hierarchy based on that of KEGG, and the KOs present in ≥18 samples were normalized to relative abundance for clustering and correlation analyses.

2.5 | Gut transcriptome analyses

RNA-seq libraries were generated using the protocol of Bost et al. (2017) modified from Wang et al. (2011). Briefly, polyA+ RNA was purified from each sample comprising 2.5 μg total RNA with DynabeadsOligo (dT) 25 (Life Technologies) and fragmented to 200- to 250-bp fragments by incubation at 94°C for 5 min. Double-stranded cDNA was synthesized using Superscript II Reverse Transcriptase (Invitrogen) following manufacturer's instructions and cleaned with RNA Clean XP magnetic beads (Agencourt). Second strand synthesis was performed providing dUTP instead of dTTP to generate strand-specific libraries, A-T ligated with universal Y-Shaped adapters, purified and size-selected using Ampure XP beads (Agencourt). The uracil-containing second strand was digested using uracil DNA glycosylase (Enzymatics), and DNA was PCR amplified with 15 cycles using barcoded Illumina index primers (Supporting Information Table S1). The amplification products were purified using Ampure XP beads (Agencourt) and eluted in 12 μl Qiagen Buffer EB. The DNA concentrations determined by Qubit were 3.23-17.9 ng/μl. Library quality was assessed on Bioanalyzer and equimolar pools were subjected to 150-bp paired-end sequencing by Illumina HiSeq.

Raw sequence reads were first processed through the Galaxy platform (Afgan et al., 2016). Quality was checked by FastQC (Galaxy Version 0.65); then, 15 bases were trimmed from the 5′ end of the sequences using FASTQ Trimmer by column (Galaxy Version

1.0.0), followed by "Filter by quality" (Galaxy Version 1.0.0) with a Quality cut-off value of 30 over 90% of specified bases. The reads were paired using FASTQ joiner (Galaxy Version 1.0.0) and then unpaired using FASTQ splitter (Galaxy Version 1.0.0) to eliminate unmatched reads. De novo transcriptome assembly was performed on sequences pooled from all conspecific samples using the Trinity and Trinotate platforms (Grabherr et al., 2011) with default settings, generating one do novo assembly per *Drosophila* species. These transcripts were then annotated to reflect *D. melanogaster* gene orthology by BLASTX to the UniProt proteome for *D. melanogaster* (UP000000803). In total, annotations were obtained for 76% of *D. falleni* transcripts, and 82% of *D. neotestacea* transcripts, and 86% of *D. putrida* transcripts. These comprised >20,000 unigene contigs with N50 > 1,900 bp annotated to 10,672 Uniprot genes, with at least 5,087 genes annotated in Uniprot in each sample (Supporting Information Data S3a–c), including 4,781 genes found in all 3 species, corresponding to 3,917 Flybase genes and 1,088 KOs.

Each sequenced sample was aligned separately to its de novo transcriptome using the BOWTIE2 utility in the Trinity suite (https://github.com/trinityrnaseq/trinityrnaseq), with default settings, and expected gene counts were calculated using the RSEM utility, also in Trinity. Expected counts were summed within samples for each KO and gene identity. KOs were organized into a functional hierarchy based on KEGG, and the values were normalized to relative abundance using genes or KOs that were identified in all three *Drosophila* species.

2.6 | Statistical analyses

Statistical analyses were conducted in R(R Development Core Team, 2015). Relationships among samples were visualized with principal coordinates analysis (PCoA) for the relative abundance of OTUs and principal components analysis (PCA) of covariance matrices for relative abundance of KEGG orthologs (KOs, comprising molecular-level functions assigned to orthologous genes) and KEGG pathways (comprising interacting gene products in metabolism, gene information processing, cellular processes, etc.) for both the gut metagenomes and host transcriptomes. Procrustes analysis (Peres-Neto & Jackson, 2001) was used to test whether the metagenome content could be predicted from the taxonomic composition (16S OTUs) of the microbiome, and whether the host transcriptome could be predicted from the metagenome. Using the R package *vegan*, Procrustes rotations were generated with the "procrustes" function and the significance of these relationships tested with permutation tests using the "protest" function (Oksanen et al., 2016).

All pairwise correlations and tests of significance were performed using an adaptation, developed for this study, of Kendall's τ_b suggested by Pimentel (2009) and analysed with R package "*mazeinda*" developed for this study (Supplementary Text S1; Supplementary Script S1). This modification of Kendall's τ_b allows for its use on data sets with a substantial proportion of zero abundance levels. The choice of Kendall's rank correlation over Spearman's was motivated by the fact that Kendall's coefficient is not affected by the distance between ranks in the pairs. The Benjamini–Hochberg procedure was then applied, using the *p.adjust* function in R, to control the maximum permissible false discovery rate (FDR) for multiple comparisons. Specifically, we calculated the *q*-value for each feature, that is, the minimum FDR attained when calling all features ordered by ascending *p*-values up to and including the feature of interest. The *q*-value threshold can be interpreted as reflecting a conservative Bayes factor (a measure of evidence to assess effect size) for FDR interpretation, but the optimal threshold is not well-established (Efron, 2007). Following Efron (2007, 2008), we adopted the threshold *q*-value value of 0.2 in our analyses.

Each of the OTU and metagenome data sets was correlated with itself to identify the groups of correlated OTUs or genes in a given sample. The correlations were processed by the heatmap.2 in the package "gplots," and then, the R packages "factoextra" and "Nbclust" were used to determine the optimal number of clusters. The dendrogram from the heatmap was divided into the determined number of clusters using the R package "dendextend," and the original ranks of the members of the clusters were averaged to determine the ranks of each cluster across the samples. These mean ranks for each cluster were then correlated with the mean ranks of the clusters in the other data sets using Kendall's τ_b (as above). *Drosophila* transcripts that were significantly correlated with any of the metagenome clusters with a *q*-value <0.2 were analysed using the DAVID Bioinformatics Database (https://david.ncifcrf.gov/) (da Huang, Sherman & Lempicki, 2009) to identify GO term enrichments for Biological Processes. Enrichments with probability <0.05 were tallied for the results.

