EFFECT OF AMONG-MICROBE INTERACTIONS ON *DROSOPHILA MELANOGASTER* NUTRITION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

John George McMullen II

December 2020

EFFECT OF AMONG-MICROBE INTERACTIONS ON *DROSOPHILA MELANOGASTER* NUTRITION

John George McMullen II, Ph. D.

Cornell University 2020

The animal gut can be colonized by diverse microorganisms that impact the nutritional physiology of the animal host. Most research has concerned the effect of individual microorganisms, which may compete for dietary constituents, provide essential nutrients or modulate host signaling pathways regulating nutrient allocation and metabolic function. The goal of this dissertation was to investigate the effect of among-microbe interactions on the nutritional physiology of the animal host, using the *Drosophila melanogaster* gut microbiome as the experimental system. The specific aims were to 1) determine the genetic capacity for metabolic function of individual bacteria associated with *Drosophila* by genome sequence analysis, 2) investigate the effect of among-microbe interactions on *Drosophila* nutrient allocation and the underlying processes, and 3) assess the effect of microbial co-associations on the *Drosophila* metabolome. Results indicated that bacteria associated with *Drosophila* differ metabolically at multiple levels of taxonomy, from order to within-species, with redundant functions predicted to influence *Drosophila* performance and physiology. Complementary experimental studies using the bacteria *Acetobacter fabarum* and *Lactobacillus brevis* and yeast *Hanseniaspora uvarum* revealed that the nutritional traits of *Drosophila* vary both with the composition of the microbiota and between

male and female flies. For example, the lipid content of male flies was negatively correlated with the titer of the microbial fermentation product acetic acid when the flies were co-associated with *A. fabarum* and *H. uvarum*. The metabolomic analysis of *Drosophila* provided evidence that the interaction between *A. fabarum* and *L. brevis* stimulates metabolic signaling to increase metabolic activity through the TCA cycle and for male flies, decrease glucose and lipid content. In addition, the microbial metabolite phenyllactic acid was identified as a statistically robust biomarker for low lipid content in *Drosophila*. Taken together, these studies reveal that nutritional traits of *Drosophila* are strongly influenced by the presence and composition of the gut microbiota, and that the effect of taxonomically-complex communities cannot be predicted from the traits of *Drosophila* bearing single microbial taxa. An imperative for future research is to determine the processes by which male and female hosts differ in their nutritional responses to microbial communities of different complexity and composition.

BIOGRAPHICAL SKETCH

John McMullen (they/them, he/him) joined Cornell University in 2015 for a Ph.D. in the field of Entomology to study the effect of the gut microbiome on *Drosophila* nutrition using a chemical ecology approach. Prior to their time at Cornell, John completed a joint B.S./M.S. program in Microbiology at the University of Arizona. During their five years at Arizona, they researched the symbiosis between insect pathogenic nematodes in the genus *Steinernema* and *Xenorhabdus* bacteria under the tutelage of Dr. S. Patricia Stock. Specifically, John investigated the effect of strain-level variation on nematode partner preference and identified genetic determinants of bacterial-mediated insect virulence in the bacterium *Xenorhabdus bovienii*. This experience opened John to the fascinating world of microbial symbioses and the chemistry that mediates these interactions, which led them to apply to Cornell University to study under Dr. Angela Douglas. John wanted to gain expertise in comparative genomics, metabolomics, and nutritional physiology techniques to tease apart the complex associations among multi-species interactions. During their ~5.5 years in the Douglas Lab, John has used the *Drosophila* gut microbiome system with experimental, statistical, and bioinformatic approaches to understand how among-microbe interactions influence fly nutritional traits and metabolic function.

For my mom and my grandma.

The two most wonderful people that have constantly supported me throughout my life.

I only wish my grandma was still with us to celebrate this milestone.

ACKNOWLEDGMENTS

I would like to first start by thanking my advisor, Dr. Angela Douglas. I am leaving her lab a more critical and thoughtful scientist and a linear-thinking writer because of her. Angela has been completely supportive of my research endeavors and has shared so much of her knowledge about symbiosis and microbiomes during my time in her lab. I also thank her for creating a diverse lab full of amazing experts in a wide range of fields. From day one in the lab, I joined what felt like a fast-moving train and I started homogenizing flies to isolate bacteria and yeasts, but I have learned more than I could have imagined possible during the past five years.

To my committee members: Dr. John Chaston, Dr. Greg Loeb, and Dr. Mike Sheehan, I give special thanks to you all for the hard and thoughtful questions on my research as well as just the positive encouragement and support. This would not be possible without them, and I appreciate how all their different backgrounds have pushed how I think about problems from spatial-temporal aspects of my questions to their application to manage insect pests.

I would like to also thank all the members of the Douglas Lab, who have been such amazing people to work with and learn from. I would like to especially thank Dr. Karen Adair, Dr. Alyssa Bost, Dr. Nana Ankrah, and Dr. Fran Blow for their amazing mentorship and for taking the time to answer a billion questions as well as their friendship. In addition, I would like to thank Eduardo Bueno, Marita Wilson, Dr. Dave Kang, Dr. Arinder Aurora, Dr. Caroline Fromont, Grace Peters-Schulze, Noah Clark, Danielle Preston, Dr. Trevor Tivey, Dr. Michael Stephens, and Dr. Sunny Yoon for their friendship and general shenanigans!

In addition, I have made so many friends during my time at Cornell. From the Department to oSTEM and programs associated with the Office of Inclusion and

Student Engagement. I literally could not have done this without your support, love, and friendship. I have a special shoutout to my cohort-mates, Dr. Matt Boucher, Dr. Natalie Bray, Dr. Silas Bossert, and Dr. Talya Shragai for accompanying me on this weird journey of graduate school!

Most importantly, I want to thank my family. My mom and grandma have been my constant supporters throughout my entire life. It was my mom in particular where I gained my love of education watching her pursue her college education before her health prevented her from completing it. It is for her that I started down this journey.

And lastly my amazing partner Ryan, I thank him for all his love and for putting up with me. I could not ask for a better person to go through life with, or spend almost the entire day with during the past eight months while cooped up inside a house during a pandemic.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

AMONG-MICROBE INTERACTIONS AND THEIR EFFECT ON ANIMAL HOSTS

Summary

Animal gut-associated microorganisms influence animal performance and fitness. Although microbiome research has predominately concerned the effect of individual microorganisms on animal traits, there is growing interest in the role of among-microbe interactions in these processes. Taking advantage of simple animal models, such as *Drosophila melanogaster*, can provide an experimental framework to address how microbiomes shape animal health and fitness.

Introduction

The animal gut is generally colonized by a community of microorganisms, referred to as the microbiome. These taxa are functionally diverse and are able to influence many animal traits, including nutrition, performance, immune system, behavior, and fitness (Herp et al., 2019; Huang et al., 2015; Karasov and Douglas, 2013; McFall-Ngai et al., 2013; Morais et al., 2020; Nicholson et al., 2012; Sommer and Bäckhed, 2013; Thaiss et al., 2016; Wong et al., 2016). Most interactions between animals and the gut microbiome have a metabolic basis, and can be defined through direct and indirect effects on animal nutritional physiology and metabolism (Ankrah and Douglas, 2018; Douglas, 2014; Engel and Moran, 2013; Hooper et al., 2002). Microbes are able to contribute directly to animal metabolism through digestion of complex substrates that are inaccessible to the host, synthesis of essential nutrients,

and recycling of metabolic waste and toxic metabolites. Indirect effects are observed through metabolite effectors that modulate animal metabolic signaling pathways regulating nutrient allocation and metabolism.

To identify genetic and metabolic determinants of microbiome-dependent animal traits, studies have relied heavily on the application of metagenomics to conventional animals, i.e. with the naturally-occurring microbiome (Armour et al., 2019; Fromont et al., 2019; Kau et al., 2011) or the study of mono-associations, i.e. experimental associations with single microbial taxa (Chaston et al., 2014; Consuegra et al., 2020b; Judd et al., 2018; Kešnerová et al., 2017). Both approaches have severe limitations. Metagenomics identifies the genetic capacity for microbial function, potentially over-estimating the realized function. The use of mono-associations precludes analysis of the effect of among-microbe interactions, which are increasing being recognized as important drivers of animal traits (Consuegra et al., 2020a; Coyte and Rakoff-Nahoum, 2019; Gould et al., 2018; Granato et al., 2019). The gut microbiome of many animals, including humans, is complex with high taxonomic diversity and includes many taxa that are not currently amenable to culture under laboratory conditions (e.g. human gut). For these reasons, simple animal models with a microbiome that is naturally of low diversity and dominated by culturable taxa play a vital role in research on microbiome-host interactions (Douglas, 2019). One simple model that is gaining increasing traction is *Drosophila melanogaster*, and this system is introduced next.

