RESEARCH ARTICLE

Metagenome-Wide Association of Microbial Determinants of Host Phenotype in *Drosophila melanogaster*

John M. Chaston,[a]* Peter D. Newell,[a]* Angela E. Douglas[a,b]

Department of Entomology[a] and Department of Molecular Biology and Genetics,[b] Cornell University, Ithaca, New York, USA

* Present address: John M. Chaston, Department of Plant & Wildlife Sciences, Genetics and Biotechnology, Brigham Young University, Provo, Utah, USA; Peter D. Newell, Department of Biological Sciences, Oswego State University of New York, Oswego, New York, USA.

J.M.C. and P.D.N. contributed equally to this article.

ABSTRACT **Animal-associated bacteria (microbiota) affect host behaviors and physiological traits. To identify bacterial genetic determinants of microbiota-responsive host traits, we employed a metagenome-wide association (MGWA) approach in two steps. First, we measured two microbiota-responsive host traits, development time and triglyceride (TAG) content, in *Drosophila melanogaster* flies monoassociated with each of 41 bacterial strains. The effects of monoassociation on host traits were not confined to particular taxonomic groups. Second, we clustered protein-coding sequences of the bacteria by sequence similarity *de novo* and statistically associated the magnitude of the host trait with the bacterial gene contents. The animals had been monoassociated with genome-sequenced bacteria, so the metagenome content was unambiguous. This analysis showed significant effects of pyrroloquinoline quinone biosynthesis genes on development time, confirming the results of a published transposon mutagenesis screen, thereby validating the MGWA; it also identified multiple genes predicted to affect host TAG content, including extracellular glucose oxidation pathway components. To test the validity of the statistical associations, we expressed candidate genes in a strain that lacks them. Monoassociation with bacteria that ectopically expressed a predicted oxidoreductase or gluconate dehydrogenase conferred reduced *Drosophila* TAG contents relative to the TAG contents in empty vector controls. Consistent with the prediction that glucose oxidation pathway gene expression increased bacterial glucose utilization, the glucose content of the host diet was reduced when flies were exposed to these strains. Our findings indicate that microbiota affect host nutritional status through modulation of nutrient acquisition. Together, these findings demonstrate the utility of MGWA for identifying bacterial determinants of host traits and provide mechanistic insight into how gut microbiota modulate the nutritional status of a model host.**

IMPORTANCE **To understand how certain gut bacteria promote the health of their animal hosts, we need to identify the bacterial genes that drive these beneficial relationships. This task is challenging because the bacterial communities can vary widely among different host individuals. To overcome this difficulty, we quantified how well each of 41 bacterial species protected *Drosophila* fruit flies from high fat content. The genomes of the chosen bacterial strains were previously sequenced, so we could statistically associate specific bacterial genes with bacterially mediated reduction in host fat content. Bacterial genes that promote glucose utilization were strongly represented in the association, and introducing these genes into the gut bacteria was sufficient to lower the animal's fat content. Our method is applicable to the study of many other host-microbe interactions as a way to uncover microbial genes important for host health.**

Received 17 July 2014 **Accepted** 8 September 2014 **Published** 30 September 2014

Citation Chaston JM, Newell PD, Douglas AE. 2014. Metagenome-wide association of microbial determinants of host phenotype in *Drosophila melanogaster*. mBio 5(5): e01631-14. doi:10.1128/mBio.01631-14.

Editor Margaret J. McFall-Ngai, University of Wisconsin

Address correspondence to John M. Chaston, john_chaston@byu.edu.

Microbial interactions with animal hosts are diverse and ubiquitous. Host-associated microbial communities (microbiota) affect multiple physiological and behavioral traits of animals, as demonstrated by the impact of eliminating the microbiota on metabolic, nutritional, vascular, hepatic, respiratory, immunological, neurobiological, and endocrine function, as well as feeding patterns and social interactions (1–10). The microbial genes that underlie the impact of the microbiota on host traits can be difficult to elucidate because microbiota are taxonomically and functionally diverse. Systems-level studies are proving to be useful to dissect the genetic and molecular bases of these complex host-microbiota interactions (11–17).

Metagenome-wide association (MGWA) approaches, which associate microbiota genetic content with host trait magnitude, are an emerging systems-level approach for predicting genetic mechanisms of gut microbiota function within a host (17–20). The associated genetic contents can range from whole-genome identity (17) to the presence or absence of specific genes (18, 19) or single nucleotide polymorphisms. When microbiota metagenomes are used as the basis for genetic content, conclusions are

TABLE 1 Bacterial strains

Strain name	Abbreviation[a]	GenBank accession no.	Preferred medium[b]	Oxygen conditions[b]
Acetobacter aceti NBRC 14818	aace	BABW00000000	mMRS	Oxic
Acetobacter malorum DmCS_005	amac	JOJU00000000[c]	mMRS	Oxic
Acetobacter pasteurianus 3P3	apa3	CADQ00000000[b,d]	mMRS	Oxic
Acetobacter pomorum DmCS_004	apoc	JOKL00000000[c]	mMRS	Oxic
Acetobacter pasteurianus NBRC 101655	apan	BACF00000000	mMRS	Oxic
Acetobacter tropicalis DmCS_006	atrc	JOKM00000000[c]	mMRS	Oxic
Acetobacter tropicalis NBRC 101654	atrn	BABS00000000	mMRS	Oxic
Bacillus subtilis subsp. subtilis strain168	bsub	NC_000964.3	LB	Oxic
Dysgonomonas mossii DSM 22836	dmos	ADLW00000000	BHI	Microoxic
Escherichia coli strain K-12 substrain MG1655	ecok	NC_000913.3	LB	Oxic
Enterococcus faecalis V583	efav	NC_004668.1	BHI	Oxic
Enterococcus faecalis OG1RF	efog	NC_017316.1	BHI	Oxic
Enterobacter hormaechei ATCC 49162	ehor	AFHR00000000	LB	Oxic
Gluconacetobacter europaeus LMG 18494	geur	CADR00000000	Potato	Oxic
Gluconacetobacter hansenii ATCC 23769	ghan	ADTV01000000	Potato	Oxic
Gluconacetobacter oboediens 174Bp2	gobo	CADT00000000	Potato	Oxic
Gluconacetobacter xylinus NBRC 3288	gxyl	NC_016037.1	Potato	Oxic
Gluconobacter frateurii NBRC 101659	gfra	BADZ00000000	Potato	Oxic
Lactobacillus animalis KCTC 3501	lani	AEOF00000000	mMRS	Microoxic
Lactobacillus brevis DmCS_003	lbrc	JOKA00000000[c]	mMRS	Microoxic
Lactobacillus brevis subsp. gravesensis ATCC 27305	lbga	ACGG00000000	mMRS	Microoxic
Lactobacillus buchneri NRRLB-30929	lbuc	NC_015428.1	mMRS	Microoxic
Lactobacillus casei W56	lcas	NC_018641.1	mMRS	Microoxic
Lactobacillus fermentum ATCC 14931	lfer	ACGI00000000	mMRS	Microoxic
Lactobacillus fructivorans DmCS_002	lfrc	JOJZ00000000[c]	mMRS	Microoxic
Lactobacillus fructivorans KCTC 3543	lfrk	AEQY00000000	mMRS	Microoxic
Lactobacillus gasseri ATCC 33323	lgas	NC_008530.1	mMRS	Microoxic
Lactococcus lactis BPL1	llac	JRFX00000000	mMRS	Microoxic
Lactobacillus malefermentans KCTC 3548	lmle	BACN00000000	mMRS	Microoxic
Lactobacillus mali KCTC 3596 = DSM 20444	lmli	BACP00000000	mMRS	Microoxic
Lactobacillus plantarum DmCS_001	lplc	JOJT00000000[c]	mMRS	Microoxic
Lactobacillus plantarum WCFS1	lplw	NC_004567.2	mMRS	Microoxic
Lactobacillus rhamnosus GG	lrha	NC_013198.1	mMRS	Microoxic
Lactobacillus versmoldensis KCTC 3814	lver	BACR00000000	mMRS	Microoxic
Leuconostoc fallax KCTC 3537	lfal	AEIZ00000000	mMRS	Microoxic
Providencia burhodogranariea DSM 19968	pbur	AKKL00000000	LB	Oxic
Pseudomonas putida F1	pput	NC_009512.1	LB	Oxic
Serratia marcescens ATCC 13880	smar	N. Perna, personal communication	LB	Oxic
Streptococcus mutans UA 159	smut	NC_004350.2	BHI	Microoxic
Streptococcus parasanguinis FW 213	spar	NC_017905.1	BHI	Microoxic
Sphingomonas wittichii RW1	swit	NC_009511.1	BHI	Oxic

[a] The abbreviations are used in Fig. 1 and in the supplemental material.
[b] Details are provided in Materials and Methods.
[c] Newell and Douglas et al., submitted for publication.
[d] Annotations were not available with whole-genome sequence data, so annotation was performed in RAST.

constrained by multiple aspects of metagenomic sequence analysis, e.g., nonsaturating sequence depth, incomplete resolution of low-abundance bacterial taxa, or inaccuracies in taxon binning. These limitations can be ameliorated by measuring the effects of clonal, genome-sequenced bacterial populations on host traits because the taxonomic and genetic composition of the microbiota is unambiguous and the specific microbial genes that affect the host can be identified with high confidence (18, 19). Despite the great potential for conducting MGWA on hosts using monoassociations, this approach has not yet been implemented in any animal host.

The fruit fly *Drosophila melanogaster* and its gut microbiota are an ideal system for MGWA with gnotobiotic monoassociation. The gut microbiota of *Drosophila* is of low taxonomic diversity, typically dominated by 2 to 5 species at a time (usually *Acetobacteraceae* or *Lactobacillus* species), each of which are readily cultured independently from the host (21–25). *Drosophila* is easily rendered axenic and readily reassociates with single species of bacteria (26–28). Also, *Drosophila* responds differently to distinct bacterial taxa for several traits, including the regulation of larval development time and host nutrient content (26–28). Finally, the *Drosophila* gut microbiota shares several characteristics with mammalian gut microbiota, suggesting that findings here may be generalizable to other animal phyla. For example, neither group displays a taxonomic core microbiota (22, 29), and the microbiota in both groups are inconstant within and across generations (22, 30).