3 | RESULTS

3.1 | Composition of the *Drosophila* gut microbiome

The first of the three molecular analyses comprised sequencing of 16S rRNA gene fragments amplified from the dissected guts of the three *Drosophila* species. The amplicons were assigned to a total of 921 OTUs at 97% sequence identity, with 238–539 OTUs per sample. The samples were dominated by three major groups: Bacteroidetes, γ-Proteobacteria and Firmicutes of the order Lactobacillales (Figure 1a, Supporting Information Data S1c). Principal coordinates analysis (PCoA) of the data revealed that the samples clustered by collection (Figure 1b). Collection-3 and Collection-4 separated along the first PCo axis, strongly influenced by OTU 65, a γ-proteobacterium of the family Enterobacteriaceae, which is abundant in all samples of *D. falleni* and *D. putrida* in Collection-4, and by two OTUs (#5 and #241) assigned to the Sphingobacteriaceae (Bacteroidetes) in all samples of Collection-3. The second PCo axis separated the samples collected in 2013 (collections 1 and 2) from 2014 (collections 3 and 4), driven partly by family-level patterns within the Firmicutes, favouring Enterococcaceae in the 2013 samples and Lactobacillaceae in the 2014 samples, as well as differences in the dominant OTUs of Enterobacteriaceae and Sphingobacteriaceae (Figure 1b).

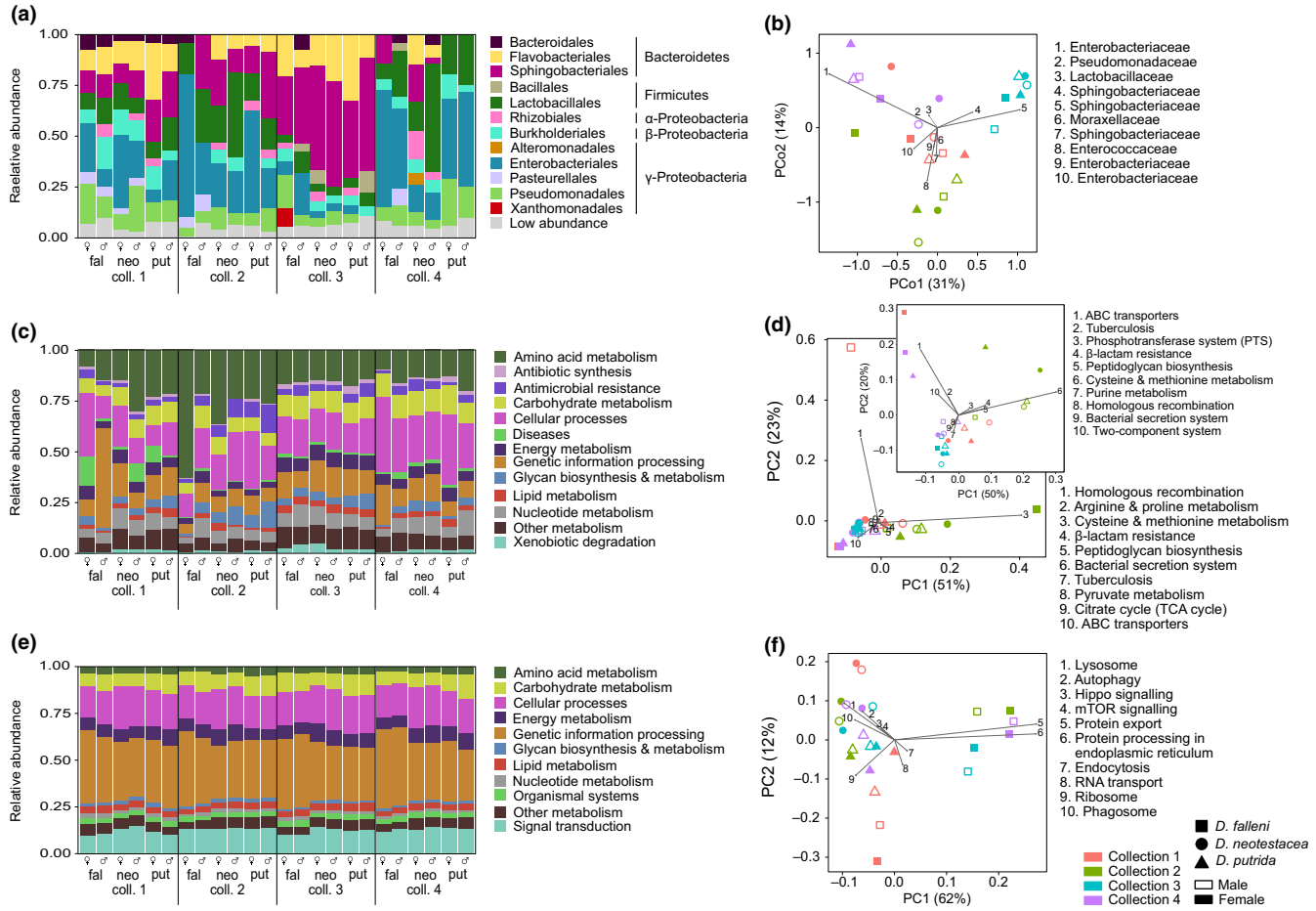

FIGURE 1 The microbiome and gut transcriptome. The taxonomic composition of the gut microbiome as (a) relative abundance of bacterial orders and (b) principal coordinates (PCo) analysis of Bray–Curtis distance between samples based on relative abundance of OTUs. The bacterial metagenome as (c) relative abundance of KEGG categories and (d) principal components (PC) analysis of KEGG pathways (inset: the same principal components analysis with the two extreme samples removed). The host transcriptome as (e) the relative abundance of KEGG categories and (f) principal components (PC) analysis of KEGG pathways. For (b), (d) and (e), the arrows represent the 10 OTUs (b) and KEGG pathways (d and f) that contribute most to among-sample variation. Key: fal, *D. falleni*; neo, *D. neotestacea*; put, *D. putrida*; coll., collection

3.2 | The gut metagenome

The metagenome analysis yielded 40–153 Mb per sample (Supporting Information Table S2), and the reads were assembled de novo. Overall 87% of reads mapped to the final assembly per sample, yielding 114,181–311,687 contigs per sample and representing 112,380–369,695 putative protein-coding genes (Supporting Information Table S2). In total, 18% of contigs and 19% of the protein-coding genes could be classified into KOs, with 3,494–7,558 KOs identified per sample (Supporting Information Table S2). Rarefaction curves of number of KOs against reads approached or reached saturation for all samples (Supporting Information Figure S1b). On average, 96% of KOs were assigned to the domain Bacteria (Supporting Information Figure S2). Most of the nonbacterial KOs were of animal (mostly host) origin and the average KO count for fungi was 0.3% of annotated genes.