Drosophila: ***a model of animal-microbiome interactions***

Drosophila melanogaster (Diptera: Drosophilidae) has been used as a biological model for over a century because of its fast generation time (< two weeks), suitability for laboratory culture, and increasingly, its superb genetic and genomic resources (Markow, 2015). In addition, the *Drosophila* gut microbiome is a fast-emerging system for investigating animal microbiomes due to its low taxonomic diversity and most of the microbiome members are culturable in the laboratory (Broderick and Lemaitre, 2012; Douglas, 2019; Erkosar et al., 2013; Ludington and Ja, 2020). The community is dominated by bacteria of the Acetobacteraceae (Alpha-Proteobacteria), Lactobacillales (Firmicutes), and Enterobacterales (Gamma-Proteobacteria), as well as yeasts of the Saccharomycetales (Ascomycota) (Adair et al., 2018; Bost et al., 2018; Chandler et al., 2012, 2011; Chaston et al., 2016; Cox and Gilmore, 2007; Quan and Eisen, 2018; Staubach et al., 2013; Wang et al., 2020; Wong et al., 2013). Acetobacteraceae and Lactobacillales bacteria are detected in most field-collected and laboratory cultures of *Drosophila*, while the incidence of Enterobacterales tends to be variable. The yeasts, although largely recognized as an important dietary component of flies, are not prevalent in laboratory lines, largely due to the addition of anti-fungal preservatives in diet (Broderick and Lemaitre, 2012).

In nature, the adult flies are attracted to microbial-derived fermentation products associated with rotting fruits, and they use the fruit substrate as a food source, courtship and mating location and, for females, an oviposition site (Becher et al., 2012). Eggs deposited on these substrates hatch and feed on microorganisms and degraded fruit tissue (Reaume and Sokolowski, 2006). *Drosophila* alter the microbial

community composition, at least under laboratory conditions, and are believed to promote microbial taxa that are beneficial for larval growth and development (Buser et al., 2014; Wong et al., 2015). Although the gut microbiome is known to be acquired by feeding on microorganisms associated with the food, and that the presence and composition of these microorganisms influence *D. melanogaster* fitness, development, and behavior, there still remain unanswered questions about the underlying processes.

Among-microbe interactions in the gut microbiome and the influence of host traits

One of the largest challenges in microbial ecology relates to the difficulties in observing how microorganisms interact with each other in real time, and therefore many studies have focused on descriptive methods to understand microbiome structure and function (Widder et al., 2016). Interactions in microbial communities are varied and complex, ranging from antagonistic to beneficial. Generally, interactions between microorganisms are described in terms of how co-associations change the abundance of individual taxa. Culture-independent methods can be used to identify and quantify microbial taxa, making use of taxonomically-informative sequences, specially the 16S rRNA gene (Adair et al., 2018; Ellegaard and Engel, 2016). There are several limitations to this approach, especially as an index of microbial function, including lack of congruence between taxon identity and functional capacity as a result of strain-level variation (Martiny, 2015), redundancy in some functions between phylogenetically-distant microbial taxa (Lozupone et al., 2012), and horizontal transfer of genes between closely and distantly related taxa (Hall et al., 2020). Community composition analyses in conjunction with imaging techniques to define

spatial organization of microorganisms (i.e. fluorescence *in situ* hybridization [FISH]) (Mark et al., 2016), functional characterization using metagenomics (functional capacity), metatranscriptomics or metaproteomics (molecular function), or metabolomics (metabolic function) (Chaston et al., 2014; Douglas, 2018), genome-scale metabolic modeling to simulate community metabolic interactions (Brunner and Chia, 2019; Mendoza et al., 2019), and community network analyses (Coyte et al., 2015) can address function more directly than 16S analyses.

As techniques focused on function are applied, it is becoming apparent that certain functions may only be realized in a community context. When microorganisms are cultured as a community, they often produce compounds that are not detected by any of the microbial taxa cultured in isolation (Wintermute and Silver, 2010). This can result from cross-feeding of metabolites between different taxa; polymicrobial functions arise from the exchange of microbial-derived metabolites that are not available in the environment and are generally by-products of metabolism or the extracellular degradation of complex substrates (Douglas, 2020; D'Souza et al., 2018; Freilich et al., 2011; Smith et al., 2019). For example, *Bacteriodes* spp. in the human gut microbiome can degrade complex carbohydrates and provide simple sugars that are used by different taxa for growth and the production of host-beneficial metabolites (Mahowald et al., 2009; Rakoff-nahoum et al., 2016; Rodriguez-Castaño et al., 2019). Similarly in the honey bee gut, cross-feeding between *Snodgrassella*, *Gilliamella*, and *Bifidobacterium* spp. has been documented (Kešnerová et al., 2017; Zheng et al., 2019).

For *Drosophila,* information is available on interactions between *Acetobacter* and

Lactobacillus spp., with evidence that *Acetobacter* and *Lactobacillus* spp. display positive and negative interactions in co-culture and in the presence of *Drosophila* (Newell and Douglas, 2014; Wong et al., 2015). However, we have limited knowledge of how they interact and their spatial organization on dietary substrates and in the *Drosophila* gut.

Evidence for competition between microorganisms and the *Drosophila* host over dietary resources comes from evidence that *Acetobacter* can consume glucose in the *Drosophila* diet, resulting in reduced *Drosophila* lipid storage (Chaston et al., 2014; Huang and Douglas, 2015). This effect can be compounded by positive interactions between *Acetobacter* and *Lactobacillus*. Specifically, the lactic acid produced by *Lactobacillus* is consumed by *Acetobacter* to expand metabolic capacity of the two microorganisms and alter their effect on *Drosophila* performance and nutrient allocation (Consuegra et al., 2020a; Sommer and Newell, 2019). In addition, the production of acetic acid, acetate esters, acetaldehyde derivatives by *Acetobacter* are dependent on cross-feeding of yeast-derived ethanol and are preferred by *Drosophila* for feeding and oviposition (Fischer et al., 2017). Acetic acid in particular is highlighted for its effects on *Drosophila* reproductive behavior and egg deposition rate, stimulating insulin-like signaling and influencing patterns of nutrient allocation, and activation of the IMD pathway (Joseph et al., 2009; Kamareddine et al., 2018; Kim et al., 2017; Shin et al., 2011).

Research objectives

The goal of this dissertation was to determine the impact of gut microorganisms on nutrient allocation and metabolic function in *Drosophila*, and to investigate the underlying processes. This was done through three separate investigations: first, to apply genome sequencing to define the metabolic capacity of individual bacteria (Chapter 2); second, to conduct an experimental analysis of the impact of co-associated microorganisms on patterns of nutrient allocation (Chapter 3); and, finally, to quantify the effect of microorganisms on the metabolome of *Drosophila* (Chapter 4). In particular for chapter 2, a comparative genomic approach was implemented to address the relationship between taxonomy and function of bacteria associated with *Drosophila* to identify functional capacity of individual bacteria. Chapter 3 concerned a combinatorial experimental approach in parallel with a metabolite feeding assay to determine whether community metabolism is present in the *Drosophila* gut microbiome and the metabolic basis for among-microbe interactions to affect host nutrient allocation. Finally, Chapter 4 used a comparative metabolomics approach to determine how co-associations shaped *Drosophila* metabolic function.

References

Adair, K.L., Wilson, M., Bost, A., Douglas, A.E., 2018. Microbial community assembly in wild populations of the fruit fly *Drosophila melanogaster*. ISME Journal 12, 959–972. https://doi.org/10.1038/s41396-017-0020-x

Ankrah, N.Y.D., Douglas, A.E., 2018. Nutrient factories: metabolic function of beneficial microorganisms associated with insects. Environmental Microbiology 20, 2002–2011. https://doi.org/10.1111/1462-2920.14097

Armour, C.R., Nayfach, S., Pollard, K.S., Sharpton, T.J., 2019. A Metagenomic Meta-analysis Reveals Functional Signatures of Health and Disease in the Human Gut Microbiome. mSystems 4. https://doi.org/10.1128/mSystems.00332-18

Becher, P.G., Flick, G., Rozpedowska, E., Schmidt, A., Hagman, A., Lebreton, S., Larsson, M.C., Hansson, B.S., Piškur, J., Witzgall, P., Bengtsson, M., 2012. Yeast, not fruit volatiles mediate *Drosophila melanogaster* attraction, oviposition and development. Functional Ecology 26, 822–828. https://doi.org/10.1111/j.1365-2435.2012.02006.x

Bost, A., Franzenburg, S., Adair, K.L., Martinson, V.G., Loeb, G., Douglas, A.E., 2018. How gut transcriptional function of *Drosophila melanogaster* varies with the presence and composition of the gut microbiota. Molecular Ecology 27, 1848–1859. https://doi.org/10.1111/mec.14413

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

Brunner, J.D., Chia, N., 2019. Metabolite-mediated modelling of microbial community dynamics captures emergent behaviour more effectively than species–species modelling. Journal of The Royal Society Interface 16, 20190423. https://doi.org/10.1098/rsif.2019.0423

Buser, C.C., Newcomb, R.D., Gaskett, A.C., Goddard, M.R., 2014. Niche construction initiates the evolution of mutualistic interactions. Ecology Letters 17, 1257–1264. https://doi.org/10.1111/ele.12331

Chandler, J.A., Eisen, J.A., Kopp, A., 2012. Yeast Communities of Diverse *Drosophila* Species : Comparison of Two Symbiont Groups in the Same Hosts 78, 7327–7336. https://doi.org/10.1128/AEM.01741-12