In this study, *D. melanogaster* was monoassociated with 41 genome-sequenced bacteria (Table 1) to identify bacterial determinants of host traits by MGWA. We focused on microbial effects

on host triglyceride (TAG) abundance, following evidence that both the unmanipulated microbiota and certain single bacteria can protect *Drosophila* from elevated TAG levels (27, 28, 31). Understanding how animals affect the allocation of nutrients, including TAG, is important since microbial effects on obesity are shared across animal phyla (e.g., see reference 32). We confirmed the prediction that candidate genes identified by MGWA affect TAG content by ectopically expressing these genes in a heterologous gut microbe *in vivo* and examining their mechanism of action. Our findings demonstrate the utility of MGWA in combination with *in vivo* monoassociation for identifying bacterial determinants of host traits and provide insight into one mechanism by which the *D. melanogaster* gut microbiota affects host nutrient acquisition.

RESULTS

Functional traits of bacterial panel. To date, research on the function of individual gut microbiota members in *D. melanogaster* has focused on a few *Acetobacteraceae* or *Lactobacillus* species (21, 26–28), despite surveys indicating that a range of bacterial species, especially *Proteobacteria* or *Lactobacillales*, are detected in the *Drosophila* microbiota (22–25). To investigate the range of bacteria capable of associating with *Drosophila* and modifying host traits, we inoculated dechorionated eggs of *D. melanogaster* CantonS individually with each of 40 bacterial strains from the *Firmicutes* (including *Lactobacillus* strains) and *Proteobacteria* (including *Acetobacter* strains) and with one *Bacteroidetes* isolate. Only 4 strains were not detected in fly homogenates, *Sphingomonas wittichii*, *Streptococcus mutans*, *Streptococcus parasanguinas*, and *Dysgonomonas mossii*. *Gluconacetobacter oboediens* was the strain with the next-to-lowest abundance, detected at 1,700 CFU fly^{-1} (see Table S1 in the supplemental material). The *Firmicutes* strains generally attained higher CFU loads than the *Proteobacteria* strains, but exceptionally high CFU loads for the gammaproteobacterium *Pseudomonas putida* and low CFU loads for several *Firmicutes* species, including *Lactobacillus brevis* and *Lactobacillus rhamnosus*, were observed. These findings demonstrate that taxa other than the *Lactobacillales* and *Acetobacteraceae* can associate abundantly with *D. melanogaster*, suggesting that *D. melanogaster* is a permissive system.

Taxonomic specificity for microbiota effects on *Drosophila* traits have been demonstrated for *Acetobacter* and *Lactobacillus*, including species-level effects on host development time and TAG content (28) and strain-level effects on host development time (27). To investigate the taxonomic distribution of bacteria that influence these host traits, *Drosophila* was monoassociated with our 41-strain panel. We tested the hypothesis that bacterial effects on host traits are constrained to particular taxonomic levels by measuring larval development time and TAG content. Axenic flies have longer development times and higher TAG contents than flies that are raised with a conventional microbiota (26, 27, 31). Within phyla, there were significant differences between strains: the *Firmicutes* species *Lactococcus lactis* and *Enterococcus faecalis* conferred faster development than other *Firmicutes*, and the *Proteobacteria* species *Gluconobacter frateurii*, *Gluconacetobacter oboediens*, and *Escherichia coli* conferred slower development than other *Proteobacteria* (Fig. 1A). We also observed within-genus differences. For example, *Gluconacetobacter xylinus* conferred faster development than *Gluconacetobacter oboediens*, and *Lactobacillus brevis* DmelCS_002 conferred faster development than *Lactobacillus versmoldensis*.

Bacterial effects on host TAG suggested that the strongest reduction of host TAG content was generally limited to the *Acetobacteraceae*, with notable exceptions (Fig. 1B). *Acetobacteraceae* (*Alphaproteobacteria*) consistently reduced TAG more significantly than other strains, driven by the strongest reductions being among *Gluconacetobacter* species. Still, only modest reductions and high variations in TAG levels were apparent among *Acetobacter* species, and non-*Proteobacteria*, including *Bacillus subtilis* and *Lactobacillus casei*, conferred TAG levels that were comparable to those observed with *Acetobacteraceae*. Taken together, these findings demonstrate that bacterial effects on host development and TAG content cannot be defined by high-level taxonomic groupings and that, even at low taxonomic levels, there are significant differences between host traits conferred by different bacterial strains. Additionally, we found no evidence for bacterial antagonism to host traits relative to their expression in axenic flies; no monoassociated flies had slower development or higher TAG content than axenic flies (data not shown).

Genetic variation in bacterial panel. To understand genetic differences between bacterial strains, we clustered amino acid sequences by sequence similarity. Genome sequences were available from public databases for all but one of the strains we studied, *L. lactis* BPL1, for which we assembled and annotated a genome sequence (accession number JRFX00000000). Sequences were parsed into 30,581 groups using OrthoMCL, identifying 12,354 clusters of orthologous groups (COGs) that contained more than 1 amino acid sequence (see Dataset S1 in the supplemental material). Each COG was present in a specific set of bacterial taxa, and most COGs (82%) were present in a unique phylogenetic distribution group (PDG), i.e., they were not present in the same set of taxa as any other COG (Fig. S1). This high frequency of unique PDGs facilitated the use of a statistical approach to identify associations between genotype and phenotype. In all, 5,157 unique PDGs were identified in this analysis.

Metagenome-wide association with host traits. We performed MGWA to identify significant relationships between PDGs and each of two host traits: the rate of development and TAG content. Since the *D. melanogaster* flies were monoassociated with genome-sequenced bacteria, each metagenome consisted of the relevant bacterial genome sequence. For each of 5,157 PDGs, the statistical difference between traits in hosts monoassociated with PDG-containing strains and with PDG-lacking strains was determined using a test appropriate to the data (survival analysis using a Cox mixed model for development time and a linear mixed model for TAG). A significant hit was identified if the *P* value of the test passed a conservative Bonferroni correction ($P < 1 \times 10^{-5}$). Since 82% of PDGs contained a single COG and 95% of PDGs contained 5 or fewer COGs, most tests associated trait changes with one or a few genes, minimizing the incidence of false positives due to within-PDG COG cooccurrence (see Fig. S1 in the supplemental material).

Seven percent (347/5,157) of PDGs were significantly associated with larval development time after Bonferroni correction (Table 2; see Table S2 in the supplemental material). The prevalent functional classes among the 100 most significant COGs included respiratory metabolism (20 COGs), core cellular processes (e.g., translation and DNA replication; 17 COGs), and stress resistance (11 COGs). Among the top 5 most significant COGs were PqqC and PqqB, enzymes involved in the synthesis of the respiratory cofactor pyrroloquinoline quinone (PQQ). PQQ is a key