The detailed analysis of the metagenome focused on the genes of bacterial origin, to obtain the greatest correspondence between the functional data and the 16S data sets. The KOs were assigned to 14 categories derived from the KEGG functional hierarchy, all of which were represented in every sample (Supporting Information Data S2a). Principal components analysis (PCA), using the summed coverage of KEGG pathways in our functional hierarchy, clustered the samples by collection, with among-sample variation dominated by Collection-2 on the first axis, and male *D. falleni* of Collection-1 on the second axis (Figure 1d). The "cysteine and methionine metabolism pathway" vector separating Collection-2 from the other collections is driven by the high variance in the abundance of K00558, a DNA cytosine methyltransferase that functions in bacterial restriction modification systems. Over half of the KOs in the hierarchy (1,832 KOs) were detected in at least 18 samples (Supporting Information Data S2b), and this set was used for the correlation analyses.

3.3 | Drosophila gut transcriptomes

The RNA-seq libraries yielded 21–54 million reads per sample (Supporting Information Data S3a–c), which were used for de novo

transcriptome assembly, with annotation to reflect *D. melanogaster* orthology. The genes were assigned to 11 categories derived from the KEGG hierarchy (Supporting Information Data S3d), with minor among-sample variation at this level (Figure 1e).

Greater among-sample variation was evident at the KEGG pathway level and PCA of the summed abundance for each pathway (Supporting Information Data S3d) revealed strong clustering by species (Figure 1f). The separation of *D. falleni* in collections 2, 3 and 4 from the other species on the first PC axis was driven particularly by the vectors for two KEGG pathways: "protein export" and "protein processing in endoplasmic reticulum." Both of these vectors were influenced strongly by large variance in the expression of K09490 (the ortholog of *D. melanogaster* Hsc70-3: CG4147) coding a heat-shock protein with elevated expression in the guts of ageing flies or those subjected to oxidative stress (Wang, Zeng, Ryoo & Jasper, 2014). The second PC axis separating *D. neotestacea*, *D. putrida* and Collection-1 of *D. falleni* was influenced strongly by three vectors with related functions in the internalization of particles ("phagosome") and intracellular degradation systems ("autophagy" and "lysosome"). The genes with large contributions to the ordination encode the Vo and V1 subunits of the V-type H^+-transporting ATPase (K02150 and K02155, respectively) which play a central role in acidification of the phagosome and lysosome, a cathepsin L which is a cysteine protease localized to the lysosome (K01365: CG6692), and the lysosomal-associated membrane protein LAMP1 (K06528: CG3305). These proteins are components of the phagosome pathway (KEGG map04145). We hypothesize that these results relate to predicted among-species variation in digestive function (see Section 4).

3.4 | Relationship between the microbiome and host traits

Our first approach to investigate how microbiome function maps onto taxonomy used Procrustes rotation between the ordinations for the relative abundance of OTUs and the two indices of metagenome function: KOs and KEGG pathways. Neither of these tests were statistically significant ($m^2 = 0.35$, $p = 0.12$ for KOs, $m^2 = 0.33$, $p = 0.16$ for pathways). We reasoned that this analysis, which treats each bacterial OTU and each functional term of the metagenome as independent, may be unrealistic. In particular, subsets of OTUs in the gut microbiome of these *Drosophila* species have been demonstrated to covary (Martinson et al., 2017), and the relative abundance of functions identified in the metagenome may also covary based on their inclusion in the same bacterial genomes. To identify the patterns of covariation, we clustered the OTUs and metagenome functions each into 5 clusters (Supporting Information Figure S3), each cluster representing a group of genes or OTUs that covary in abundance across the samples.

The correlation analysis of the five clusters of bacterial OTUs revealed multiple significant associations (Figure 2a (left)) that could be linked loosely to high-order taxonomy (Figure 2b, Supporting Information Figure S4a). In particular, the strongly negatively correlated clusters A and E include many OTUs of Bacteroidetes and Enterobacteriales, respectively; and cluster B (which includes representatives of Actinobacteria, Bacteroidetes and α- and γ-Proteobacteria) is also negatively correlated with cluster D with representatives in the Lactobacillales (Firmicutes) and Enterobacteriales and Pseudomonadales (γ-Proteobacteria) The five clusters of KEGG pathways in the metagenomes were not strictly defined by function (Figure 2c, Supporting Information Figure S4b). Nevertheless, the positively correlated clusters A and B contained over half of the metabolic functions and housekeeping functions, while the positively correlated clusters D and E contained most of the genes conferring functions relating to interactions with the environment, for example, transport functions (ABC transporters and phosphotransferase systems), motility (flagellar assembly) and biofilm formation. Cluster C, which was positively correlated with all other clusters, comprised a small proportion of all of the functional groups.

We then investigated the pattern of correlations between the OTU clusters and metagenome clusters (Figure 2a right). Two patterns are apparent: that the metabolic functions of the metagenome (strongly represented by metagenome clusters A and B) are positively associated with various bacteria, including members of the Bacteroidetes represented in OTU clusters A and B, but negatively associated with members of the Firmicutes and γ-Proteobacteria associated with OTU clusters D and E.

To investigate the correspondence between microbiome function and host function, we conducted Procrustes analysis. None of the tests, using both KOs and KEGG pathways for each of the metagenomes and transcriptomes, yielded statistically significant results (for metagenome KOs, $m^2 = 0.25$, $p = 0.43$ with transcriptome pathways, and $m^2 = 0.24$, $p = 0.50$ with transcriptome KOs; for metagenome pathways, $m^2 = 0.24$, $p = 0.48$ with transcriptome pathways and $m^2 = 0.26$, $p = 0.43$ with transcriptome KOs). Following the same reasoning as for the relationship between metagenomes and bacterial OTUs (see above), we then correlated the abundance of the five metagenome clusters (Supporting Information Data S2b) with the transcript abundance of expressed host genes (Supporting Information Table S3e) across the 24 samples. In total, the expression of 241 host genes (6.2% of the gene set) was significantly correlated with the relative abundance of KOs in at least one metagenome cluster after Benjamini–Hochberg correction for multiple tests with the threshold q-value set at 0.2 (and 109 genes at $q = 0.15$) (Supporting Information Data S3f). Most of the correlations involved metagenome clusters A and B (151 and 174 genes, respectively), and these included 31 of the 34 genes with the strongest correlations (Kendall's $|\tau| > 0.5$) (Supporting Information Data S3f). Many of these 34 genes function in metabolism and its regulation, for example, *chico*, *ns1* and *Tsf2* (insulin and TOR signalling), *gcn2* (control of amino acid metabolism), and in gut homeostasis and repair, for example, Wnt (*tws*) and BMP (*fu*) signalling (Buchon, Broderick & Lemaitre, 2013).