Chandler, J.A., Lang, J.M., Bhatnagar, S., Eisen, J.A., 2011. Bacterial Communities of Diverse *Drosophila* Species: Ecological Context of a Host – Microbe Model System 7. https://doi.org/10.1371/journal.pgen.1002272

Chaston, J.M., Dobson, A.J., Newell, P.D., Douglas, A.E., 2016. Host genetic control of the microbiota mediates the *Drosophila* Nutritional Phenotype. Applied and Environmental Microbiology 82, 671–679. https://doi.org/10.1128/AEM.03301-15.Editor

Chaston, J.M., Newell, P.D., Douglas, A.E., 2014. Metagenome-wide association of microbial determinants of host phenotype in *Drosophila melanogaster*. mBio 5, 1–12. https://doi.org/10.1128/mBio.01631-14

Consuegra, J., Grenier, T., Akherraz, H., Rahioui, I., Gervais, H., da Silva, P., Leulier, F., 2020a. Metabolic cooperation among commensal bacteria supports *Drosophila* juvenile growth under nutritional stress. iScience 101232.

https://doi.org/10.1016/j.isci.2020.101232
Consuegra, J., Grenier, T., Baa-Puyoulet, P., Rahioui, I., Akherraz, H., Gervais, H., Parisot, N., Silva, P. da, Charles, H., Calevro, F., Leulier, F., 2020b. *Drosophila*-associated bacteria differentially shape the nutritional requirements of their host during juvenile growth. PLOS Biology 18, e3000681. https://doi.org/10.1371/journal.pbio.3000681
Cox, C.R., Gilmore, M.S., 2007. Native Microbial Colonization of *Drosophila melanogaster* and Its Use as a Model of *Enterococcus faecalis* Pathogenesis. Infect Immun 75, 1565–1576. https://doi.org/10.1128/IAI.01496-06
Coyte, K.Z., Rakoff-Nahoum, S., 2019. Understanding Competition and Cooperation within the Mammalian Gut Microbiome. Current Biology 29, R538–R544. https://doi.org/10.1016/j.cub.2019.04.017
Coyte, K.Z., Schluter, J., Foster, K.R., 2015. The ecology of the microbiome: Networks, competition, and stability. Science 350, 663–666. https://doi.org/10.1126/science.aad2602
Douglas, A.E., 2020. The microbial exometabolome: ecological resource and architect of microbial communities. Philosophical Transactions of the Royal Society B: Biological Sciences 375, 20190250–20190250. https://doi.org/10.1098/rstb.2019.0250
Douglas, A.E., 2019. Simple animal models for microbiome research. Nature Reviews Microbiology 17, 764–775. https://doi.org/10.1038/s41579-019-0242-1
Douglas, A.E., 2018. Omics and the metabolic function of insect–microbial symbioses. Current Opinion in Insect Science, Global change biology * Molecular physiology 29, 1–6. https://doi.org/10.1016/j.cois.2018.05.012
Douglas, A.E., 2014. Symbiosis as a General Principle in Eukaryotic Evolution. Cold Spring Harb Perspect Biol 6, a016113. https://doi.org/10.1101/cshperspect.a016113
D'Souza, G., Shitut, S., Preussger, D., Yousif, G., Waschina, S., Kost, C., 2018. Ecology and evolution of metabolic cross-feeding interactions in bacteria. Nat. Prod. Rep. 35, 455–488. https://doi.org/10.1039/C8NP00009C
Ellegaard, K.M., Engel, P., 2016. Beyond 16S rRNA Community Profiling: Intra-Species Diversity in the Gut Microbiota. Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.01475
Engel, P., Moran, N.A., 2013. The gut microbiota of insects - diversity in structure and function. FEMS Microbiology Reviews 37, 699–735. https://doi.org/10.1111/1574-6976.12025
Erkosar, B., Storelli, G., Defaye, A., Leulier, F., 2013. Host-intestinal microbiota mutualism: "learning on the fly." Cell Host and Microbe 13, 8–14. https://doi.org/10.1016/j.chom.2012.12.004
Fischer, C., Trautman, E.P., Crawford, J.M., Stabb, E. V., Handelsman, J., Broderick, N.A., 2017. Metabolite exchange between microbiome members produces compounds that influence drosophila behavior. eLife 6. https://doi.org/10.7554/eLife.18855
Freilich, S., Zarecki, R., Eilam, O., Segal, E.S., Henry, C.S., Kupiec, M., Gophna, U., Sharan, R., Ruppin, E., 2011. Competitive and cooperative metabolic

interactions in bacterial communities. Nature Communications 2, 589–589. https://doi.org/10.1038/ncomms1597
Fromont, C., Adair, K.L., Douglas, A.E., 2019. Correlation and causation between the microbiome, *Wolbachia* and host functional traits in natural populations of *Drosophilid* flies. Mol. Ecol. 28, 1826–1841. https://doi.org/10.1111/mec.15041
Gould, A.L., Zhang, V., Lamberti, L., Jones, E.W., Obadia, B., Korasidis, N., Gavryushkin, A., Carlson, J.M., Beerenwinkel, N., Ludington, W.B., 2018. Microbiome interactions shape host fitness. Proceedings of the National Academy of Sciences of the United States of America 115, E11951–E11960. https://doi.org/10.1073/pnas.1809349115
Granato, E.T., Meiller-Legrand, T.A., Foster, K.R., 2019. The Evolution and Ecology of Bacterial Warfare. Current Biology 29, R521–R537. https://doi.org/10.1016/j.cub.2019.04.024
Hall, R.J., Whelan, F.J., McInerney, J.O., Ou, Y., Domingo-Sananes, M.R., 2020. Horizontal Gene Transfer as a Source of Conflict and Cooperation in Prokaryotes. Front Microbiol 11. https://doi.org/10.3389/fmicb.2020.01569
Herp, S., Brugiroux, S., Garzetti, D., Ring, D., Jochum, L.M., Beutler, M., Eberl, C., Hussain, S., Walter, S., Gerlach, R.G., Ruscheweyh, H.J., Huson, D., Sellin, M.E., Slack, E., Hanson, B., Loy, A., Baines, J.F., Rausch, P., Basic, M., Bleich, A., Berry, D., Stecher, B., 2019. *Mucispirillum schaedleri* Antagonizes *Salmonella* Virulence to Protect Mice against Colitis. Cell Host and Microbe 25, 681-694.e8. https://doi.org/10.1016/j.chom.2019.03.004
Hooper, L.V., Midtvedt, T., Gordon, J.I., 2002. How Host-Microbial Interactions Shape the Nutrient Environment of the Mammalian Intestine. Annual Review of Nutrition 22, 283–307. https://doi.org/10.1146/annurev.nutr.22.011602.092259
Huang, J., Douglas, A.E., 2015. Consumption of dietary sugar by gut bacteria determines *Drosophila* lipid content. Biology Letters 11, 20150469.
Huang, J.-H., Jing, X., Douglas, A.E., 2015. The multi-tasking gut epithelium of insects. Insect Biochemistry and Molecular Biology 1–6. https://doi.org/10.1016/j.ibmb.2015.05.004
Joseph, R.M., Devineni, A. V, King, I.F.G., Heberlein, U., 2009. Oviposition preference for and positional avoidance of acetic acid provide a model for competing behavioral drives in *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America 106, 11352–11357. https://doi.org/10.1073/pnas.0901419106
Judd, A.M., Matthews, M.K., Hughes, R., Veloz, M., Sexton, C.E., Chaston, J.M., 2018. Bacterial Methionine Metabolism Genes Influence *Drosophila melanogaster* Starvation Resistance. Appl Environ Microbiol 84, e00662-18, /aem/84/17/e00662-18.atom. https://doi.org/10.1128/AEM.00662-18
Kamareddine, L., Robins, W.P., Berkey, C.D., Mekalanos, J.J., Watnick, P.I., 2018. The *Drosophila* Immune Deficiency Pathway Modulates Enteroendocrine Function and Host Metabolism. Cell Metabolism 28, 449-462.e5. https://doi.org/10.1016/j.cmet.2018.05.026