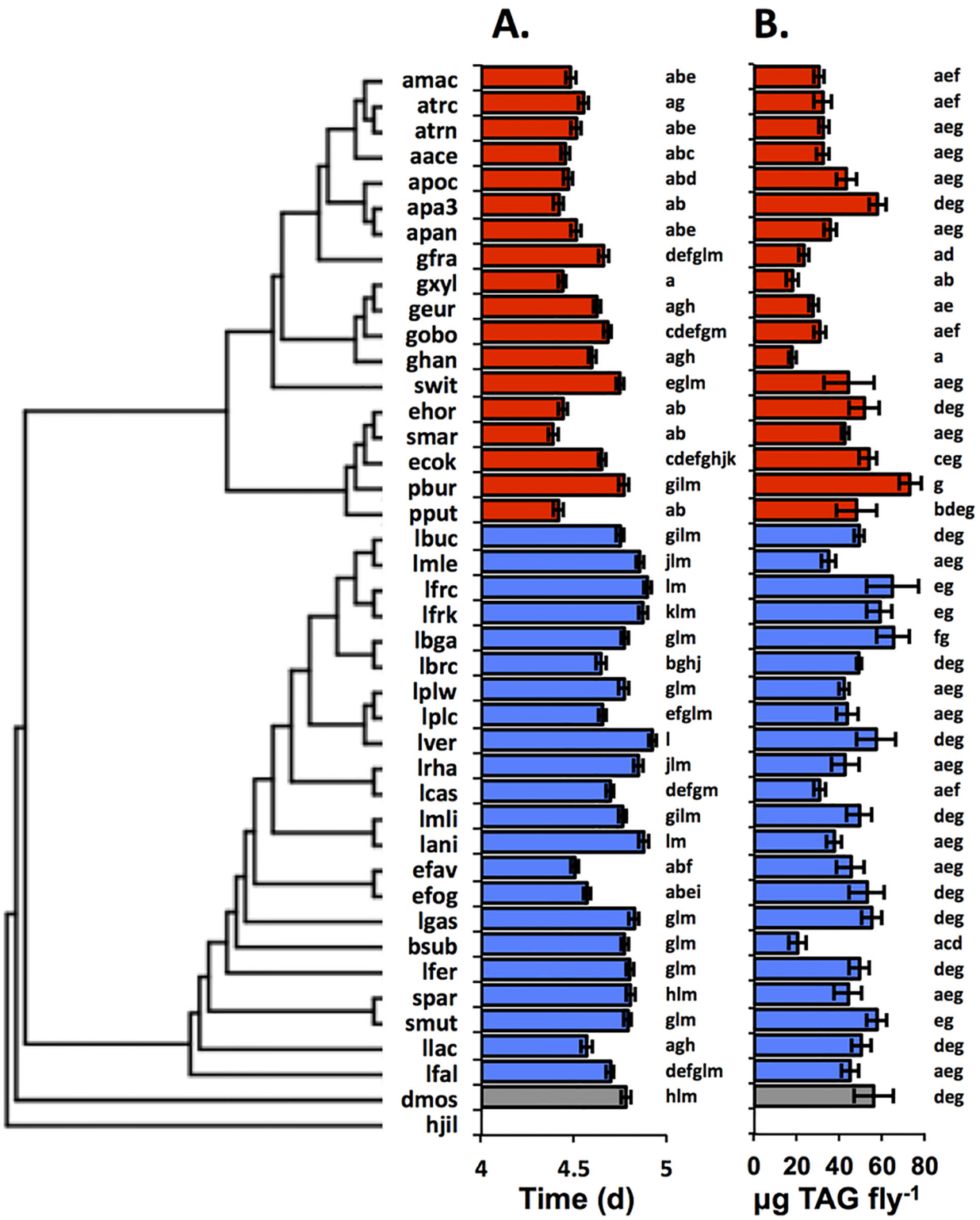

FIG 1 Larval development time and TAG content in monoassociated *D. melanogaster* flies. Traits for *D. melanogaster* that were monoassociated with each of 41 bacterial strains were measured. Phylogenetic trees were calculated using 16S sequences with unweighted branch lengths. Taxon abbreviations are defined in Table 1. Significant differences between treatments after Bonferroni correction ($P < 0.05$) are indicated by different letters next to bars. (A) Differences in bacterial effects on larval time to pupariation (development time) were observed between strains. Survival analysis using a Cox mixed model was used to identify significant differences between treatments, with experimental replicate and vial as random effects. To facilitate visualization, data are presented as the mean times to development ± standard errors of the means (SEM). (B) Differences in bacterial effects on TAG content were observed between strains. A linear mixed model was used to identify significant differences between treatments, with experimental replicate as a random effect. Data are presented as mean TAG content ± SEM (mean of experimental means). Red, *Proteobacteria*; blue, *Firmicutes*; gray, *Bacteroidetes*.

TABLE 2 Top significant associations between larval development rate and phylogenetic distribution group[a]

P value	No. of COGs	Notable annotated genes	PDG (no. of *Acetobacter* strains, no. of *Gluconacetobacter* strains, and other taxa containing COG[s])
2.20E−16	1	YciL protein	7, 4, *E. hormaechei*, *E. coli*, *P. putida*, *S.* marcescens
1.10E−15	3	Coenzyme PQQ synthesis B, C	7, 4, *G. frateurii*, *E. coli*, *S.* marcescens, *P. putida*, *E. hormaechei*
2.70E−15	1	Hypothetical protein	7, 4, *G. frateurii*, *E. coli*, *S.* marcescens, *S.* wittichii, *P. putida*, *P. burhodogranariea*, *E. hormaechei*, 2 *Enterococcus* strains
4.30E−14	1	Ankyrin-like protein	7, 4, *G. frateurii*, *E. coli*, *S.* marcescens, *P. putida*, *P. burhodogranariea*, *E. hormaechei*
7.80E−13	11	Malate:quinone oxidoreductase	7, 4, *G. frateurii*, *S. marcescens*, *S.* wittichii, *P. putida*
8.00E−13	3	Zinc uptake regulation protein ZUR	6, 4, *G. frateurii*, *E. coli*, *S.* marcescens, *P. putida*, *P. burhodogranaria*, *E. hormaechei*, 2 *Enterococcus* strains
8.80E−13	1	Deoxyribodipyrimidine photolyase	7, 4, *G. frateurii*, *E. coli*, *S.* marcescens, *S.* wittichii, *P. putida*, *E. hormaechei*
1.00E−12	6	Glucose dehydrogenase (PQQ dependent)	6, 4, *S. marcescens*, *S. wittichii*, *P. putida*, *E. hormaechei*
2.70E−12	1	Hypothetical protein	7, 2, *E. coli*, *S. marcescens*, *P. putida*, *P. burhodogranariea*, *E. hormaechei*
2.90E−12	1	Major facilitator superfamily transporter	7, 3, *G. frateurii*, *E. coli*, *S.* marcescens, *P. putida*, *P. burhodogranariea*, *E. hormaechei*
5.20E−12	1	Paraquat-inducible protein B	7, 4, *E. coli*, *S. marcescens*, *S.* wittichii, *P. putida*, *P. burhodogranariea*, *E. hormaechei*
3.40E−11	4	Succinate dehydrogenase iron-sulfur protein	7, 3, *G. frateurii*, *E. coli*, *S.* marcescens, *S.* wittichii, *P. putida*, *P. burhodogranariea*, *E. hormaechei*
5.00E−11	1	ATP-dependent helicase HrpB	7, 4, *E. coli*, *S. marcescens*, *E. hormaechei*
6.00E−11	1	Hypothetical protein	7, 3, *E. coli*, *S. marcescens*, *B. subtilis*, *P. putida*, *P. burhodogranariea*, *E. hormaechei*
6.10E−11	1	Aminodeoxychorismate lyase	7, 4, *G. frateurii*, *E. coli*, *S.* marcescens, *S.* wittichii, *P. putida*, *P. burhodogranariea*, *E. hormaechei*
6.70E−11	19	Flavin mononucleotide reductase YcdH	7, 4, *G. frateurii*, *S. marcescens*, *S.* wittichii, *P. putida*, *P. burhodogranariea*, *E. hormaechei*
7.20E−11	3	Multidrug resistance transporter HlyD	7, 4, *E. coli*, *S. marcescens*, *S.* wittichii, *P. putida*, *P. burhodogranariea*, *E. hormaechei*, 2 *Enterococcus* strains, *Dysgonomonas*
8.40E−11	2	NADH-ubiquinone oxidoreductase chain E, F	*B. subtilis*, 2 *Streptococcus* strains, 16 *Lactobacillus* strains, 1 *Enterococcus* sp., *Leuconostoc*
1.00E−10	2	Phenylalanine-tRNA ligase subunit beta	7, 4, *G. frateurii*, *E. coli*, *S.* marcescens, *S.* wittichii, *P. putida*, *P. burhodogranariea*, *E. hormaechei*, 1 *Enterococcus* sp., *Dysgonomonas*
1.00E−10	1	Deoxyguanosinetriphosphate triphosphohydrolase	*B. subtilis*, 2 *Streptococcus* strains, 16 *Lactobacillus* strains, *Leuconostoc*
1.00E−10	1	Thiamine pyrophosphokinase	7, 4, *G. frateurii*, *S. marcescens*, *P. putida*
1.30E−10	5	Oxidoreductase; alcohol dehydroenase cytochrome *c* subunit	7, 3, *E. coli*, *S. marcescens*, *S.* wittichii, *P. putida*, *P. burhodogranariea*, *E. hormaechei*
1.60E−10	1	Succinate dehydrogenase cytochrome *b* subunit	7, 4, *G. frateurii*, *S. marcescens*, *S.* wittichii, *P. putida*, *P. burhodogranariea*, *E. hormaechei*, 2 *Enterococcus*, *Dysgonomonas*
1.70E−10	1	Putative metal (Zn) chaperone	7, 4, *G. frateurii*, *E. coli*, *S.* marcescens, *B. subtilis*, *P. putida*, *P. burhodogranariea*, *E. hormaechei*
1.80E−10	3	Cytochrome *o* ubiquinol oxidase subunit II, III	7, 4, *G. frateurii*, *S. marcescens*, *B. subtilis*, *P. putida*, *P. burhodogranariea*, *E. hormaechei*

[a] For a full list, see Table S2.

component of oxidative metabolism in *Acetobacteraceae*, serving as the redox cofactor in several dehydrogenase enzymes, including glucose dehydrogenase (GDH) and alcohol dehydrogenase (ADH) (33). Transposon mutations in PqqB, PqqC, and PQQ-Adh have previously been shown to abolish the ability of *Acetobacter pomorum* to promote the development of monocolonized flies (26). Thus, MGWA with the host phenotype recapitulated the findings of a genetic screen for bacterial factors that promote development in *Drosophila*.

The COG with the highest-ranked association with TAG content was a predicted NADH-dependent, single-domain oxidoreductase (Table 3; see Table S3 in the supplemental material). This class of proteins has a broad range of candidate functions, including metabolism and redox sensing (34). As with development, several TAG-associated COGs were involved in oxidative metabolism, including ubiquinol oxidase (a.k.a. cytochrome *c* reductase), a Sco1-like cytochrome *c* oxidase biogenesis protein, and gluconate-2 dehydrogenase (GnDH; a cytochrome *c* subunit) (Table 3). These proteins may function in the same pathway of glucose oxidation, producing electrons that are channeled to the electron transport chain for energy generation by oxidative phosphorylation (Fig. S2). Of particular interest is GnDH, a three-subunit enzyme that oxidizes gluconate to 2-ketogluconate, usually after extracellular oxidation of glucose by GDH (33). After Bonferroni correction, one GnDH subunit was significant ($P = 0.002$, ranked number 4) (Table 3), and the remaining two GnDH subunits approached statistical significance ($P = 0.08$, ranked number 30) (Table 3).

Only one PDG was significantly associated with both larval development rate and TAG content. The two COGs represented by this PDG, cob(II)yrinic acid reductase and urate oxidase, were found in *B. subtilis*, *P. putida*, and all the *Acetobacteraceae*. They were associated with fast larval development and low TAG content.