As a complementary analysis, we quantified the GO term enrichment for expressed host genes that were correlated with the mean ranks of the five metagenome clusters (Figure 3, Supporting

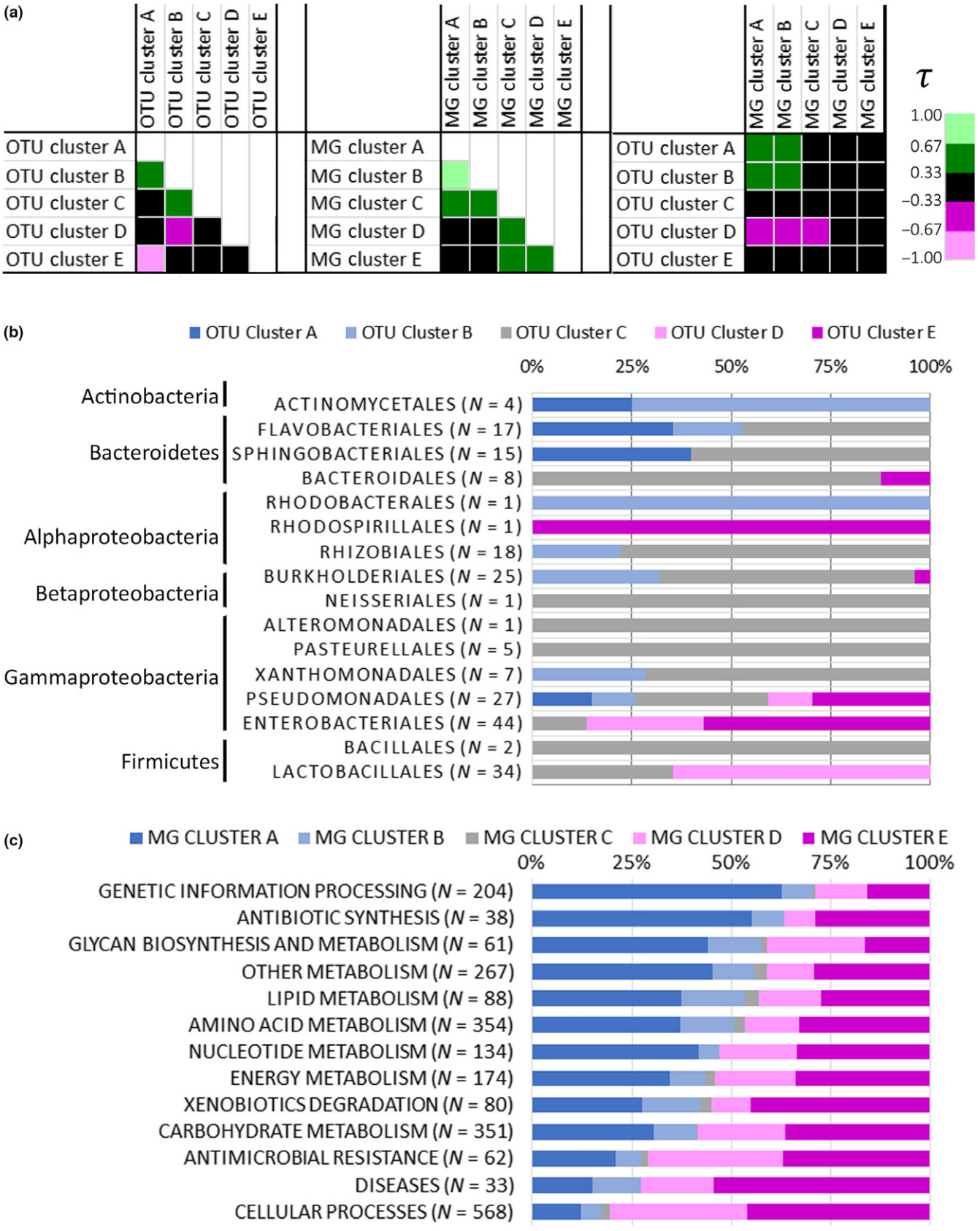

FIGURE 2 Clusters of bacterial OTUs and metagenome KOs. (a) Heatmap of correlations between mean ranks of each OTU cluster (left), each metagenome (MG) cluster of KOs (centre) and between OTUs and metagenome KOs. (b) Taxon representation in each OTU cluster (phyla and orders, *N* = number of OTUs). (c) KEGG category representation in each metagenome cluster (*N* = number of KOs)

Information Data S3g). As for the analysis at the host gene level, positive correlations with metagenome clusters A and B were strongly related to various host functions, including cell division and differentiation, proteolysis and stress response, and negatively correlated with energy/lipid metabolism. In addition, metagenome clusters C and D were correlated negatively with cell migration, morphogenesis and general signalling.

4 | DISCUSSION

The foundation of this study was the substantial among-sample variation in the gut microbiome of the field-collected *Drosophila*, by the criteria of both taxonomic composition of the bacteria and their functional traits (Figure 1a–d). Both of these indices varied principally with the period of collection, rather than with our molecular index of host function (the gut transcriptome), which varied by species (Figure 1e–f). Here, we discuss, in turn, the patterns of microbiome and host transcriptome variation and then the implications of correspondence between the functional traits of the microbiome and host.