Karasov, W.H., Douglas, A.E., 2013. Comparative digestive physiology. Comprehensive Physiology 3, 741–783. https://doi.org/10.1002/cphy.c110054
Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L., Gordon, J.I., 2011. Human nutrition, the gut microbiome and the immune system. Nature 474, 327–336. https://doi.org/10.1038/nature10213
Kešnerová, L., Mars, R.A.T., Ellegaard, K.M., Troilo, M., Sauer, U., Engel, P., 2017. Disentangling metabolic functions of bacteria in the honey bee gut. PLoS Biology 15, 1–28. https://doi.org/10.1371/journal.pbio.2003467
Kim, G., Hsin, J., Mcmullen, J.G., Newell, P.D., Douglas, A.E., 2017. Physiological responses of insects to microbial fermentation products: Insights from the interactions between *Drosophila* and acetic acid. Journal of Insect Physiology 0–1. https://doi.org/10.1016/j.jinsphys.2017.05.005
Lozupone, C., Stombaugh, J., Gordon, J.I., Jansson, J., Knight, R., 2012. Diversity, stability and resilience of the human gut microbiota. Nature 489, 220–230.
Ludington, W.B., Ja, W.W., 2020. *Drosophila* as a model for the gut microbiome. PLoS Pathog 16. https://doi.org/10.1371/journal.ppat.1008398
Mahowald, M.A., Rey, F.E., Seedorf, H., Turnbaugh, P.J., Fulton, R.S., Wollam, A., Shah, N., Wang, C., Magrini, V., Wilson, R.K., Cantarel, B.L., Coutinho, P.M., Henrissat, B., Crock, L.W., Russell, A., Verberkmoes, N.C., Hettich, R.L., Gordon, J.I., 2009. Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. Proceedings of the National Academy of Sciences 106, 5859–5864. https://doi.org/10.1073/pnas.0901529106
Mark, J.L., Rossetti, B.J., Rieken, C.W., Dewhirst, F.E., Borisy, G.G., 2016. Biogeography of a human oral microbiome at the micron scale 791–800. https://doi.org/10.1073/pnas.1522149113
Markow, T.A., 2015. The secret lives of *Drosophila* flies. eLife 4, 1–9. https://doi.org/10.7554/eLife.06793
Martiny, 2015. Microbiomes in light of traits: A phylogenetic perspective 350. https://doi.org/10.1126/science.aac9323
McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H.V., Domazet-Lošo, T., Douglas, A.E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S.F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A.H., Kremer, N., Mazmanian, S.K., Metcalf, J.L., Nealson, K., Pierce, N.E., Rawls, J.F., Reid, A., Ruby, E.G., Rumpho, M., Sanders, J.G., Tautz, D., 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, 3229–36. https://doi.org/10.1073/pnas.1218525110
Mendoza, S.N., Olivier, B.G., Molenaar, D., Teusink, B., 2019. A systematic assessment of current genome-scale metabolic reconstruction tools. Genome Biology 20, 158. https://doi.org/10.1186/s13059-019-1769-1
Morais, L.H., Schreiber, H.L., Mazmanian, S.K., 2020. The gut microbiota–brain axis in behaviour and brain disorders. Nature Reviews Microbiology 1–15. https://doi.org/10.1038/s41579-020-00460-0
Newell, P.D., Douglas, A.E., 2014. Interspecies Interactions Determine the Impact of

the Gut Microbiota on Nutrient Allocation in *Drosophila melanogaster*. Applied and Environmental Microbiology 80, 788–796. https://doi.org/10.1128/AEM.02742-13

Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., Pettersson, S., 2012. Host-Gut Microbiota Metabolic Interactions. Science 336, 1262–1267. https://doi.org/10.1126/science.1223813

Quan, A.S., Eisen, M.B., 2018. The ecology of the *Drosophila*-yeast mutualism in wineries. PLoS ONE 13, 1–25. https://doi.org/10.1371/journal.pone.0196440

Rakoff-nahoum, S., Foster, K.R., Comstock, L.E., 2016. The evolution of cooperation withing the gut microbiota. Nature 533, 255–259. https://doi.org/10.1038/nature17626

Reaume, C.J., Sokolowski, M.B., 2006. The nature of *Drosophila melanogaster*. Current Biology 16, 623–628. https://doi.org/10.1016/j.cub.2006.07.042

Rodriguez-Castaño, G.P., Dorris, M.R., Liu, X., Bolling, B.W., Acosta-Gonzalez, A., Rey, F.E., 2019. *Bacteroides thetaiotaomicron* starch utilization promotes quercetin degradation and butyrate production by *Eubacterium ramulus*. Frontiers in Microbiology 10, 1–8. https://doi.org/10.3389/fmicb.2019.01145

Shin, S.C., Kim, S.-H., You, H., Kim, B., Kim, A.C., Lee, K.-A., Yoon, J.-H., Ryu, J.-H., Lee, W.-J., 2011. *Drosophila* microbiome modulates host developmental and metabolic homeostasis via insulin signaling. Science 334, 670–4. https://doi.org/10.1126/science.1212782

Smith, N.W., Shorten, P.R., Altermann, E., Roy, N.C., McNabb, W.C., 2019. The Classification and Evolution of Bacterial Cross-Feeding. Front. Ecol. Evol. 7. https://doi.org/10.3389/fevo.2019.00153

Sommer, A.J., Newell, P.D., 2019. Metabolic Basis for Mutualism between Gut Bacteria and Its Impact on the *Drosophila melanogaster* Host. Applied and environmental microbiology 85, 1–15. https://doi.org/10.1128/AEM.01882-18

Sommer, F., Bäckhed, F., 2013. The gut microbiota-masters of host development and physiology. Nature Reviews Microbiology 11, 227–238. https://doi.org/10.1038/nrmicro2974

Staubach, F., Baines, J.F., Künzel, S., Bik, E.M., Petrov, D. a., 2013. Host Species and Environmental Effects on Bacterial Communities Associated with *Drosophila* in the Laboratory and in the Natural Environment. PLoS ONE 8, e70749–e70749. https://doi.org/10.1371/journal.pone.0070749

Thaiss, C.A., Zmora, N., Levy, M., Elinav, E., 2016. The microbiome and innate immunity. Nature 535, 65–74. https://doi.org/10.1038/nature18847

Wang, Y., Kapun, M., Waidele, L., Kuenzel, S., Bergland, A.O., Staubach, F., 2020. Common structuring principles of the *Drosophila melanogaster* microbiome on a continental scale and between host and substrate. Environmental Microbiology Reports 12, 220–228. https://doi.org/10.1111/1758-2229.12826

Widder, S., Allen, R.J., Pfeiffer, T., Curtis, T.P., Wiuf, C., Sloan, W.T., Cordero, O.X., Brown, S.P., Momeni, B., Shou, W., Kettle, H., Flint, H.J., Haas, A.F., Laroche, B., Kreft, J., 2016. Challenges in microbial ecology: building predictive understanding of community function and dynamics. 1–12. https://doi.org/10.1038/ismej.2016.45

Wintermute, E.H., Silver, P.A., 2010. Emergent cooperation in microbial metabolism. Molecular Systems Biology 6, 1–7. https://doi.org/10.1038/msb.2010.66

Wong, A.C.-N., Chaston, J.M., Douglas, A.E., 2013. The inconstant gut microbiota of *Drosophila* species revealed by 16S rRNA gene analysis. The ISME journal 7, 1922–32. https://doi.org/10.1038/ismej.2013.86

Wong, A.C.-N., 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

Wong, A.C.N., Vanhove, A.S., Watnick, P.I., 2016. The interplay between intestinal bacteria and host metabolism in health and disease: lessons from *Drosophila melanogaster*. Disease Models & Mechanisms 9. https://doi.org/10.1242/dmm.023408

Zheng, H., Perreau, J., Elijah Powell, J., Han, B., Zhang, Z., Kwong, W.K., Tringe, S.G., Moran, N.A., 2019. Division of labor in honey bee gut microbiota for plant polysaccharide digestion. Proceedings of the National Academy of Sciences of the United States of America 116, 25909–25916. https://doi.org/10.1073/pnas.1916224116

CHAPTER 2

GENOME-INFERRED CORRESPONDENCE BETWEEN PHYLOGENY AND METABOLIC TRAITS IN THE WILD *DROSOPHILA* GUT MICROBIOME[1]

Abstract

Annotated genome sequences provide valuable insight into the functional capabilities of members of microbial communities. Although this approach is applied extensively to infer function of the microbiomes in animal guts, most studies use metagenomic data, hampering the assignment of genes to specific microbial taxa. Here, we make use of the readily culturable bacterial communities in the gut of the fruit fly *Drosophila melanogaster* to obtain draft genome sequences for 96 isolates from wild flies. These include 81 new *de novo* assembled genomes, assigned to three order (Enterobacterales, Lactobacillales, and Rhodospirillales) with 80% of strains identified to species using average nucleotide identity and phylogenomic reconstruction. Based on annotations by the RAST pipeline, among-isolate variation in metabolic function partitioned strongly by bacterial order, particularly by amino acid metabolism (Rhodospirillales), fermentation and nucleotide metabolism (Lactobacillales) and arginine, urea and polyamine metabolism (Enterobacterales). Seven bacterial species, comprising 2-3 species in each order, were well-represented among the isolates and included ≥ 5 strains, permitting analysis of metabolic functions in the accessory genome (i.e. genes

[1] Article in preparation for journal submission by McMullen, J.G., Bueno, E., Blow, F., and Douglas, A.E.
Supplemental material is found in Appendix A.
Bueno, E. contributed to fly collections, microorganism isolations, and 16S rRNA gene amplicon sequencing and Blow, F. contributed to the genome assembly pipeline.

not present in every strain). Overall, the metabolic function in the accessory genome partitioned by bacterial order. Two species, *Gluconobacter cerinus* (Rhodospirillales) and *Lactobacillus plantarum* (Lactobacillales) had large accessory genomes, and metabolic functions were dominated by amino acid metabolism (*G. cerinus*) and carbohydrate metabolism (*L. plantarum*). The patterns of variation in metabolic capabilities at multiple phylogenetic scales provide the basis for future studies of the ecological and evolutionary processes shaping the diversity of microorganisms associated with natural populations of *Drosophila*.