Genes identified by MGWAs reduce host TAG content. We next sought to test the validity of statistical associations identified by MGWA. We focused on TAG-associated genes because several genes that we found to be associated with larval development rate have previously been shown to affect larval development rate (26). We first tested whether the gene encoding the COG with the most

TABLE 3 Significant associations between adult TAG content and phylogenetic distribution group[a]

P value	No. of COGs	Notable annotated genes	PDG (no. of *Acetobacter* strains, no. of *Gluconacetobacter* strains, other taxon[s] containing COG[s])
1.93E−009	1	NADH-dependent oxidoreductase	4, 4, *G. frateurii*, *B. subtilis*
2.34E−007	1	Gamma-glutamyltranspeptidase	7, 4, *G. frateurii*, *S. wittichii*, *B. subtilis*, *L. malefermentans*
2.41E−007	1	Dihydrolipoamide dehydrogenase	7, 4, *G. frateurii*, *S. wittichii*, *B. subtilis*, *L. casei*, 2 *Enterococcus* strains
3.36E−007	4	Gluconate 2-dehydrogenase subunit	4, 4, *G. frateurii*
4.06E−007	3	Copper efflux ATPase; metallo-dependent_hydrolases	7, 4, *G. frateurii*, *B. subtilis*
4.06E−007	1	MreC	All but *Acetobacteraceae* family and 1 *Enterococcus* sp.
5.43E−007	5	Cytochrome *c* oxidase biogenesis; CarD-like transcriptional regulator; nitrogen fixation	7, 4, *G. frateurii*, *S. wittichii*, *B. subtilis*
6.45E−007	5	Thioredoxin peroxidase (phytoene)	4, 4, *G. frateurii*, *S. wittichii*
1.02E−006	1	Phytoene biosynthesis	4, 4, *G. frateurii*, *L. plantarum*, *B. subtilis*
1.15E−006	3	MotA	*3*, 4, *G. frateurii*
1.30E−006	2	5-Aminolevulinate synthase	4, 4, *G. frateurii*, *S. marcescens*, *S.* wittichii, *B. subtilis*
1.71E−006	1	TonB-dependent outer membrane channel	4, 3, *G. frateurii*
1.73E−006	2	Multidrug efflux pump acriflavin resistance protein	3, 4, *G. frateurii*
2.01E−006	2	Transcriptional regulator	3, 4, *G. frateurii*, *S. wittichii*
2.09E−006	1	Putative hexosyltransferase	6, 4, *G. frateurii*, *B. subtilis*
2.38E−006	1	PEBP family protein	3, 4, *G. frateurii*, *S. wittichii*
2.94E−006	1	Hydroxyacylglutathione	2, 4, *B. subtilis*
2.94E−006	2	Cob(II)yrinic acid reductase	7, 4, *G. frateurii*, *B. subtilis*, *P. putida*
3.12E−006	1	Hypothetical	2, 4, *G. frateurii*
3.46E−006	1	Hypothetical	5, 4, *B. subtilis*
3.59E−006	2	TPR superfamily; ubiquinol oxidase	7, 4, *G. frateurii*, *B. subtilis*, *S.* wittichii, *P. putida*
7.27E−006	1	Putative tRNA modifying enzyme	7, 4, *G. frateurii*, *B. subtilis*, *S.* wittichii, *D. mossii*
7.98E−006	1	*N*-Acetyltransferase	11 *Lactobacillus* strains, 2 *Enterococcus* strains, 2 *Streptococcus* strains, *S. marcescens*, *E. coli*, *E. hormaechei*
8.35E−006	1	Transcriptional regulator	2, 4, *G. frateurii*
1.46E−005[b]	2	Gluconate 2-dehydrogenase subunits	4, 4, *G. frateurii*, *S. marcescens*, *P. putida*

[a] For a full list, see Table S3

[b] Not significant after Bonferroni correction for multiple tests ($P < 0.05$).

significant association with TAG content, a predicted single-domain oxidoreductase (SDR), was sufficient to impact host TAG. We introduced the gene into an *Acetobacter* strain that lacks it (*Acetobacter pasteurianus* 3p3) under the control of a constitutive promoter. Flies monoassociated with *A. pasteurianus* expressing SDR had significantly reduced TAG compared to those monoassociated with the empty vector control strain (Fig. 2A). Similar results were obtained when SDR was ectopically expressed in *Acetobacter tropicalis* DmCS_006, the strain from which the gene was cloned (Fig. 2A). These results validate the identification of SDR by MGWA as a COG associated with host TAG.

Several components of the GDH pathway were also among the most significant MGWA associations with host TAG, including GnDH, which had the 4th most significant association. To test whether GnDH was sufficient to modulate host TAG, a plasmid expressing all three subunits from a constitutive promoter was introduced into *A. pasteurianus* 3p3, which lacks the enzyme. Flies monoassociated with *A. pasteurianus* 3p3 expressing GnDH had significantly lower TAG contents than those bearing the empty vector (Fig. 2A), indicating that GnDH is sufficient to affect host TAG levels. Interestingly, introducing the same construct into *A. tropicalis* DmCS_006 (a fly isolate from which the genes were cloned) did not significantly reduce TAG in monoassociated hosts (Fig. 2A). This result indicated that a gain of function was not possible in this strain and suggested that elevating the activity of the GDH pathway rather than that of GnDH itself may be required to have an impact on the host TAG level. To test this hypothesis, we expressed GDH, the first enzyme in the pathway, in a similar fashion and found that this was sufficient to reduce TAG in both *A. tropicalis*- and *A. pasteurianus*-monoassociated flies (Fig. 2A). These results are consistent with a model in which consumption of glucose via the GDH pathway contributes to the reduction of host TAG levels by the gut microbiota. The results corroborate the association of multiple GDH pathway components with this trait in the MGWA. In support of the expectation that cloned genes were expressed, we detected reduced glucose content in the diet when the genes but not the empty vector were expressed (Fig. 2B; see discussion below). Species-specific glucose reduction by GnDH in *A. pasteurianus* but not *A. tropicalis* also supports the conclusion that the effects were gene specific and were not due to unexpected effects from the presence or induction of the construct. Parallel experiments demonstrated that these genetic manipulations of *Acetobacter* species did not significantly affect development time (data not shown), bacterial abundance on the fly diet (see Fig. S3A in the supplemental material), or the abundance of the bacteria in the flies (Fig. S3B).

Bacteria affect host TAG content through modification of dietary glucose content. Taken together, our results argue that glucose oxidation by the GDH pathway in *Acetobacteraceae* reduces host TAG levels. This result is consistent with previous findings that the influence of gut microbiota on host TAG levels is more pronounced on diets with high glucose content (4). To test whether dietary glucose abundance is affected by the ectopic expression of GDH pathway components, we measured the glucose content of the remaining diet after gnotobiotic rearing with each recombinant bacterial strain. For each instance where the recom-

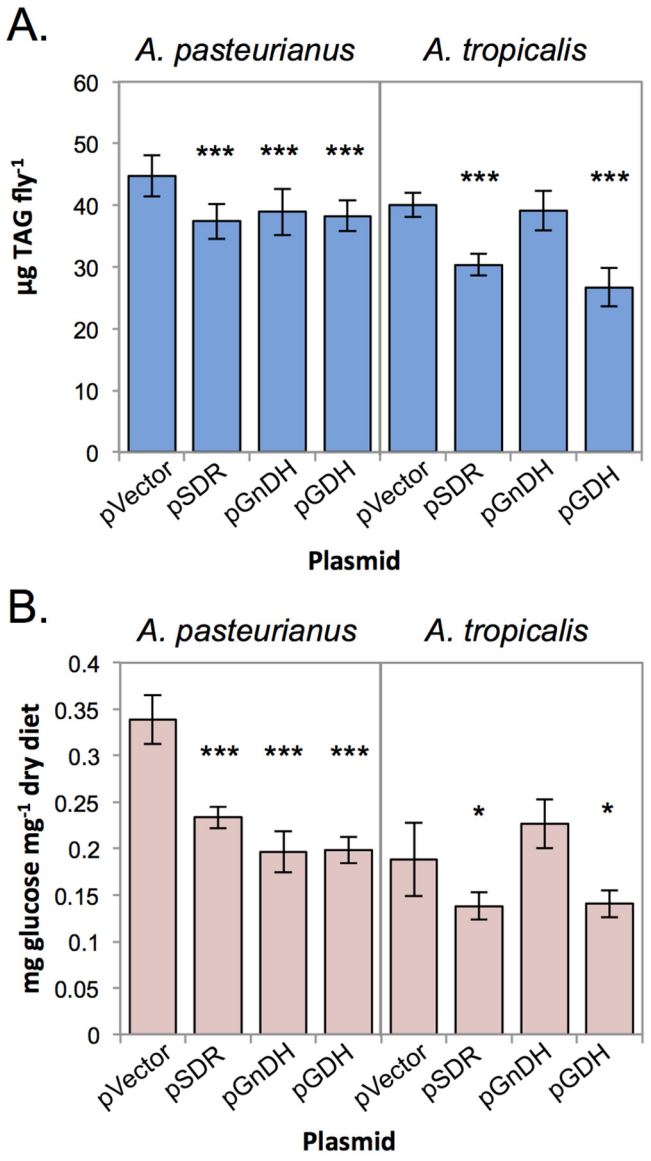

FIG 2 Experimental validation of candidate microbiota genes associated with host TAG level. (A) TAG content is shown for gnotobiotic flies monoassociated with either *A. pasteurianus* 3P3 or *A. tropicalis* DmCS_006 bearing the plasmids indicated. (B) Glucose content of fly diet after gnotobiotic rearing from egg to adulthood with recombinant strains bearing the indicated plasmids. Values are means ± standard errors for 3 experiments with 7 to 9 technical replicates each. Significant difference from the results for the control were determined by Dunnet's test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