The demonstration in this study that microbiome variation does not partition principally by host species suggests that environmental factors play an important role in shaping the gut microbiome. Although this result is fully consistent with the analysis of Martinson et al. (2017) for the taxonomic composition of the microbiome in individual flies of four mycophagous drosophilid species, including the three studied here, these results are surprising from the broad perspective of gut microbiome research. The default prediction is that the composition and functional traits of the gut microbiome would be shaped by the conditions and resources in the gut habitat and that the gut habitat would differ between host species, depending on diet and phylogeny (Brooks et al., 2016; Duar et al., 2017; Ley, Luzupone, Hamady, Knight & Gordon, 2008; Moeller et al., 2016). The latter expectation is borne out by the gut transcriptome data in this study, with gene expression differentiating by species on the first two PC axes (Figure 1f). Furthermore, the greatest distinction on both PC axes is between *D. falleni* and the other two species. This separation corresponds to both phylogeny (*D. falleni* is in subgroup quinaria, and *D. neotestacea*/*D. putrida* in subgroup testacea, and the two subgroups diverged an estimated 27 My ago (Izumitani et al., 2016)) and ecology, with *D. falleni* preferentially feeding on fresh portions of a mushroom and the other species on rotting parts of the mushroom substrate (Grimaldi, 1985). Interestingly, the transcriptional patterns that separate the species relate principally to the trafficking of macromolecules and particles between the gut epithelium and the external milieu. The gut transcriptome of *D. falleni* is enriched in transcripts with function related to protein processing and export, which is particularly associated with the production of digestive enzymes that function in the gut lumen (Simpson & Douglas, 2012). Digestive function is anticipated to be important to degrade the complex macromolecules, including proteins, glycoproteins and chitin, which dominate fresh mushroom ingested by *D. falleni*. The ingesta of the drosophilid species that preferentially feed on rotting mushroom substrate is expected to contain elevated concentrations of small organic molecules, derived from the microbial degradation of the mushroom substrate, together with many microbial cells. This diet requires less investment in the production of digestive enzymes and greater function in the internalization of microbial cells and cell fragments (phagocytosis and endocytosis) and the trafficking of this material to the lysosome for intracellular digestion. Our interpretation that the elevated trafficking of cargo from the extracellular environment to the lysosome in the gut of *D. putrida* and *D. neotestacea* has primarily a nutritional rather than defensive role comes from the parallel observation that expression of immune-related genes in these guts is generally low and does not vary significantly by species, although we recognize that nutrition and immunological functions are often overlapping (Broderick, 2015). This general interpretation that the nutritional ecology of *D. falleni* is associated with physiological adaptations for extracellular digestion in the gut lumen, while *D. neotestacea* and *D. putrida* are adapted for intracellular digestion can be tested empirically. From the perspective of the microbiome, however, the important point is that the genes which differentiate the gut function of the different host species are not the principal drivers of the relationship between host gut function and the taxonomic and functional composition of the microbiome. This result is fully consistent with other studies encompassing a greater number and phylogenetic diversity of *Drosophila* species that report either no or limited statistical relationship between host phylogeny and gut microbiota community composition (Brooks et al., 2016; Wong et al., 2013).

Despite the differences between the main determinants of the observed variation between the microbiome traits (collection) and host traits (species), this study did reveal statistically significant correspondence between the functional traits of the microbiome and both microbial taxonomy and host traits. However, these relationships are diffuse, that is, the patterns do not point to simple one-to-one-to-one relationships between any individual bacterial taxon, specific microbiome function and defined host trait. Interestingly, a similar conclusion was reached for an analysis of the gut microbiomes in natural populations of fruit-feeding *D. melanogaster* (Bost et al., 2017), despite major differences in diet and composition of the gut microbiota between *D. melanogaster* and the mycophagous drosophilids (Wong et al., 2013).

We hypothesize that the key biological reason for the diffuse relationship between the functional traits of the microbiome and host may be that the gut microbiome of drosophilids is an open system, subject to regular inputs of microorganisms with the food and loss by shedding in the faeces (Blum, Fischer, Miles & Handelsman, 2013; Fink, Staubach, Kuenzel, Baines & Roeder, 2013; Inamine et al., 2018; Wong et al., 2015). This trait is predicted to limit the opportunity for coevolution between microbial and host functional traits. However, the results may also reflect, in part, a failure of our indices to capture the functional traits accurately. There are caveats for all of our indices. The procedure to amplify the DNA may have introduced bias in the taxonomic composition (16S) and functional

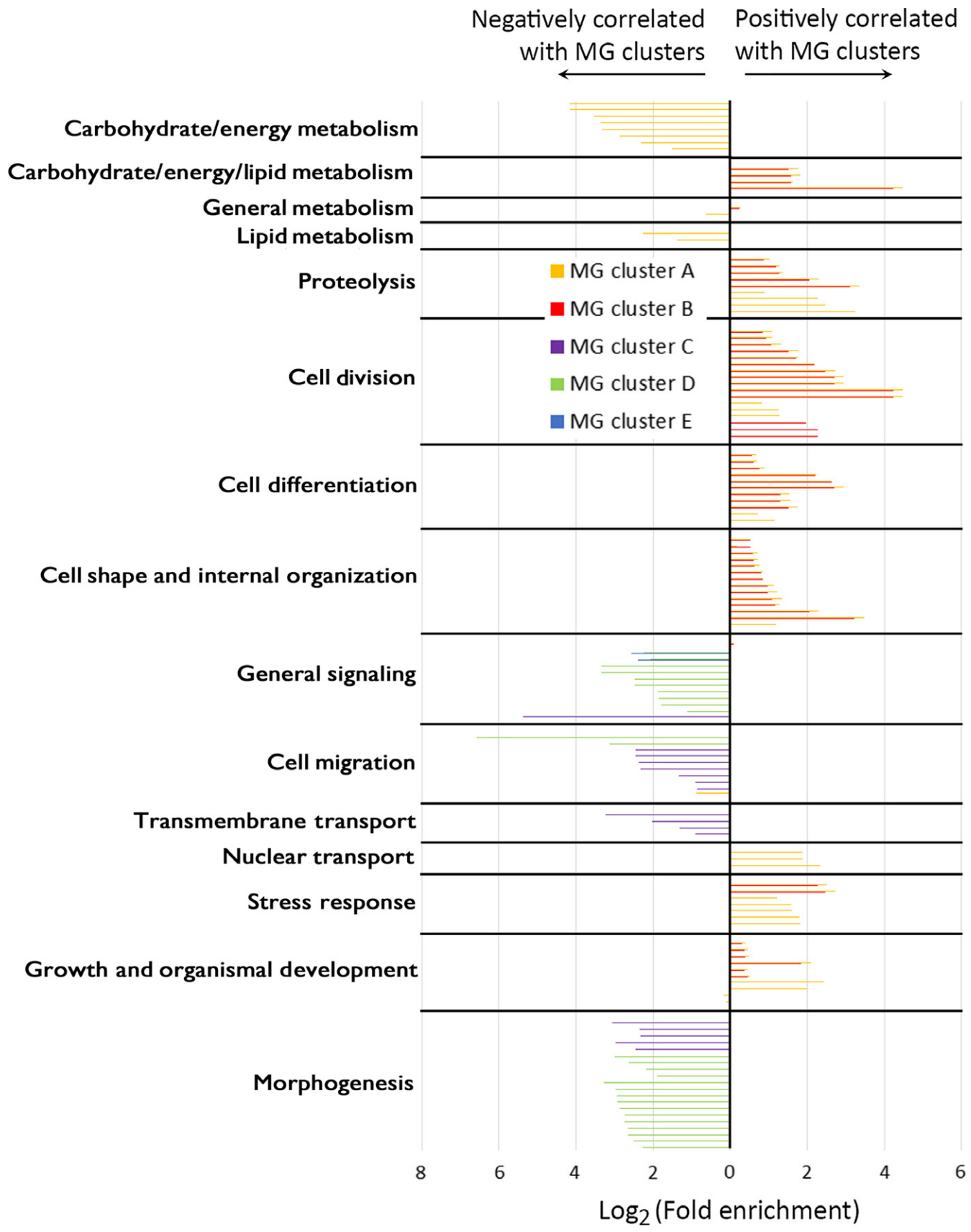

FIGURE 3 Gene ontology terms enriched in *Drosophila* gut transcripts that are correlated with metagenome clusters