Introduction

Animal gut microbiomes are complex assemblages of microorganisms which mediate diverse functions that impact host physiology, behavior, and fitness (Nicholson et al. 2012; Read & Holmes 2017; Rolhion & Chassaing 2016; Sommer & Bäckhed 2013; Thaiss et al. 2016; Huang et al. 2015; Qiao et al. 2019; Turkiewicz et al. 2019). Most interactions between the microbiome and the animal host are based on the metabolic capabilities of microbiome members, with traits ranging from degradation and fermentation of host-inaccessible substrates to synthesis of key nutrients for the host, detoxification of harmful dietary constituents and recycling of metabolic waste products, and effects on host signaling pathways (Ankrah & Douglas 2018; Hooper et al. 2002; Engel & Moran 2013). Investigation of the relationship between the traits and taxonomic identity among gut microorganisms has shown that many metabolic traits are functionally redundant and can be shared by closely and distantly related microbiome members (Louca et al. 2018; Heintz-Buschart & Wilmes

2018). This finding is largely based on metagenomic studies, where the taxonomic composition of the microbiome is uncontrolled and variable (Lozupone et al. 2012; Huttenhower et al. 2012).

Functional redundancy can ensure sustained function (also known as ecosystem resilience) of the gut microbiome during perturbations that reduce the abundance or function of specific taxa and alter the overall microbiome composition (Heintz-Buschart & Wilmes 2018; Allison & Martiny 2008). Evolutionary changes, which can occur within ecological timeframes, can also affect the relationship between taxonomy and function. In particular, phylogenetically-divergent taxa may share a metabolic trait by gain of function through horizontal gene transfer (HGT), and closely-related taxa may differ in functional traits by differential gene deletions and by functional divergence of a recently-duplicated gene (Louca et al. 2018). Two examples illustrate these processes. The first is the bile salt hydrolase gene, which is involved in lipid homeostasis and antimicrobial effects. This gene is widespread across bacterial taxa in the human microbiome (most prevalent among the Firmicutes) with evidence of HGT events among different *Lactobacillus* spp. and *Listeria monocytogenes* (Chand et al. 2017; Jones et al. 2008; Kumar et al. 2012). Secondly, in the honey bee gut microbiome, the distribution of a glucoside hydrolase gene family (genes involved in degradation of hemicellulose in pollen) in *Bifidobacterium* spp. is the result of gene duplication and deletion events (Zheng et al. 2019).

The apparent ubiquity of functional redundancy, however, is open to question. Functional composition analyses often rely on broad metabolic annotations that can encompass multiple pathways (Langille 2018). These methods can fail to detect

biologically important differences in metabolic function of gene families, as demonstrated, for example, in Proteobacteria of the human gut microbiome (Bradley & Pollard 2017). Compounding these problems, within-species variation in metabolic function can be widespread, such that metabolic traits important to the host are displayed by only a subset of strains or are mediated by pathways distributed across two or more different strains (Douglas 2020). For example, *Bifidobacterium longum*, a member of the microbiome of the human infant, has a large accessory genome with variable incidence of genes involved in transport and degradation of human milk oligosaccharides, implicating some, but not all, strains of this species as important to human milk metabolism (Vatanen et al. 2019). Intraspecific variation requires identification of not only the pangenome (i.e. total genetic capabilities) of a species, but also how the functional traits are distributed across different strains (Tettelin et al. 2005; Brockhurst et al. 2019; Van Rossum et al. 2020).

The goal of this study was to investigate how primary metabolism functions of a gut microbiome map onto bacterial phylogeny. We used the gut microbiome of *Drosophila melanogaster* for this analysis because, unlike the microbiome of many animals, most of the *Drosophila*-associated bacteria are readily culturable (Douglas 2019). Relative to metagenome-assembled genomes, genome sequences of the individual bacterial isolates enable higher quality assembly and increased resolution of phylogenomic patterns (Van Rossum et al. 2020). More generally, *Drosophila* is a fast-emerging system to investigate ecological and evolutionary questions regarding animal-associated microbiomes (Broderick & Lemaitre 2012; Douglas 2019; Erkosar et al. 2013; Wong et al. 2016) and there are indications that, as for the mammalian gut

microbiome, the *Drosophila* metagenome displays incongruence between functional traits and taxonomic composition (Newell et al. 2014; Petkau et al. 2016; Adair et al. 2018; Consuegra et al. 2020; Kang & Douglas 2020). However, the relationship between taxonomy and distribution of traits has not been robustly tested.

For our analysis, we focused on bacterial taxa isolated from natural populations of *Drosophila*. The gut microbiome of wild *Drosophila* is dominated by members of the bacterial orders Enterobacterales, Lactobacillales, and Rhodospirillales, although the relative abundance of the different taxa varies among individuals and collections (Chandler et al. 2011; Adair et al. 2018; Bost et al. 2018; Kang & Douglas 2020; Wang et al. 2020; Walters et al. 2020). Long-term laboratory cultures of *Drosophila* were not used because their gut microbiome is of low diversity (Cox & Gilmore 2007; Wong et al. 2013; Obadia et al. 2018) and can be functionally different from wild populations (Winans et al. 2017; Bost et al. 2018). The great majority of published studies on the genome sequences of *Drosophila* gut microorganisms have concerned bacterial taxa derived from laboratory lines (e.g. Winans et al. 2017; Newell et al. 2014; Petkau et al. 2016) with few sequences available from field-isolates (Table 1.1). Therefore, this study was initiated by the isolation of bacteria from field-collected *Drosophila*. In total, we isolated and sequenced the genomes of 81 bacterial strains associated with wild *Drosophila*. We performed comparisons of metabolic traits among all field-isolated strains, and then examined the metabolic pangenomes of prevalent species to assess the scale of within-species variation. For this panel of bacteria, the three bacterial orders were strongly differentiated by primary metabolic function, and a subset of species also displayed strain level variation in metabolism

related genes. The taxonomically-variable traits include functions likely to be adaptive for utilization of sugar-rich rotting fruit environment and predicted to influence *Drosophila* physiology and performance.

Table 1.1. Bacterial strains used in comparative genomics analyses. Prevalent species (detected in 4 or more flies and represented by >4 strains in our dataset) used for pangenome analyses are in bold.

Order	Family	Genus	Species (strain ID)	No. strains sequenced (no. flies)	Publicly available strains
Enterobacterales	Enterobacteriaceae	*Citrobacter*	sp. (C)	1 (1)	
		Enterobacter	*asburiae* (Ea)	1 (1)	
			ludwigii (El)	1 (1)	
			mori (Em)	1 (1)	
			sp. (E)	1 (1)	
		Klebsiella	*michiganensis* (Km)	1 (1)	
			variicola (Kv)	1 (1)	
	Erwiniaceae	*Pantoea*	*dispersa* (PAd)	2 (1)	
			sp. (PA)	1 (1)	
		Tatumella	**sp. #1 (T)**	**6 (6)**	
			sp. #2 (T)	1 (1)	
	Morganellaceae	*Providencia*	*alcalifaciens* (PRa)		1[a]
			burhodogranariea (PRb)		1[a]
			***rettgeri* (PRr)**	**4 (4)**	**1[a]**
			sneebia (PRs)		1[a]
			sp. (PR)	3 (3)	
	Yersiniaceae	*Nissabacter*	*archeti* (Na)	1 (1)	
		Serratia	*rubidaea* (Sr)	1 (1)	
Lactobacillales	Lactobacillaceae	***Lactobacillus***	***brevis* (LAb)**	**5 (5)**	
			paracasei (LApa)	1 (1)	1[b]
			***plantarum* (LApl)**	**5 (5)**	1[c]
		Leuconostoc	*citreum* (LEc)		1[d]
			mesenteroides (LEm)	1 (1)	

			pseudomesenteroides (LEp)	1 (1)	
			suionicum (LEs)	1 (1)	
		Weissella	*cibaria* (Wc)		1[e]
			minor (Wm)	1 (1)	
	Streptococcaceae	*Lactococcus*	*lactis* (Ll)		1[f]
Rhodospirillales	Acetobacteraceae	*Acetobacter*	*cibinongensis* (Ac)		1[g]
			indonesiensis (Ai)		1[g]
			okinawensis (Aok)	2 (1)	
			orientalis (Aor)		2[g]
			persici (Ap)	3 (2)	
			***thailandicus* (Ath)**	**4 (4)**	**1[g]**
			tropicalis (Atr)		1[g]
		Gluconobacter	*albidus* (Ga)	1 (1)	
			***cerinus* (Gc)**	**13 (5)**	
			japonicus (Gj)	1 (1)	
			***kondonii* (Gk)**	**6 (5)**	
			sp. #1 (G)	3 (2)	
			sp. #2 (G)	1 (1)	
			sphaericus (Gs)	3 (2)	
			wancherniae (Gw)	3 (1)	

[a]Galac and Lazzaro 2012; [b]Hammer et al. 2017; [c]Petkau et al. 2016; [d]Wright et al. 2017; [e]Ricks et al. 2017; [f]Chaston et al. 2014; [g]Winans et al. 2017