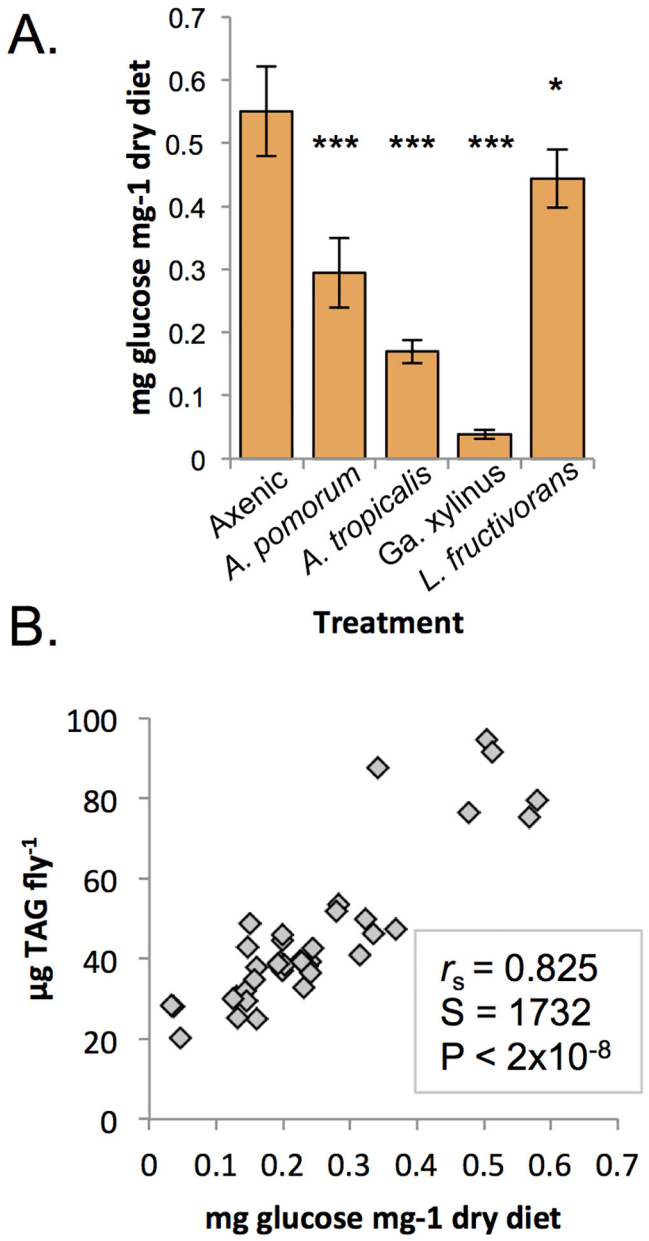

FIG 3 Microbiota effects on diet. (A) The glucose contents of fly diet after gnotobiotic rearing from egg to adulthood with a subset of strains from the 41-strain panel are shown. Microbiota treatments are indicated along the *x* axes. Values are means ± standard errors for 3 experiments with 7 to 9 technical replicates each. Significant differences from the results for the control were determined by Dunnet's test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). (B) Correlation between TAG contents of gnotobiotic flies and glucose contents of food remaining after rearing. Statistics are from Spearman's rank order test.

binant microbiota treatment reduced host TAG contents, there was a concomitant reduction in dietary glucose content (Fig. 2B). These data indicate that the ectopic expression of GDH in both *Acetobacter* species and of GnDH in *A. pasteurianus* can functionally reduce glucose in the food. The results also implicate glucose metabolism as a putative role for SDR in *Acetobacter*. The water and soluble protein contents of the diet did not differ significantly across treatments (see Table S4 in the supplemental material).

Since TAG reduction by recombinant strains correlates negatively with dietary glucose content, we tested for this correlation in flies monoassociated with strains at extreme ends of the host response spectrum; *G. xylinus*, which supports the lowest mean TAG level, and *L. fructivorans* DmCS_002, which supports the highest (Fig. 1B). Two *Acetobacter* species that result in intermediate host TAG levels were also included. All microbiota treatments significantly reduced the glucose content of the food compared to the glucose content using the axenic control (Fig. 3A), with *G. xylinus* having the most pronounced effect and *L. fructivorans* the least. Across all samples tested, there was a strong posi-

tive correlation between dietary glucose content and *D. melanogaster* TAG content (Fig. 3B). Uniquely, *G. xylinus* significantly increased the concentration of soluble protein in the diet (see Fig. S4B in the supplemental material). *L. fructivorans* resulted in a reduction in soluble protein that was not statistically significant but was sufficient to depress the protein/glucose ratio to a level equivalent to that under the axenic condition (Fig. S4C). These results support the hypothesis that microbiota effects on host TAG levels are determined by bacterial diet modifications.

DISCUSSION

A major challenge in gut microbiota research is identifying the genetic determinants of bacteria that determine host traits. Several studies have approached this challenge by correlating host phenotype with the taxonomic composition of the microbiota (11) or the metagenomic content of the microbiota (17). Other approaches have measured bacterial effects on host traits during *in vitro* cell culture (18–20). In this study, we expanded on these approaches through monoassociation of 41 sequenced strains with an animal host *in vivo*. This approach ensured that effects of microbiota on the host could be unambiguously assigned to bacterial taxon and genotype. Our results reveal that bacteria detected infrequently or in low abundance in *D. melanogaster* can monoassociate with the host and affect host traits. Utilizing a strain panel with broad representation of *Acetobacteraceae* and *Lactobacillales*, we observed that significant variation in host traits can be conferred by closely related (including conspecific) taxa within these groups. Furthermore, MGWA robustly identified bacterial genes that were experimentally demonstrated to affect host development (26) and lipid storage (this work). Genes that affect host TAG content cause correlated reductions in dietary glucose, suggesting that the effect of the microbiota on nutrient availability is an important factor in microbiota-dependent host nutrition.

We identified at least two benefits to an MGWA/monoassociation study relative to traditional mutant screens. First, mutant screens require individual hosts for association with thousands of mutants (this limitation does not apply when implementing selection regimes on mutant pools, as has been elegantly demonstrated previously; e.g., see references 14 and 35). Here, only 41 different preparations were used, providing a substantial technical advantage. Second, monoassociation with a bacterial panel provided insight into taxonomic specificity for host effects. We confirmed and extended previous reports of genus-, species-, and strain-specific effects of bacteria on microbiota-dependent *Drosophila* traits (27, 28), indicating that bacterial effects on host TAG and development are neither limited to nor consistent within a specific taxonomic group. These findings support the currently favored model that the metagenomic content of the microbiota is a better predictor of function than taxonomic composition (29, 36). For example, *B. subtilis* stood out among the *Firmicutes* as conferring the lowest TAG level, which could be because it is the only *Firmicute* we tested that utilizes aerobic respiration to consume glucose. *B. subtilis* also possesses a copy of the COG with the most significant association with TAG, a previously uncharacterized oxidoreductase, which our experiments implicate in glucose metabolism. Other notable examples of bacterial taxa that do not match their taxonomic neighbors include *Serratia marcescens* and *P. putida*, both of which promote rapid development. These bacteria share in common with *Acetobacteraceae* the production of PQQ, ADH, and GDH, which contributed in part to the identification of the relevant COGs in the MGWA for development.

D. melanogaster supported many *Proteobacteria* and *Firmicutes* strains at bacterial loads above 1,500 CFU fly^{-1}, including taxa that have never been detected in deep sequencing surveys of the gut microbiota (22–25). Thus, the host environment is not intrinsically antagonistic to taxa outside the groups that are frequently detected in high abundance (e.g., *Lactobacillales* and *Acetobacteraceae*). Instead, other factors must play important roles in shaping the gut microbiota composition, including availability in the *Drosophila* habitat and bacterial competition for the host niche or diet.

MGWA implicated oxidative metabolism of the gut microorganisms as a driver of host lipid storage. Multiple extracellular glucose oxidation pathway components were identified and confirmed to reduce host TAG levels and dietary glucose content. This result is congruent with important attributes of metabolism in some *Acetobacteraceae*, which utilize extracellular sugar and alcohol oxidation when carbon is abundant. For example, in *G. xylinus*, extracellular oxidation of glucose to gluconate by GDH accounted for >40% of glucose consumption (37). However, not all *Acetobacteraceae* that possess GDH produce acid from glucose; this activity is weak or variable among *Acetobacter* species, including *A. pasteurianus* and *A. pomorum* (38). Notably, GnDH shows a more limited distribution across genome-sequenced *Acetobacteraceae* than does GDH, which coincides with acid production from glucose (see Table S5 in the supplemental material). Our results indicate that strains bearing GnDH utilize more glucose in the fly diet, suggesting that this enzyme could be considered a marker for higher glucose oxidation potential in *Acetobacteraceae*.

As with TAG content, MGWA identified significant associations between components of the glucose oxidation pathway and faster larval development times, including GDH and gluconolactonase. This association is supported by evidence that excess dietary sugar prolongs development (4, 39). Inconsistent with this interpretation, ectopic expression of GDH or GnDH did not confer a reduction in development time. While this result could indicate that GDH was falsely associated with development, other explanations are also plausible. For example, further reduction in development time may be constrained by factors other than dietary glucose content. Timing could also be important: if diet modification by the bacteria is achieved gradually over the course of rearing, effects on adults (in which TAG was measured) would be more pronounced than effects on larvae. Further investigation is required to determine conclusively whether GDH and the glucose oxidation pathway are important for the promotion of development by microbiota.

This study provides several lines of evidence that bacterial modification of diet affects host traits (4, 40). First, microbiota that reduce host TAG contents produced correlated reductions in dietary glucose content over the course of rearing. This observation is consistent with prior studies that showed a positive correlation between the amount of glucose administered in the diet and the resulting TAG contents of *D. melanogaster* flies raised under both axenic and conventional conditions (4, 39). Specifically, doubling the dietary glucose concentration (while maintaining a constant yeast concentration) resulted in a 117% ± 40% increase in the TAG contents of conventionally raised female flies (4), which is roughly consistent with the 68% increase predicted from this study's correlation in host TAG and final dietary glucose concentration after gnotobiotic rearing (Fig. 3B). Second, bacterial

strains that did not associate with the fly still affected host traits. Four bacterial strains in this study did not associate with *D. melanogaster* above our limit of detection (0.5 CFU/fly). Despite this, *Drosophila* flies cultured with one of the strains, *S. wittichii*, displayed significantly reduced larval development time or TAG content relative to the results for axenic flies ($P = 4.5 \times 10^{-4}$). These effects could therefore be mediated by bacterial modification of the host diet. Overall, it is unclear to what extent dietary modification is attributable to microbial metabolism by microorganisms in the food versus microorganisms within the *Drosophila* gut. Modification of the fly rearing regimen to eliminate reingestion, for example by capillary feeding (41), would help distinguish between these possibilities. Additionally, further investigation is necessary to determine whether other *D. melanogaster* microbiota-responsive traits can also be mediated through dietary modifications or whether dietary effects are limited to a subset of microbiota-responsive host traits, including TAG content. The finding that host TAG responds quantitatively to dietary glucose could serve as a useful tool for unraveling the effects of microbiota on other traits; by altering the concentration of glucose in the diet, one could, in principal, normalize host TAG levels across microbiota treatments.