traits (metagenome) of the microbiome. Sequence variation in the 16S rRNA gene is insufficient to represent functionally important strain variation caused by horizontal acquisition and deletion of genes in closely related bacteria (Gordienko, Kazanov & Gelfand, 2013; Jaspers & Overmann, 2004; Poretsky, Rodriguez, Luo, Tsementzi & Konstantinidis, 2014; Roux, Enault, Bronner & Debroas, 2011). The metagenome quantifies the genetic capacity for function, and some functional traits may remain undetected because of inadequate annotation (Quince, Walker, Simpson, Loman & Segata, 2017) or be inflated artificially by genes that are not expressed or are localized in metabolically inactive or recently killed microbial cells (Aguiar-Pulido et al., 2016; Franzosa et al., 2014). Finally, the host gene expression profile is not a perfectly accurate representation of the conditions and resources encountered by the microorganisms in the gut because some genes are expressed in regions of the gut that are poorly colonized by microorganisms or are regulated post-translationally. For example, the dual oxidase Duox, which generates microbicidal hydrogen peroxide, hypochlorite and reactive oxygen species in the gut of *D. melanogaster* (Ha, Oh, Bae & Lee, 2005; Lee et al., 2015).

Despite the potential limitations to the indices used, statistically significant correlations between the functional traits of the microbiome and host were obtained. These relationships encompassed multiple functional categories, with a repeated pattern of metabolism-related microbiome functions that are strongly represented in metagenome clusters A and B, and positively correlated with cell division and differentiation functions in the host transcriptome. This robust relationship may be driven by variation in the quality and abundance of food, with high food availability promoting metabolic function in microorganisms and host gut epithelial cell division, differentiation and turnover (Buchon et al., 2013; Lemaitre & Miguel-Aliaga, 2013); that is, the microbiome and host may be responding independently to diet (and possibly other environmental variables). Additionally or alternatively, metabolic activity of both the host and the microbiome may be enhanced by high rates of exchange of metabolites between the partners. For the host, microbial metabolites may contribute directly to nutrition and also act as signalling molecules that promote gut cell proliferation and differentiation (Rath & Dorrestein, 2012; Sharon et al., 2014). For example, the genes of mycophagous drosophilids with expression levels that correlated with microbial metabolic function include key genes in the insulin/TOR pathway, which in *D. melanogaster* is promoted by a specific microbial fermentation product, acetic acid (Shin et al., 2011). Our data further indicate that many of the bacterial taxa that contribute to these interactions are members of the Bacteroidetes (Figure 2), including Sphingobacteriaceae, Weeksellaceae and Flavobacteriaceae, and not the γ-Proteobacteria and Firmicutes which tend to dominate the microbial communities in many individual flies (Martinson et al., 2017) and in some collections (this study). Further research is required to establish the processes shaping these correlations. However, as with mammalian microbiomes (Johnson, Heaver, Walters & Ley, 2017; Spor, Koren & Ley, 2011), the relationships between taxonomy, function and host traits in the gut microbiome of mycophagous drosophilids include considerable within-phylum variation. It appears that broad generalizations about the functional traits of high-level taxonomic groupings cannot be made for the gut microbiome of many animals.

To conclude, this study has demonstrated that correspondence between functional traits of the gut microbiome and host can be detected in natural animal populations, without imposing the tight experimental controls over environment, host genotype and microbiome composition that are standard protocols in most laboratory studies. Furthermore, the functional traits contributing to the microbiome–host interaction in this system are dominated by metabolism and growth-related functions. Our analysis was conducted entirely at the molecular level, and a key priority for future research is to establish how the nexus of host–microbiome interactions influences the performance and fitness of the host under field conditions. Drosophilid flies are well suited to such analysis.

ACKNOWLEDGEMENTS

We thank Leanne Donahue and Marita Wilson for technical assistance and NSF grant BIO 1241099 for financial support.

DATA ACCESSIBILITY

16S amplicon sequences, metagenome sequences and transcriptome sequences are deposited in NCBI SRA under Accession No PRJNA428838. The R-script for Kendall's correlations is published on figshare (https://doi.org/10.6084/m9.figshare.6067256.v1). The metadata associated with the sequencing data are also available at figshare (https://doi.org/10.6084/m9.figshare.6013988.v1).

AUTHOR CONTRIBUTIONS

V.G.M., S.F. and A.E.D. designed the research; V.G.M. conducted the field work; A.B., V.G.M., S.F. and K.L.A. conducted the molecular biology and analysis; A.A. and M.T.W. devised and conducted the modified Kendall's test; A.B., S.F., K.L.A. and A.E.D. wrote the manuscript; all authors participated in discussions and revised the manuscript.

ORCID


Angela E. Douglas http://orcid.org/0000-0001-5212-6826

REFERENCES

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., … Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*, *44*, W3–W10. https://doi.org/10.1093/nar/gkw343

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., & Narasimhan, G. (2016). Metagenomics, metatranscriptomics, and

metabolomics approaches for microbiome analysis. *Evolutionary Bioinformatics Online*, *12*, 5–16.

Blum, J. E., Fischer, C. N., Miles, J., & Handelsman, J. (2013). Frequent replenishment sustains the beneficial microbiome of *Drosophila melanogaster*. *mBio*, *4*, e00860-00813.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bost, A., Franzenburg, S., Adair, K. L., Martinson, V. G., Loeb, G., & Douglas, A. E. (2017). How gut transcriptional function of *Drosophila melanogaster* varies with the presence and composition of the gut microbiota. *Molecular Ecology*, *27*, 1848–1859.

Broderick, N. A. (2015). A common origin for immunity and digestion. *Frontiers in Immunology*, *6*, 72.