Materials and methods

Isolation of Drosophila*-associated bacteria*

Wild *D. melanogaster* flies were collected from compost bins or other food waste from five domestic kitchens in Ithaca, NY, USA and from a dumpster containing rotting fruits at the Cornell Orchards, Ithaca, NY, USA from 2015 to 2019 (see Table S1.1A for collection details). Flies were starved for 1-3 h to allow any food in the gut to be eliminated, and then anesthetized with CO_2 and sorted by sex (distinguished visually by genitalia morphology) and species to obtain *D. melanogaster* adults (no *Drosophila simulans* males were observed in collections). The flies were washed in sterile phosphate buffered-saline (PBS; Cold Spring Harbor, 2018), and hand-homogenized in 100 µl PBS (except 200 µl for 2019 Lactobacillales collections) with a disposable pestle (Kontes/Kimble-Chase, Vineland, NJ, USA) using aseptic technique. Each homogenate was inoculated onto an agar medium (yeast-peptone-dextrose [YPD] or modified De Man, Rogosa, and Sharpe [mMRS]) and incubated at 30°C for up to one week under aerobic or high CO_2 conditions by placing a lit candle in a glass jar (Fan & Li 1997) (Table S1.1A&B). YPD is a nutrient rich medium that supports the growth of sugar-rich environment microorganisms, while mMRS is a more selective medium that promotes the growth of acetic acid bacteria (Rhodospirillales) and Lactobacillales associated with *Drosophila*. In 2019, the procedure was modified to enhance the efficacy of isolating lactobacilli, which tend to have low relative abundance in wild fly guts (Adair et al. 2018; Chandler et al. 2011; Kang & Douglas 2020). Specifically, the homogenates were allowed to settle for 5-10 mins (allowing large microorganisms, e.g. yeasts, to settle) and 75 µl supernatant was

inoculated on agar plates. The mMRS medium was also supplemented with azide, tween-80, and bromocreosol purple (Table S1.1B) to select for Lactobacillales taxa (Choi et al. 2016). Individual colonies representative of different morphologies were isolated and streaked onto fresh agar (same medium as initial growth but lacking any antibiotics or dyes). A single representative colony was grown in broth of the same medium, visually confirmed as a bacterium by light microscopy (DM5000 B, Leica Microsystems, Buffalo Grove, IL, USA), and stored in 20% glycerol (Sigma, St. Louis, MO, USA) at -80°C.

DNA extraction of bacterial isolates

A chunk of frozen glycerol stock was inoculated either onto mMRS or YPD agar and a single colony was obtained to grow in 5 ml broth until turbid (see Table S1.1B for media). A 1 ml sample of the cell suspension was centrifuged at 19,000x*g* for 5 min and cells were re-suspended in 678 µl cell lysis buffer (108 mM Tris-HCl, pH 8.0; 1.5 M NaCl; 21.6 mM EDTA; Sigma) and 16 U proteinase K (Qiagen, Hilden, Germany) with either 30 µl 1 mm diameter glass beads (Scientific Industries, Bohemia, NY, USA) and 250 µl 2.3 mm diameter zirconia beads (BioSpec, Bartlesville, OK, USA) or 200 µl 1 mm diameter glass beads. Samples were homogenized for 35 s at 5.5 m/s with a FastPrep-24 instrument (MP Biomedicals, Santa Ana, CA, USA) and incubated at 56°C for 2 h. Homogenates were incubated overnight at 37°C with 35 U RNaseA (Qiagen). DNA was extracted from homogenate with 750 µl phenol:chloroform:isoamyl alcohol (25:24:1; Thermo Fisher Scientific, Waltham, MA, USA) and centrifuged at 19,000x*g* for 15 min at 4°C. To precipitate

DNA from 450 µl aqueous layer, 900 µl ethanol and 45 µl 3 M sodium acetate (pH 5.2; Sigma) were added to each sample and incubated overnight at -20°C. Following centrifugation at 19,000x*g* for 15 min at 4°C, DNA pellet was washed in 75% ethanol, centrifuged at 19,000x*g* for 10 min at 4°C, air-dried for 10 min, and re-suspended in 50 ul nuclease-free water (Ambion, Austin, TX, USA). DNA was stored at -20°C until PCR amplification and whole-genome sequencing.

Molecular identification of bacteria

Molecular characterization was first performed by Sanger sequencing of bacterial 16S rRNA gene amplicons obtained by PCR with the primers 16SA1 (forward: 5'-AGAGTTTGATCMTGGCTCAG-3') and 16SB1 (reverse: 5'-TACGGYTACCTTGTTACGACTT-3') from Fukatsu & Nikoh (1998). Approximately 1 μg DNA template (quantified using Nanodrop; Thermo Fisher Scientific) was added to 0.2 µM primers and 1 U OneTaq 2x Master Mix with Standard Buffer (New England BioLabs, Ipswich, MA, USA). PCR reaction conditions were 94°C for 30 seconds, 30 amplification cycles of 94°C for 30 s, 55.3°C for 60 s, and 68°C for 60 s with a final extension for 5 min at 68°C. PCR products were purified using ExoSAP-IT PCR Clean Up Reagent (Applied Biosystems, Waltham, MA, USA) and submitted for Sanger sequencing (both forward and reverse directions) at Cornell University Genomics Facility using Applied Biosystems 3730xl. Consensus sequences were generated from forward and reverse sequences and taxonomic identity was assigned using BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the NCBI non-redundant nucleotide

collection with Geneious Prime 2019.2.1 (Biomatters, Auckland, New Zealand). Bacterial isolates were selected for genome sequencing by maximizing taxonomic, fly replicate, and collection diversity within Enterobacterales, Lactobacillales, and Rhodospirillales.

Sequencing and genome assembly

Genomic DNA (0.2 ng/µl; quantified by Qubit 2.0 fluorimeter; Invitrogen, Waltham, MA, USA) was submitted to Cornell University Genomics Facility for whole-genome shotgun sequencing using an Illumina NextSeq500 Platform with the Nextera XL DNA Library Preparation kit (Illumina, San Diego, CA, USA) to generate 150 bp paired-end reads according to manufacturer's protocol. Libraries were pooled in equal proportions across three runs and their quality was assessed with a Fragment Analyzer (Advanced Analytical Technologies, Ames, IA, USA). A Blue Pippin device (Sage Science, Beverley, MA, USA) was used for further size-selection of pooled libraries to target fragments ≤ 800 bp, if required.

Between 1,033,730 and 28,432,172 reads were obtained for 81 bacterial genomes (Table S1.1A). Read quality was assessed using FastQC v0.11.3 (www.bioinformatics.babraham.ac.uk/) and were trimmed with trimmomatic v0.36 (Bolger et al. 2014). Reads were trimmed on the ends if the quality score was <3 or the terminal base was unidentified ('N'), and sequences were only retained if they had a quality score of ≥15 over a 4 bp moving window and length of 125 bp. Then, SPADES v3.11.1 (Bankevich et al. 2012) was used to assemble reads into contigs (k-mer lengths 21, 33, 55, and 77 were used) following default parameters. The careful

option was included for genome polishing. Low k-mer coverage contigs were filtered to reduce contamination following Douglass et al. (2019); see Table S1.1A for cutoffs applied to each genome. SSPACE v3.0 (Boetzer et al. 2011) was used for contig extension and scaffolding following default parameters with a minimum 100 bp (insert size was estimated from subsampling 1,000,000 reads). Genome assembly statistics were obtained using Quast v4.6.3 (Gurevich et al. 2013) with contigs less than 500 bp removed. To assess average sequence depth, reads were mapped to final contigs using Bowtie2 v2.2.6 (Langmead & Salzberg 2012) following default parameters and the samtools v0.1.19 (Li et al. 2009) depth function. ConEst16S (Lee et al. 2017) was used to identify bacterial contamination when more than one 16S rRNA gene was detected for a genome (Table S1.1A); none of the genomes were found to have bacterial contamination.

Genome annotation

Genomes were annotated using the RASTtk pipeline on RAST server with error correcting (Brettin et al. 2015; Overbeek et al. 2014). For analysis of primary metabolism genes, the following RAST categories were extracted: amino acids and derivatives; carbohydrates; cofactors, vitamins, prosthetic groups and pigments; fatty acids, lipids, and isoprenoids; nitrogen metabolism; and nucleosides and nucleotides. The RAST subsystems associated with secondary metabolism (cyanate hydrolysis, hopanes, polyhydroxybutyrate metabolism, nitrilase, and nitrosative stress), and the nucleosides and nucleotides subcategories detoxification and 'no subcategory' were removed to retain the main nucleotide biosynthesis, conversion, and degradation

genes. For the selected primary metabolism functions, all genes in the 'no subcategory' subsystems were combined into an 'other' subcategory for each RAST category (apart from nucleoside and nucleotide category). For analyses, each RAST role (or gene function) was counted once, although there may be several genes (or RAST features) that are annotated with each function. Due to the large variation in total number of coding sequences for each strain (Table S1.1A), relative counts were generated for the number of functions found in each RAST subcategory (scaled to the total number of primary metabolism related functions). The GenBank flat file of publicly available genomes for other wild *Drosophila*-associated bacteria were downloaded from NCBI (Table S1.1C) and were re-annotated using RAST to obtain functional trait data. For pangenome analysis, metabolic genes were extracted using a custom R script for species with more than 4 strains and were re-annotated using PROKKA v1.14.6 (Seemann 2014). For orthogroup analysis, Eggnog mapper v2 (Huerta-Cepas et al. 2019) was implemented to annotate representative sequences from each orthogroup as a general annotation, while a custom R script was used to associate RAST metabolic functions with metabolism-related orthogroups for statistical analyses.