Taken together, findings that host TAG content correlates with dietary glucose content and that bacteria modify the composition of the diet suggest that the effects of microbiota on host nutrition come predominantly through changes in nutrient acquisition. This result contrasts with those of previous studies that implicate microbiota in nutrient allocation by host signaling pathways, e.g., the insulin pathway modulating TAG levels (26). However, this discrepancy is not irresolvable. Nutrient acquisition and allocation are regulated by linked regulatory networks (42–44) that are highly responsive to available dietary nutrients, and the microbiota are likely to interact with the host in multiple ways. The priority for future research is to understand how the host nutritional phenotype is shaped by the complex network of interactions between multiple bacterial functions and the host regulatory networks controlling nutrient acquisition and allocation.

This study demonstrates that gene-based genotype-phenotype association is a valuable tool for identifying bacterial determinants of host traits when employing monoassociated gnotobiotic animals. *D. melanogaster* is a superb model for this approach because diverse bacterial taxa can associate with the host and gnotobiotic rearing is cost effective and can be performed on a large scale. As implemented here, MGWA compares favorably to transposon mutagenesis, a more traditional method for assigning phenotype to bacterial genotype. Monoassociation of fewer than 50 bacterial genotypes with an animal host provided sufficient phenotypic and genetic resolution to identify bacterial genes that were successfully validated. Additionally, MGWA sampled a much larger group of genes (>12,000) by utilizing multiple strains and species. MGWA can also assign importance to essential genes or genes with redundant functions and does not require genetic manipulation of the strains employed. Future adaptations of this approach could explore bacterial effects on multiple host genotypes, employ defined multispecies communities, or vary environmental (e.g., dietary) conditions. This approach also has great potential to identify microbial genetic determinants of host traits in a variety of animal-bacterial symbioses.

MATERIALS AND METHODS

Cultivation of bacteria and flies. *D. melanogaster* Canton S. (*Wolbachia* free) flies were reared at 25°C in a 12-h-light/12-h-dark cycle on a yeast-glucose diet comprised of 100 g liter^{-1} brewer's yeast (inactive) (MP Biomedicals), 100 g liter^{-1} glucose (Sigma), 12 g liter^{-1} agar (Apex), and preservatives (0.04% phosphoric acid and 0.42% propionic acid [Sigma]). Axenic and gnotobiotic flies were reared on an autoclaved yeast-glucose diet without the addition of preservative. Gnotobiotic flies prepared with *Acetobacter* strains carrying plasmids were raised on food with 20 mg/liter chlortetracycline (Sigma).

The bacteria used in this study are listed in Table 1, along with culture conditions. The following media were used, with the addition of 1.5% agar (Apex) when solid medium was needed: modified MRS medium (28); potato medium (.5% glucose, 1% yeast extract, 1% peptone, 0.8% potato extract [Fluka 07915]); Luria broth (LB) (1% tryptone, 0.5% yeast extract, 0.5% sodium choloride); brain heart infusion (BHI) (3.7%; Becton Dickinson). All strains were cultured at 30°C. The oxygen conditions were created as follows: for oxic conditions, liquid culture was aerated by shaking at 200 rpm and solid-phase culture was used without atmospheric treatment, and for microoxic conditions, liquid culture was static and solid-phase culture was used in a CO_2-flooded airtight chamber. Transformed *E. coli* and *Acetobacter* strains were cultured with 5 mg/liter and 20 mg/liter chlortetracycline, respectively.

Preparation of axenic and gnotobiotic flies. Axenic and gnotobiotic flies were prepared as described previously (28). Briefly, eggs ≤24 hours old were collected from grape juice agar plates, surface sterilized by two 2.5-min washes in 0.6% hypochlorite, rinsed 3 times with sterile water, and aseptically transferred to sterile food in a laminar flow hood. Control axenic eggs were left undisturbed thereafter. Gnotobiotic eggs were inoculated with monocultures of each of the strains listed in Table 1. Innocula were prepared from PBS-washed cells harvested from stationary-phase cultures, normalized to a standard optical density at 600 nm (OD_{600}) of 0.1, and added directly to the food surface in a 50-μl volume. This amounted to the addition of ~1×10^7 CFU/vial. Bacteria were added within 3 h of egg transfer to sterile yeast-glucose medium.

Insect development. Insect development from egg to eclosion was observed three times daily, at 0, 6, and 11.5 h post-circadian light cycle initiation. Puparium formation and eclosion were recorded separately for each individual. Data from 3 separate experiments conducted in triplicate were collected. Data were analyzed in R Software for Statistical Computing, version 3.0.2 (45), using the Survival (46), coxme (47), and multcomp (48) packages as described previously (28).

Triglyceride content. TAG content was determined as described previously (28). Briefly, 3 to 5 mated females were collected under light CO_2 anesthesia 5 or 6 days posteclosion and homogenized in 125 μl TET buffer (10 mM Tris, pH 8, 1 mM EDTA, 0.1% Triton X-100) using a FastPrep-24 instrument (MP Biomedicals) with the default settings for 30 s. Forty microliters of cleared lysate was heated at 72°C for 20 min to inactivate endogenous enzymes prior to storage at −80°C. TAG was measured using a free glycerol detection kit in combination with triglyceride reagent, following the manufacturer's instructions (Sigma). For some experiment blocks, the lipase from *Pseudomonas* (20 U/ml in 20 mM potassium phosphate, 20 mM EDTA, and 20 mM magnesium chloride, pH 7.5) (product number L9518; Sigma) was substituted for the triglyceride reagent (product number T2449; Sigma). Four separate experiments were each conducted in triplicate (one replicate per vial), and data from all experiments were pooled for analysis by a linear mixed model (R, Nlme [49]) with microbial treatment as the fixed effect and experiment as a random effect. Multiple comparisons by Tukey's test were performed using the multcomp (48) package. All statistics were performed in R version 3.0.2. This approach allowed us to account for any blockwise variation among experiments. Pairwise comparisons were made via Tukey's test.

CFU determination. In each experiment described above, a sample of 3 to 5 female flies was homogenized in 100 μl of the culture medium preferred by each bacterial strain and 100 μl of lysis matrix D (MP Bio-

medicals) for 30 s on a FastPrep-24 instrument (MP Biomedicals). The homogenate was diluted to 1 ml with culture medium and spiral plated on agar medium using the WASP-2 apparatus (Microbiology International). CFU counts were made with a Protocol 3 colony counter (Microbiology International).

DNA isolation and sequencing. The genome of *Lactococcus lactis* BPL1, isolated from wild-caught *D. melanogaster* (B. Lazzaro, 2006), was sequenced, assembled, and annotated. Genomic DNA was isolated with the Qiagen DNeasy blood and tissue kit according to the manufacturer's recommendations, with pretreatment lysozyme digestion according to the protocol for Gram-positive bacteria. The Cornell Life Sciences Core Facility performed Illumina library preparation and sequencing using a 100-bp paired-end approach on an Illumina HiSeq 2000 instrument, obtaining 22,261,565 read pairs that passed quality filtering (1,930× coverage). We assembled the genome using Velvet 1/2/03 (50). A random subset of read pairs was destructively sampled into 10 subsets (sequence sets) of 1,600,000 read pairs (138× coverage each), estimated to be best assembled by a kmer length of 79 (http://dna.med.monash.edu.au/~torsten/velvet_advisor/). Each sequence set was assembled into contigs using a range of kmer lengths of around 79 (increments of 2), and for each sequence set assembly, an optimal kmer length was manually selected that minimized the contig number, maximized the N_{50} score and maximum contig length, and converged upon a common genome coverage across kmer lengths (in all cases, a length of 83 to 87). We also manually trimmed high-abundance, low-coverage reads and estimated actual kmer coverage from the assembly. The output contig file from each curated sequence set was used as the input in a second Velvet run with all other sequence sets for that genome to create a final assembly with a kmer length of 87 (after manual curation as described above). The final assembly was 2,306,222 bp in length, represented by 67 contigs, with a maximum contig size of 234,126 bp and N_{50} of 114,801 bp. Annotation using the RAST (Rapid Annotation using Subsystem Technology) server (51) produced an annotated genome sequence containing 2,299 features (59 RNAs and 2,240 open reading frames).

Phylogenetic analysis. To build reference trees for visualization and comparisons, full-length 16S rRNA gene sequences were downloaded from NCBI. When creating trees above the species level, a single representative sequence from each taxonomic level was selected. Multiple sequence alignments and unweighted-pair group method using average linkages (UPGMA) trees were created using MUSCLE (52). The trees were visualized and formatted in R version 3.0.2 (45) using the ape package (53).

Metagenome-wide association. Amino acid sequences from predicted open reading frames in each genome were clustered into clusters of orthologous groups (COGs) *de novo* relative to predicted open reading frames in all other taxa used in this study. Amino acid sequence files were extracted from NCBI. For *Acetobacter pasteurianus* 3P3, only nucleotide contig files were available, so we called open reading frames in RAST. ORFs were also called in RAST for *A. pomorum* DmCS_004, *A. malorum* DmCS_005, *A. tropicalis* DmCS_006, *L. brevis*, DmCS_003, *L. fructivorans* DmCS_002, and *L. plantarum* DmCS_001, sequenced previously (P. D. Newell and A. E. Douglas, unpublished data); and for *L. lactis* BPL1, sequenced in this study. Amino acid sequences were formatted for OrthoMCL and subsequently searched against all amino acid sequences in the taxon pool using a custom sequence database and blastall (Blast 2.2.26; database updated March 2013).

COGs were called using default instructions for OrthoMCL with an inflation factor of 1.5 (54–57). Custom perl scripts were used to define the representation of sequences from each COG in and across bacterial taxa. A representative gene for each COG was selected using HMMer (61). Briefly, an HMM profile for each COG was built from an alignment of all COG sequences created using Muscle (52). The best match for the HMM profile against all protein sequences from taxa in this study was identified using HMMsearch and selected as a representative sequence for the cluster. The annotation of the best match was retained as the annotation for the cluster. These representative sequences are listed in Dataset S1.