Broderick, N. A., & Lemaitre, B. (2012). Gut-associated microbes of *Drosophila melanogaster*. *Gut Microbes*, *3*, 307–321. https://doi.org/10.4161/gmic.19896

Brooks, A. W., Kohl, K. D., Brucker, R. M., van Opstal, E. J., & Bordenstein, S. R. (2016). Phylosymbiosis: Relationships and functional effects of microbial communities across host evolutionary history. *PLoS Biology*, *14*, e2000225.

Buchon, N., Broderick, N. A., & Lemaitre, B. (2013). Gut homeostasis in a microbial world: Insights from *Drosophila melanogaster*. *Nature Reviews Microbiology*, *11*, 615–626. https://doi.org/10.1038/nrmicro3074

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*, 335–336. https://doi.org/10.1038/nmeth.f.303

da Huang, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*, 44–57. https://doi.org/10.1038/nprot.2008.211

Doolittle, W. F., & Booth, A. (2017). It's the song, not the singer: An exploration of holobiosis and evolutionary theory. *Biology and Philosophy*, *32*, 5–24. https://doi.org/10.1007/s10539-016-9542-2

Duar, R. M., Lin, X. B., Zheng, J., Martino, M. E., Grenier, T., Perez-Munoz, M. E., … Walter, J. (2017). Lifestyles in transition: Evolution and natural history of the genus *Lactobacillus*. *FEMS Microbiology Reviews*, *41*, S27–S28. https://doi.org/10.1093/femsre/fux030

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Efron, B. (2007). Size, power and false discovery rates. *Annals of Statistics*, *35*, 1351–1377. https://doi.org/10.1214/009053606000001460

Efron, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statistical Science*, *23*, 1–22. https://doi.org/10.1214/07-STS236

Fink, C., Staubach, F., Kuenzel, S., Baines, J. F., & Roeder, T. (2013). Non-invasive analysis of microbiome dynamics in the fruit fly *Drosophila melanogaster*. *Applied and Environmental Microbiology*, *79*, 6984–6988. https://doi.org/10.1128/AEM.01903-13

Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., Giannoukos, G., Boylan, M. R., Ciulla, D., Gevers, D., … Huttenhower, C. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, E2329–E2338. https://doi.org/10.1073/pnas.1319284111

Gordienko, E. N., Kazanov, M. D., & Gelfand, M. S. (2013). Evolution of pan-genomes of *Escherichia coli*, *Shigella* spp., and *Salmonella enterica*. *Journal of Bacteriology*, *195*, 2786–2792. https://doi.org/10.1128/JB.02285-12

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*, 644–652. https://doi.org/10.1038/nbt.1883

Grimaldi, D. (1985). Niche separation and competitive coexistence in mycophagous *Drosophila* (Diptera: Drosophilidae). *Proceedings of the Entomological Society of Washington*, *87*, 498–511.

Ha, E. M., Oh, C. T., Bae, Y. S., & Lee, W. J. (2005). A direct role for dual oxidase in *Drosophila* gut immunity. *Science*, *310*, 847–850. https://doi.org/10.1126/science.1117311

Hamilton, P. T., Peng, F., Boulanger, M. J., & Perlman, S. J. (2016). A ribosome-inactivating protein in a *Drosophila* defensive symbiont. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 350–355. https://doi.org/10.1073/pnas.1518648113

Inamine, H., Ellner, S. P., Newell, P. D., Luo, Y., Buchon, N., & Douglas, A. E. (2018). Spatio-temporally heterogeneous population dynamics of gut bacteria inferred from fecal time-series data. *mBio*, *9*, e01453-17. https://doi.org/10.1128/mBio.01453-17

Izumitani, H. F., Kusaka, Y., Koshikawa, S., Toda, M. J., & Katoh, T. (2016). Phylogeography of the Subgenus Drosophila (Diptera: Drosophilidae): Evolutionary history of faunal divergence between the Old and the New Worlds. *PLoS ONE*, *11*, e0160051. https://doi.org/10.1371/journal.pone.0160051

Jaenike, J., & James, A. C. (1991). Aggregation and the coexistence of mycophagous *Drosophila*. *Journal of Animal Ecology*, *60*, 913–928. https://doi.org/10.2307/5421

Jaspers, E., & Overmann, J. (2004). Ecological significance of microdiversity: Identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies. *Applied and Environmental Microbiology*, *70*, 4831–4839. https://doi.org/10.1128/AEM.70.8.4831-4839.2004

Johnson, E. L., Heaver, S. L., Walters, W. A., & Ley, R. E. (2017). Microbiome and metabolic disease: Revisiting the bacterial phylum Bacteroidetes. *Journal of Molecular Medicine*, *95*, 1–8. https://doi.org/10.1007/s00109-016-1492-2

Keller, A. (2007). Drosophila melanogaster's history as a human commensal. *Current Biology*, *17*, R77–R81. https://doi.org/10.1016/j.cub.2006.12.031

Kohl, K. D., Weiss, R. B., Cox, J., Dale, C., & Dearing, M. D. (2014). Gut microbes of mammalian herbivores facilitate intake of plant toxins. *Ecology Letters*, *17*, 1238–1246. https://doi.org/10.1111/ele.12329

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*, *79*, 5112–5120. https://doi.org/10.1128/AEM.01043-13

Lee, K. A., Kim, B., Bhin, J., Kim, D. H., You, H., Kim, E. K., … Lee, W. J. (2015). Bacterial uracil modulates *Drosophila* DUOX-dependent gut immunity via Hedgehog-induced signaling endosomes. *Cell Host & Microbe*, *17*, 191–204. https://doi.org/10.1016/j.chom.2014.12.012

Lemaitre, B., & Miguel-Aliaga, I. (2013). The digestive tract of *Drosophila melanogaster*. *Annual Review of Genetics*, *47*, 377–404. https://doi.org/10.1146/annurev-genet-111212-133343

Ley, R. E., Luzupone, C. A., Hamady, M., Knight, R., & Gordon, J. I. (2008). Worlds within worlds: Evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology*, *6*, 776–788. https://doi.org/10.1038/nrmicro1978

Louca, S., Parfrey, L. W., & Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. *Science*, *353*, 1272–1277. https://doi.org/10.1126/science.aaf4507

Macpherson, A. J., & McCoy, K. D. (2015). Standardised animal models of host microbial mutualism. *Mucosal Immunology*, *8*, 476–486. https://doi.org/10.1038/mi.2014.113

Martinson, V. G., Douglas, A. E., & Jaenike, J. (2017). Community structure of the gut microbiota in sympatric species of wild *Drosophila*. *Ecology Letters*, *20*, 629–639. https://doi.org/10.1111/ele.12761