Orthologous group gene clustering and pangenome analysis

OrthoFinder v2.4.0 (Emms & Kelly 2015, 2019) was implemented to cluster protein-coding sequences into orthogroups for all *Drosophila*-associated bacteria with default settings. Several reference genomes were included for analysis of species tree (Table S1.1C). For metabolism-related clusters, reference genomes were pruned from

orthogroup list, and a custom R script was used to extract orthogroups containing relevant metabolic functions (based on RAST annotations). Hmmer v3.3.1 was used to identify representative amino acid sequence for each orthgroup using 'hmmbuild' and 'hmmsearch' functions (hmmer.org). In addition, Roary v3.13.0 (Page et al. 2015) was used to assess variation in metabolic repertoire of prevalent species using PROKKA annotations. The pangenome distribution index was calculated as a corrected proportion of the number of core genes (subtracting the accessory gene count from the core gene count of each strain) scaled to the total number of genes found in the pangenome.

Phylogenetic and phylogenomic reconstructions

Sequences for single-gene and multi-locus phylogenies were aligned using MUSCLE (Edgar 2004) with default settings in Geneious Prime and phylogenetically informative sites were selected with GBlocks v0.91b (Castresana 2000) using less stringent options (b1-b5 settings: 0.5, 0.55, 8, 5, half). Maximum likelihood phylogenies were generated using IQ-TREE v1.6.12 (Nguyen et al. 2015) with model of evolution chosen by lowest BIC score with ModelFinder (Kalyaanamoorthy et al. 2017). Bootstrap replicates (10,000 replicates with ultrafast bootstrap approximation method) were performed to identify node support using UFBoot2 (Hoang et al. 2018). For the phylogenomic reconstruction, single orthologous gene clusters identified using OrthoFinder (52 amino acid sequences, see Table S1.2) were concatenated with SequenceMatrix v1.8 (Vaidya et al. 2011) for a partitioned model (proportional branch lengths implemented) with IQ-TREE (Chernomor et al. 2016). For statistical analyses,

reference taxa (Table S1.1C) were removed from phylogenies.

Species boundaries of sampled taxa were determined using a 95% average nucleotide identity (ANI) score threshold using JSpecies v1.2.1 (Richter & Rosselló-Móra 2009) with MuMmer v3.23 (Kurtz et al. 2004) at default settings. Taxa identities were confirmed by comparing each strain to related genomes (type specimens accessed from NCBI) and a BLASTn search for genome extracted 16S rRNA gene sequences (Table S1.1A). Individual phylogenies for each bacterial order were drawn by extracting each clade from the entire reconstruction using the packages ape v5.4 (Paradis & Schliep 2019) and ggtree v2.2.4 (Yu et al. 2017) with *Vibrio cholerae,* *Bacillus subtilis,* and Rhodospirillaceae spp. (*Magnetospirillum magneticum* and *Rhodospirillum rubrum*) used to root phylogenies of the Enterobacterales, Lactobacillales, and Rhodospirillales, respectively.

Statistics

All analyses were performed using R v4.0.2 (R Core Team 2018) with a significance α threshold of 0.05. Genome features (number of coding sequences (CDS), genome size, metabolic function count, and GC content) were assessed for phylogenetic signal using two different univariate methods. First, Pagel's λ was imputed to determine whether genomic features could be explained by phylogenetic relatedness as compared to a Brownian motion model of evolution using a likelihood ratio test (null hypothesis: $\lambda = 0$ or completely random) with the package phytools v0.7.47 (Revell 2012). Then, an analysis of variance (ANOVA) was implemented to assess the categorical effect of taxonomy on genomic features with patristic distance (sum of branch lengths from

root tip) as a covariate using the car package v3.0.8 (Fox & Weisberg 2019), except a logistic regression (quasibinomial distribution with logit link) was implemented to analyze GC content with a Wald's X^2 test for the omnibus test. Patristic distance was calculated with the 'distRoot' function from the package adephylo v1.1.11 (Jombart et al. 2010) using a *Lactococcus lactis* Bpl1 rooted tree. This taxon was selected because it is the most ancestral strain represented in the dataset (Hug et al. 2016; Zheng et al. 2020). Normality and homoscedasticity of residuals were visually assessed for each model. For all models, genome size and CDS were $\log_{10}$-transformed.

Several multivariate methods were implemented to identify relationships among bacterial metabolic traits and orthogroups. First, RAST subcategories were visualized by bacterial strain using principal coordinates analysis (PCoA) with Bray-Curtis dissimilarities on relative counts (proportions were based on total number of function counts in selected RAST subcategories related to primary metabolism pathways) using the 'capscale' function in the vegan package v2.5.6 (Oksanen et al. 2019). Orthogroup incidence was visualized with a PCoA using a Jaccard similarity coefficient for presence-absence data. Second, a permutational multivariate analysis of variance (PERMANOVA) was performed with the 'adonis' function to determine whether metabolic traits and orthogroup incidence varied by bacterial taxonomy with 999 permutations and Bray-Curtis dissimilarities on relative count data or Jaccard similarity coefficient for presence-absence data. A *post hoc* pairwise PERMANOVA was implemented using the 'adonis.pair' function from the EcolUtils package v0.1 (Salazar 2020) with 999 permutations and Benjamini-Hochberg false discovery rate p-value correction method (FDR). Then, a Ward's linkage agglomerative hierarchical

cluster was applied to relative count data with Bray-Curtis dissimilarities to generate a dendrogram by bacterial strains. The pvclust package v2.2.0 (Suzuki et al. 2019) was implemented to identify significant clusters in the hierarchical cluster with approximately unbiased p-values and bootstrap probability support values (n = 10,000). Finally, PCoA of the metabolism-related orthogroups were correlated with the PCoA of all orthogroups using a Procrustean randomization test (999 permutations) in the vegan package with the function 'protest.'

Correlation between dendrograms was determined using two metrics. First, normalized Robinson-Foulds metric was calculated using the phangorn package v2.5.5 (Schliep 2011) to test for congruence between dendrogram topologies. nRF values are bounded between 0 and 1, corresponding to complete congruence to incongruence. Then, a Mantel test was performed to associate two distance matrices using Spearman's rank correlation with 999 permutations using vegan. For phylogenies, cophenetic distances (pairwise sum of branch lengths) were calculated using the 'cophenetic.phylo' function in ape. Bray-Curtis dissimilarities were used for the relative function counts. Tanglegrams were generated using the dendextend package v1.13.4 (Galili 2015) with the 'step2side' aligner.

For the analysis of prevalent species, represented in at least four flies and comprising >4 strains, several methods were used to compare pangenome distribution and functional content. Differences in the pangenome distribution index were examined with a beta regression using the betareg package v3.1.3 (Cribari-Neto & Zeileis 2010). A likelihood ratio test was used to assess the effect of species by comparing the regression to an intercept-only model with the package lmtest v0.9.37 (Zeileis &

Hothorn 2002) and a *post hoc* Tukey's test was implemented with the emmeans package. Pearson's product-moment correlation coefficient was used to assess linear association between strain diversity and pangenome distribution gene count. Strain diversity was scored in two ways: first with Shannon's entropy on the concatenated amino acid sequence alignment used in the phylogenomics analysis with the Bio3d package (Grant et al. 2006) and then nucleotide diversity was obtained for 16S rRNA gene alignment with the pegas package (Paradis 2010). A two-sided Fisher's exact test was used to compare orthogroup incidence between species, while a one-sided Fisher's exact test was used for enrichment of subsystems in the accessory genome compared to the core genome of each species. The odds ratios (OR) were calculated based of the function count of a given subsystem in the accessory genome relative to the rest of the function counts in the core genome. FDR was used to correct for multiple Fisher's exact tests.

Results

Sequencing and characterization of bacterial genomes

In this study, we assessed whether primary metabolism functions found in gut bacterial microbiome members of wild *Drosophila* can be mapped onto bacterial taxonomy. First, we characterized the genomic features of the strains found in each bacterial order. Given that few bacterial species associated with wild *Drosophila* have been isolated and sequenced previously, we collected and sequenced 81 newly isolated strains that are members of the three dominant bacterial orders (i.e. Enterobacterales, Lactobacillales, and Rhodospirillales) found within the fly gut to complement the 15

genomes currently available (Table 1.1 and S1.1A&B). Genome features (genome size, number of coding sequences (CDS), and GC content) of all newly-sequenced taxa (Table S1.1A) were similar to publicly available species. The estimated genome sizes of the strains sequenced ranged from 1.8 to 5.8 Mbp with 1,879-5,983 CDS and GC content of 37-60%. Average coverage (i.e. sequence depth) of genomes ranged from 53x to 1,390x (Table S1.1A). The number of annotated metabolic functions by RAST ranged from 286 to 968. Comparisons of genomic features indicated that all measures significantly differed by bacterial order (Fig. 1.1A-D). In addition, phylogenetic signal was found for all four genomic features scored using Pagel's λ and patristic distance (based on phylogenomic analysis) as a covariate in ANOVA and logistic regression analyses (Fig. 1.1A-D), indicating that closely related taxa tend to have similar genome characteristics.