MGWA of microbiota-dependent variation in host trait magnitude with microbiota gene content (defined as COGs with least 2 amino acid sequences per category) was performed in R. In each case, the significance of the fixed effect of PDG presence/absence identified genes statistically associated with changes in microbiota-dependent host trait magnitude. For TAG, the data were normalized by square root transformation and analyzed with the presence/absence of the PDG as a fixed effect and experiment nested within bacterial species as random effects in a linear mixed model in R (49). For development, a survival model with the same parameters was used (46, 47); vial was omitted as an additional nested random factor due to prohibitive time requirements to calculate the model. In each case, a Bonferroni correction threshold (at $P = 0.05$) was applied to identify significantly associated genes.

Cloning and expression of genes in *Acetobacter*. Genes were amplified by PCR from *A. tropicalis* DmCS_006 genomic DNA and cloned using enzymes from New England Biolabs according to the manufacturer's recommendations. All products were amplified using Phusion polymerase, cut with restriction enzymes as indicated below, and ligated with T4 DNA ligase to the plasmid pCM62 (59) (prepared by cutting with the same enzymes as used for the product to be cloned) such that the P*lac* promoter would drive the expression of the cloned gene. Glucose dehydrogenase (GDH) was amplified with forward (5′ TAACTGCAGCTCTGAGAGAA ACACATCATGCAAGAGAG 3′; underline indicates restriction sites here and in subsequent primers) and reverse (5′ TAAGGATCCGTCTGCTTA TTGGGCGTTGGAGCC 3′) primers with an annealing temperature of 60.5°C and extension time of 1 min 45 s, and the product was cloned with PstI and BamHI. Gluconate dehydrogenase is encoded by three genes, which form an operon in the *A. tropicalis* chromosome. The three genes were cloned as one unit, with forward (5′ TATAAGCTTGCACAATCTG GACGCTTGGTACAAGG 3′) and reverse (5′ TAAGGATCCCAGAACT TGTTATTCCGTTGCCTGTCCG 3′) primers with an annealing temperature of 61°C and extension time of 3 min, and the product cloned with HindIII and BamHI. The single domain oxidoreductase (SDR) was amplified with forward (5′ TAACTGCAGGCTTGTCAGCCGCTTGGTGG C 3′) and reverse (5′ TAAGGATCCCCCGTTATCCACCATCCGCCAG 3′) primers with an annealing temperature of 60°C and extension time of 1 min, and the product cloned with PstI and BamHI. Ligated products were used to transform *E. coli* S17 λ-Pir via electroporation. Clones were confirmed by restriction digest and sequencing. Plasmids were introduced into *Acetobacter* cells by conjugation as described previously (60). Briefly, bacteria were cultured overnight in Luria broth (*E. coli*) or potato medium (*Acetobacter*) and cells from 0.5 ml of culture were harvested by centrifugation. Cells of donor and receiver were washed separately in sterile growth medium twice, resuspended in a final volume of 50 μl of potato medium, and then mixed together and transferred to a fresh potato medium plate. After incubation at 30°C for 16 h, cells were harvested and plated onto YPG medium (0.5% yeast extract, 0.5% peptone, 1% glycerol, 1.5% agar) containing 0.2% acetic acid and 20 mg/liter chlortetracycline. Colonies that appeared after 48 h of incubation were subcultured on potato medium supplemented with 20 mg/liter chlortetracycline. Transformation of *Acetobacter* was confirmed by plasmid isolation and subsequent transformation of *E. coli*.

Measurement of food contents. At 5 or 6 days posteclosion, flies were removed from the vial, and a small aliquot of spent food (~25 mg) was removed from the top, taking care to exclude eggs and larvae. The food was lyophilized at −80°C (FreeZone lyophilizer; Labconco) and then weighed to the nearest microgram using a Mettler Toledo MX5 microbalance. The sample was then homogenized in 100 μl of TET buffer with 100 μl of lysis matrix D (MP Biomedicals) for 1 min on a FastPrep-24 instrument (MP Biomedicals). Debris was pelleted by centrifugation, and the supernatant removed and frozen immediately. After thawing the sample on ice, the glucose content was determined by the glucose oxidase method, as implemented previously (28), and the soluble protein content

determined with the Bio-Rad DC kit according to the manufacturer's instructions. Three replicate vials were tested for each of three biological replicates.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01631-14/-/DCSupplemental.

Figure S1, TIF file, 2.5 MB.
Figure S2, TIF file, 5 MB.
Figure S3, TIF file, 4.5 MB.
Figure S4, TIF file, 6.8 MB.
Table S1, DOCX file, 0.1 MB.
Table S2, CSV file, 0.5 MB.
Table S3, CSV file, 0.5 MB.
Table S4, DOCX file, 0.05 MB.
Table S5, DOCX file, 0.1 MB.
Dataset S1, TXT file, 4.7 MB.

ACKNOWLEDGMENTS

We thank the following for generously sharing strains or plasmids: F. Barja, University of Geneva; J. Helmann, Cornell University; D. Hogan, Geisel School of Medicine at Dartmouth; J. Keagle, Mansfield University; B. Lazzaro, Cornell University; M. Lidstrom, University of Washington; Siquing Liu, USDA; J. March, Cornell University; K. Matsushita, Yamaguchi University; E. Martens, University of Michigan; J. Peters, Cornell University; M. Tien, Pennsylvania State University; H. Wu, University of Alabama; and Winclove b.v., Netherlands.

We acknowledge the support of the National Institutes of Health (NIH), National Institute for General Medical Science (NGIMS), via the Ruth L. Kirschstein NRSA postdoctoral fellowship 1F32GM099374-01 to P.D.N. and NIH grant 1R01GM095372 to A.E.D. We also acknowledge support from the Sarkaria Institute for Insect Physiology and Toxicology.

The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the NGIMS or the NIH.

REFERENCES

1. **Bercik P, Denou E, Collins J, Jackson W, Lu J, Jury J, Deng Y, Blennerhassett P, Macri J, McCoy KD, Verdu EF, Collins SM.** 2011. The intestinal microbiota affect central levels of brain-derived neurotropic factor and behavior in mice. Gastroenterology **141:**599–609.e3. http://dx.doi.org/10.1053/j.gastro.2011.04.052.
2. **Smith K, McCoy KD, Macpherson AJ.** 2007. Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. Semin. Immunol. **19:**59–69. http://dx.doi.org/10.1016/j.smim.2006.10.002.
3. **Venu I, Durisko Z, Xu J, Dukas R.** 2014. Social attraction mediated by fruit flies' microbiome. J. Exp. Biol. **217:**1346–1352. http://dx.doi.org/10.1242/jeb.099648.
4. **Wong AC, Dobson AJ, Douglas AE.** 2014. Gut microbiota dictates the metabolic response of *Drosophila* to diet. J. Exp. Biol. **217:**1894–1901. http://dx.doi.org/10.1242/jeb.101725.
5. **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.
6. **Lize A, McKay R, Lewis Z.** 2014. Kin recognition in *Drosophila*: the importance of ecology and gut microbiota. ISME J. **8:**469–477. http://dx.doi.org/10.1038/ismej.2013.157.
7. **Hsiao EY, McBride SW, Hsien S, Sharon G, Hyde ER, McCue T, Codelli JA, Chow J, Reisman SE, Petrosino JF, Patterson PH, Mazmanian SK.** 2013. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. Cell **155:**1451–1463. http://dx.doi.org/10.1016/j.cell.2013.11.024.
8. **Neufeld KM, Kang N, Bienenstock J, Foster JA.** 2011. Reduced anxiety-like behavior and central neurochemical change in germ-free mice. Neurogastroenterol. Motil. **23:**255–264, e119. http://dx.doi.org/10.1111/j.1365-2982.2010.01620.x.
9. **Neufeld KA, Kang N, Bienenstock J, Foster JA.** 2011. Effects of intestinal microbiota on anxiety-like behavior. Commun. Integr. Biol. **4:**492–494. http://dx.doi.org/10.4161/cib.4.4.15702.
10. **Cryan JF, Dinan TG.** 2012. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. Nat. Rev. Neurosci. **13:**701–712. http://dx.doi.org/10.1038/nrn3346.
11. **Faith JJ, Ahern PP, Ridaura VK, Cheng J, Gordon JI.** 2014. Identifying gut microbe-host phenotype relationships using combinatorial communities in gnotobiotic mice. Sci. Transl. Med. **6:**220ra211. http://dx.doi.org/10.1126/scitranslmed.3008051.
12. **Greenblum S, Turnbaugh PJ, Borenstein E.** 2012. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proc. Natl. Acad. Sci. U. S. A. **109:**594–599. http://dx.doi.org/10.1073/pnas.1116053109.
13. **Degnan PH, Barry NA, Mok KC, Taga ME, Goodman AL.** 2014. Human gut microbes use multiple transporters to distinguish vitamin B_{12} analogs and compete in the gut. Cell Host Microbe **15:**47–57. http://dx.doi.org/10.1016/j.chom.2013.12.007.
14. **Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI.** 2009. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host Microbe **6:**279–289. http://dx.doi.org/10.1016/j.chom.2009.08.003.
15. **Maurice CF, Haiser HJ, Turnbaugh PJ.** 2013. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. Cell **152:**39–50. http://dx.doi.org/10.1016/j.cell.2012.10.052.
16. **Wier AM, Nyholm SV, Mandel MJ, Massengo-Tiassé RP, Schaefer AL, Koroleva I, Splinter-Bondurant S, Brown B, Manzella L, Snir E, Almabrazi H, Scheetz TE, Bonaldo MdeF, Casavant TL, Soares MB, Cronan JE, Reed JL, Ruby EG, McFall-Ngai MJ.** 2010. Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical and metabolic change in a beneficial symbiosis. Proc. Natl. Acad. Sci. U. S. A. **107:**2259–2264. http://dx.doi.org/10.1073/pnas.0909712107.
17. **Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto JM, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, Wang J, Ehrlich SD, Nielsen R, Pedersen O, Kristiansen K, Wang J.** 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature **490:**55–60. http://dx.doi.org/10.1038/nature11450.
18. **van Hemert S, Meijerink M, Molenaar D, Bron PA, de Vos P, Kleerebezem M, Wells JM, Marco ML.** 2010. Identification of *Lactobacillus plantarum* genes modulating the cytokine response of human peripheral blood mononuclear cells. BMC Microbiol. **10:**293. http://dx.doi.org/10.1186/1471-2180-10-293.
19. **Meijerink M, van Hemert S, Taverne N, Wels M, de Vos P, Bron PA, Savelkoul HF, van Bilsen J, Kleerebezem M, Wells JM.** 2010. Identification of genetic loci in *Lactobacillus plantarum* that modulate the immune response of dendritic cells using comparative genome hybridization. PLoS One **5:**e10632. http://dx.doi.org/10.1371/journal.pone.0010632.
20. **Spinler JK, Sontakke A, Hollister EB, Venable SF, Oh PL, Balderas MA, Saulnier DM, Mistretta TA, Devaraj S, Walter J, Versalovic J, Highlander SK.** 2014. From prediction to function using evolutionary genomics: human-specific ecotypes of *Lactobacillus reuteri* have diverse probiotic functions. Genome Biol. Evol. **6:**1772–1789. http://dx.doi.org/10.1093/gbe/evu137.
21. **Ryu JH, Kim SH, Lee HY, Bai JY, Nam YD, Bae JW, Lee DG, Shin SC, Ha EM, Lee WJ.** 2008. Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in *Drosophila*. Science **319:**777–782. http://dx.doi.org/10.1126/science.1149357.
22. **Wong AC, Chaston JM, Douglas AE.** 2013. The inconstant gut microbiota of *Drosophila* species revealed by 16S rRNA gene analysis. ISME J. **7:**1922–1932. http://dx.doi.org/10.1038/ismej.2013.86.
23. **Chandler JA, Lang JM, Bhatnagar S, Eisen JA, Kopp A.** 2011. Bacterial communities of diverse *Drosophila* species: ecological context of a host-microbe model system. PLoS Genet. **7:**e1002272. http://dx.doi.org/10.1371/journal.pgen.1002272.
24. **Staubach F, Baines JF, Kunzel S, Bik EM, Petrov DA.** 2013. Host species and environmental effects on bacterial communities associated with *Drosophila* in the laboratory and in the natural environment. PLoS One **8:**e70749. http://dx.doi.org/10.1371/journal.pone.0070749.