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C., Carey, H. V., Domazet-Loso, T., Douglas, A. E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S. F., … Wernegreen, J. J. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 3229–3236. https://doi.org/10.1073/pnas.1218525110

Moeller, A. H., Caro-Quintero, A., Mjungu, D., Georgiev, A. V., Lonsdorf, E. V., Muller, M. N., … Ochman, H. (2016). Cospeciation of gut microbiota with hominids. *Science*, *353*, 380–382. https://doi.org/10.1126/science.aaf3951

Moya, A., & Ferrer, M. (2016). Functional redundancy-induced stability of gut microbiota subjected to disturbance. *Trends in Microbiology*, *24*, 402–413. https://doi.org/10.1016/j.tim.2016.02.002

Newell, P. D., & Douglas, A. E. (2014). Among-species interactions determine the impact of gut microbiota on nutrient allocation in *Drosophila melanogaster*. *Applied and Environmental Microbiology*, *80*, 788–796. https://doi.org/10.1128/AEM.02742-13

Nguyen, T. L., Vieira-Silva, S., Liston, A., & Raes, J. (2015). How informative is the mouse for human gut microbiota research? *Disease Model Mechanisms*, *8*, 1–16. https://doi.org/10.1242/dmm.017400

Oksanen, J., Guillaume Blanchet, F., Kidt, R., Legendre, P., Minchin, P. R., & O'Hara, R. B. (2016). *vegan: community ecology package*. R package version 2.3-4.

Peres-Neto, P. R., & Jackson, D. A. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*, *129*, 169–178. https://doi.org/10.1007/s004420100720

Pimentel, R. S. (2009). *Kendall's Tau and Spearman's Rho for zero-inflated data*. Western Michigan University Dissertation 721. Retrieved from http://scholarworks.wmich.edu/dissertations/721.

Poretsky, R., Rodriguez, R. L., Luo, C., Tsementzi, D., & Konstantinidis, K. T. (2014). Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS ONE*, *9*, e93827. https://doi.org/10.1371/journal.pone.0093827

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, *35*, 833–844. https://doi.org/10.1038/nbt.3935

R Development Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rath, C. M., & Dorrestein, P. C. (2012). The bacterial chemical repertoire mediates metabolic exchange within gut microbiomes. *Current Opinion in Microbiology*, *15*, 147–154. https://doi.org/10.1016/j.mib.2011.12.009

Rideout, J. R., He, Y., Navas-Molina, J. A., Walters, W. A., Ursell, L. K., Gibbons, S. M., Chase, J., McDonald, D., Gonzalez, A., Robbins-Pianka, A., … Caporaso, J. G. (2014). Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *Peer Journal*, *2*, e545. https://doi.org/10.7717/peerj.545

Roux, S., Enault, F., Bronner, G., & Debroas, D. (2011). Comparison of 16S rRNA and protein-coding genes as molecular markers for assessing microbial diversity (Bacteria and Archaea) in ecosystems. *FEMS Microbiology Ecology*, *78*, 617–628. https://doi.org/10.1111/j.1574-6941.2011.01190.x

Sampson, T. R., Debelius, J. W., Thron, T., Janssen, S., Shastri, G. G., Ilhan, Z. E., Challis, C., Schretter, C. E., Rocha, S., Gradinaru, V., … Mazmanian, S. K. (2016). Gut microbiota regulate motor deficits and neuroinflammation in a model of Parkinson's disease. *Cell*, *167*, 1469–1480 e1412.

Sharon, G., Garg, N., Debelius, J., Knight, R., Dorrestein, P. C., & Mazmanian, S. K. (2014). Specialized metabolites from the microbiome in health and disease. *Cell Metabolism*, *20*, 719–730. https://doi.org/10.1016/j.cmet.2014.10.016

Shin, S. C., Kim, S. H., You, H., Kim, B., Kim, A. C., Lee, K. A., … Lee, W. J. (2011). Drosophila microbiome modulates host developmental and metabolic homeostasis via insulin signaling. *Science*, *334*, 670–674. https://doi.org/10.1126/science.1212782

Simpson, S. J., & Douglas, A. E. (2012). *The insects: Structure and function*, 5th edn. Cambridge, UK: Cambridge University Press).

Spor, A., Koren, O., & Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. *Nature Reviews Microbiology*, *9*, 279–290. https://doi.org/10.1038/nrmicro2540

Surana, N. K., & Kasper, D. L. (2017). Moving beyond microbiome-wide associations to causal microbe identification. *Nature*, *552*, 244–247.

Wang, L., Si, Y., Dedow, L. K., Shao, Y., Liu, P., & Brutnell, T. P. (2011). A low-cost library construction protocol and data analysis pipeline for Illumina-based strand-specific multiplex RNA-seq. *PLoS ONE*, *6*, e26426. https://doi.org/10.1371/journal.pone.0026426

Wang, L., Zeng, X., Ryoo, H. D., & Jasper, H. (2014). Integration of UPRER and oxidative stress signaling in the control of intestinal stem cell proliferation. *PLoS Genetics*, *10*, e1004568. https://doi.org/10.1371/journal.pgen.1004568

Werner, T., & Jaenike, J. (2017). *Drosophilids of the Midwest and Northeast*. Open Access Books. 1. Retrieved from http://digitalcommons.mtu.edu/oabooks/1.

Wong, A. C., Chaston, J. M., & Douglas, A. E. (2013). The inconstant gut microbiota of *Drosophila* species revealed by 16S rRNA gene analysis. *The ISME Journal*, *7*, 1922–1932. https://doi.org/10.1038/ismej.2013.86

Wong, A. C., Luo, Y., Jing, X., Franzenburg, S., Bost, A., & Douglas, A. E. (2015). The host as the driver of the microbiota in the gut and external environment of *Drosophila melanogaster*. *Applied and Environmental Microbiology*, *81*, 6232–6240. https://doi.org/10.1128/AEM.01442-15

Zerbino, D. R., & Birney, E. (2008). VELVET: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*, 821–829. https://doi.org/10.1101/gr.074492.107

Zhang, F., Berg, M., Dierking, K., Felix, M. A., Shapira, M., Samuel, B. S., & Schulenburg, H. (2017). *Caenorhabditis elegans* as a model for microbiome research. *Frontiers in Microbiology*, *8*, 485.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Bost A, Martinson VG, Franzenburg S, et al. Functional variation in the gut microbiome of wild *Drosophila* populations. *Mol Ecol*. 2018;27:2834–2845. https://doi.org/10.1111/mec.14728