25. **Wong CN, Ng P, Douglas AE.** 2011. Low-diversity bacterial community in the gut of the fruitfly *Drosophila melanogaster*. Environ. Microbiol. **13:**1889–1900. http://dx.doi.org/10.1111/j.1462-2920.2011.02511.x.
26. **Shin SC, Kim SH, You H, Kim B, Kim AC, Lee KA, Yoon JH, Ryu JH, Lee WJ.** 2011. *Drosophila* microbiome modulates host developmental and metabolic homeostasis via insulin signaling. Science **334:**670–674. http://dx.doi.org/10.1126/science.1212782.
27. **Storelli G, Defaye A, Erkosar B, Hols P, Royet J, Leulier F.** 2011. *Lactobacillus plantarum* promotes *Drosophila* systemic growth by modulating hormonal signals through TOR-dependent nutrient sensing. Cell Metab. **14:**403–414. http://dx.doi.org/10.1016/j.cmet.2011.07.012.
28. **Newell PD, Douglas AE.** 2014. Interspecies interactions determine the impact of the gut microbiota on nutrient allocation in *Drosophila melanogaster*. Appl. Environ. Microbiol. **80:**788–796. http://dx.doi.org/10.1128/AEM.02742-13.
29. **Microbiome, Human Project C.** 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207–214. doi:10.1038/nature11234. PubMed.
30. **Rogers GB, Kozlowska J, Keeble J, Metcalfe K, Fao M, Dowd SE, Mason AJ, McGuckin MA, Bruce KD.** 2014. Functional divergence in gastrointestinal microbiota in physically separated genetically identical mice. Sci. Rep. **4:**5437.
31. **Ridley EV, Wong AC, Westmiller S, Douglas AE.** 2012. Impact of the resident microbiota on the nutritional phenotype of *Drosophila melanogaster*. PLoS One 7:e36765. http://dx.doi.org/10.1371/journal.pone.0036765.
32. **Baker KD, Thummel CS.** 2007. Diabetic larvae and obese flies—emerging studies of metabolism in *Drosophila*. Cell Metab. **6:**257–266. http://dx.doi.org/10.1016/j.cmet.2007.09.002.
33. **Mamlouk D, Gullo M.** 2013. Acetic acid bacteria: physiology and carbon sources oxidation. Indian J. Microbiol. **53:**377–384. http://dx.doi.org/10.1007/s12088-013-0414-z.
34. **Kavanagh KL, Jörnvall H, Persson B, Oppermann U.** 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families: the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. Cell. Mol. Life Sci. **65:**3895–3906. http://dx.doi.org/10.1007/s00018-008-8588-y.
35. **Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW.** 1995. Simultaneous identification of bacterial virulence genes by negative selection. Science **269:**400–403. http://dx.doi.org/10.1126/science.7618105.
36. **Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI.** 2009. A core gut microbiome in obese and lean twins. Nature **457:**480–484. http://dx.doi.org/10.1038/nature07540.
37. **Zhong C, Zhang GC, Liu M, Zheng XT, Han PP, Jia SR.** 2013. Metabolic flux analysis of *Gluconacetobacter xylinus* for bacterial cellulose production. Appl. Microbiol. Biotechnol. **97:**6189–6199. http://dx.doi.org/10.1007/s00253-013-4908-8.
38. **Boone DR, Castenholz RW, Garrity GM.** 2001. Bergey's manual of systematic bacteriology, p. 41–54. In The Proteobacteria, vol. **2**, 2nd ed. Springer Verlag, New York, NY.
39. **Matzkin LM, Johnson S, Paight C, Bozinovic G, Markow TA.** 2011. Dietary protein and sugar differentially affect development and metabolic pools in ecologically diverse *Drosophila*. J. Nutr. **141:**1127–1133. http://dx.doi.org/10.3945/jn.111.138438.
40. **Broderick NA, Lemaitre B.** 2012. Gut-associated microbes of *Drosophila melanogaster*. Gut Microbes **3:**307–321. http://dx.doi.org/10.4161/gmic.19896.
41. **Ja WW, Carvalho GB, Mak EM, de la Rosa NN, Fang AY, Liong JC, Brummel T, Benzer S.** 2007. Prandiology of *Drosophila* and the CAFE assay. Proc. Natl. Acad. Sci. U. S. A. **104:**8253–8256. http://dx.doi.org/10.1073/pnas.0702726104.
42. **Danielsen ET, Moeller ME, Rewitz KF.** 2013. Nutrient signaling and developmental timing of maturation. Curr. Top. Dev. Biol. **105:**37–67. http://dx.doi.org/10.1016/B978-0-12-396968-2.00002-6.
43. **Oldham S.** 2011. Obesity and nutrient sensing TOR pathway in flies and vertebrates: functional conservation of genetic mechanisms. Trends Endocrinol. Metab. **22:**45–52. http://dx.doi.org/10.1016/j.tem.2010.11.002.
44. **Grewal SS.** 2009. Insulin/TOR signaling in growth and homeostasis: a view from the fly world. Int. J. Biochem. Cell Biol. **41:**1006–1010. http://dx.doi.org/10.1016/j.biocel.2008.10.010.
45. **Core Team R.** 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
46. **Therneau T.** 2014. package for survival analysis in S. R package version 2.37-7. http://CRAN.R-project.org/package=survival.
47. **Therneau T.** 2012. coxme: mixed effects Cox models. R package version **2.**-2–3. http://cran.r-project.org/web/packages/coxme/index.html.
48. **Hothorn T, Bretz F, Westfall P.** 2008. Simultaneous inference in general parametric models. Biom. J. **50:**346–363. http://dx.doi.org/10.1002/bimj.200810425.
49. **Pinheiro J, Bates D, DebRoy S, Sarkar D, Core Team R.** 2014. nlme: linear and nonlinear mixed effects models. R package version **3.**-1–117. http://cran.r-project.org/web/packages/nlme/citation.html.
50. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
51. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics **9:**75. http://dx.doi.org/10.1186/1471-2164-9-75.
52. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:**1792–1797. http://dx.doi.org/10.1093/nar/gkh340.
53. **Paradis E, Claude J, Strimmer K.** 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics **20:**289–290. http://dx.doi.org/10.1093/bioinformatics/btg412.
54. **Chen F, Mackey AJ, Vermunt JK, Roos DS.** 2007. Assessing performance of orthology detection strategies applied to eukaryotic genomes. PLoS One **2:**e383. http://dx.doi.org/10.1371/journal.pone.0000383.
55. **Chen F, Mackey AJ, Stoeckert CJ, Jr, Roos DS.** 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. **34:**D363–D368. http://dx.doi.org/10.1093/nar/gkj123.
56. **Li L, Stoeckert CJ, Jr, Roos DS.** 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13:**2178–2189. http://dx.doi.org/10.1101/gr.1224503.
57. **Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, Shanmugam D, Roos DS, Stoeckert CJ, Jr..** 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Curr. Protoc. Bioinformatics Chapter 6:Unit 6.12.1–19. http://dx.doi.org/10.1002/0471250953.bi0612s35.
58. **Eddy SR.** 2011. Accelerated profile HMM searches. PLoS Comp. Biol. **7**(10):e1002195.
59. **Marx CJ, Lidstrom ME.** 2001. Development of improved versatile broad-host-range vectors for use in methylotrophs and other gram-negative bacteria. Microbiology **147**(Pt 8):2065–2075.
60. **Deeraksa A, Moonmangmee S, Toyama H, Yamada M, Adachi O, Matsushita K.** 2005. Characterization and spontaneous mutation of a novel gene, polE, involved in pellicle formation in *Acetobacter tropicalis* SKU1100. Microbiology **151:**4111–4120. http://dx.doi.org/10.1099/mic.0.28350-0.
61. **Finn RD, Clements J, Eddy SR.** 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. **39:**W29–W37. http://dx.doi.org/10.1093/nar/gkr367.