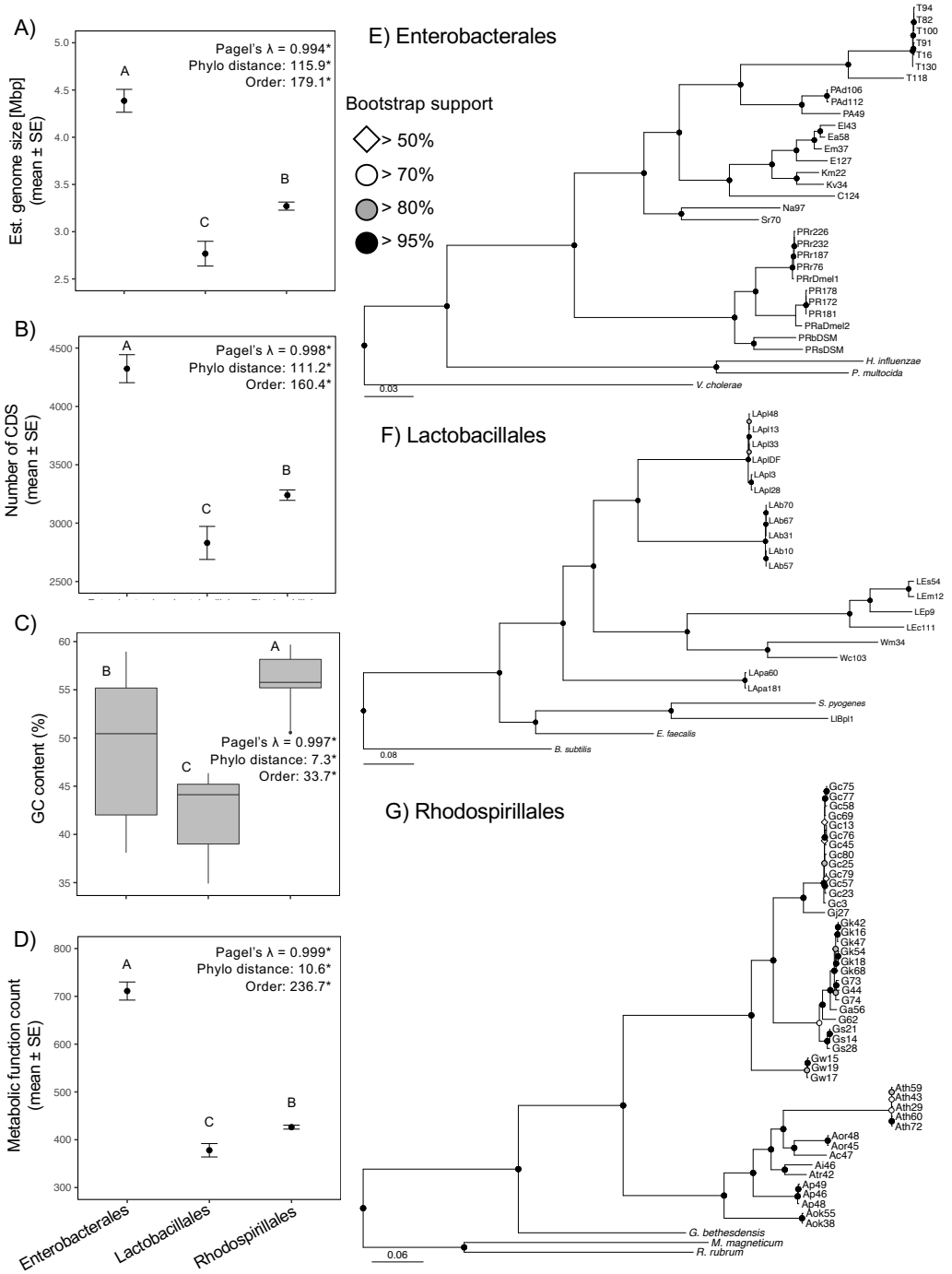

A)
Est. genome size [Mbp] (mean ± SE)
Pagel's λ = 0.994*
Phylo distance: 115.9*
Order: 179.1*
B)
Number of CDS (mean ± SE)
Pagel's λ = 0.998*
Phylo distance: 111.2*
Order: 160.4*
C)
GC content (%)
Pagel's λ = 0.997*
Phylo distance: 7.3*
Order: 33.7*
D)
Metabolic function count (mean ± SE)
Pagel's λ = 0.999*
Phylo distance: 10.6*
Order: 236.7*
Enterobacterales
Lactobacillales
Rhodospirillales
E) Enterobacterales
Bootstrap support
> 50%
> 70%
> 80%
> 95%
T94
T82
T100
T91
T16
T130
T118
PAd106
PAd112
PA49
EI43
Ea58
Em37
E127
Km22
Kv34
C124
Na97
Sr70
PRr226
PRr232
PRr187
PRr76
PRrDmel1
PR178
PR172
PR181
PRaDmel2
PRbDSM
PRsDSM
H. influenzae
P. multocida
V. cholerae
0.03
F) Lactobacillales
LApl48
LApl13
LApl33
LAplDF
LApl3
LApl28
LAb70
LAb67
LAb31
LAb10
LAb57
LEs54
LEm12
LEp9
LEc111
Wm34
Wc103
LApa60
LApa181
S. pyogenes
LIBpl1
E. faecalis
B. subtilis
0.08
G) Rhodospirillales
Gc75
Gc77
Gc58
Gc69
Gc13
Gc76
Gc45
Gc80
Gc25
Gc79
Gc57
Gc23
Gc3
Gj27
Gk42
Gk16
Gk47
Gk54
Gk18
Gk68
G73
G44
G74
Ga56
G62
Gs21
Gs14
Gs28
Gw15
Gw19
Gw17
Ath59
Ath43
Ath29
Ath60
Ath72
Aor48
Aor45
Ac47
Ai46
Atr42
Ap49
Ap46
Ap48
Aok55
Aok38
G. bethesdensis
M. magneticum
R. rubrum
0.06

Figure 1.1. Genomic features and phylogenomic analysis of *Drosophila*-associated bacteria. (A) Estimated genome size, (B) number of CDS (coding sequences), (C) GC content, and (D) RAST metabolic function counts by bacterial order. (E-F) Phylogenomic reconstruction for (E) Enterobacterales, (F) Lactobacillales, and (G) Rhodospirillales. For all genomic features, the raw means and standard error are displayed, except for box plots used in panel C to show GC content. Pagel's λ and F-statistic or X^2 statistics for model predictors (order and phylogenetic distance) are displayed for each panel (residual df = 92). * All p values are less than <0.01. Phylogenetic distance is calculated from branch lengths of phylogenomic reconstruction. Letters represent statistical grouping from *post hoc* Tukey's test. Phylogenomic analysis is based on the concatenated sequence (length = 13,238 amino acids) of 52 genes (details in Table S7). Dendrograms are scaled to amino acid divergence.

The taxonomy of the newly isolated bacterial strains was characterized by two methods: genome comparisons of average nucleotide identity (ANI) to genomes of type specimens, and BLAST search of genome extracted 16S rRNA gene against the non-redundant NCBI database. Based on ANI scores, 80% of the strains were identified to the species level (Table S1.1A). The remainder of the strains were identified to the genus level using 16S rRNA gene sequence where no close ANI match was available (Table S1.1A). In addition, a previously sequenced genome *Acetobacter* sp. DmW-043 (Winans et al. 2017) was identified as *Acetobacter thailandicus* (98.9% ANI to *A. thailandicus* LMG 30826, accession: GCA_011516655, which was not available at the time of publishing this genome sequence). A phylogenomic analysis of 52 single-orthologs supported the ANI species boundaries with strong bootstrap node support (generally >95%, although some of the Rhodospirillales species had node support >70%; Fig. 1.1E-G & S1.1). Most species and genera included formed monophyletic clades. Exceptionally, *Leuconostoc* and *Weissella* spp. were embedded within the paraphyletic *Lactobacillus* genus, as found previously in larger phylogenomic analyses of this group (Salvetti et al. 2018; Zheng et al. 2020). In addition, the evolutionary relationships between taxa of the Enterobacterales and Rhodospirillales were consistent with published datasets containing additional species from each order (Adeolu et al. 2016; Matsutani et al. 2011; Baek et al. 2020; Yukphan et al. 2020).

Association of 16S rRNA gene with phylogenomic relationships

As 16S rRNA gene sequencing is widely used in taxonomic surveys for microbiome

studies, we investigated how well 16S sequence predicted species identity and phylogenomic relationships of the strains used in this study. In the BLAST top matches with 16S rRNA genes, 77% of the strains had more than one species match (Table S1.1A). Similarly, many of the 16S rRNA genes yielded ≥97% identity matches (general threshold for species boundaries) between strains of the different species isolated. This applied especially to *Gluconobacter* and *Leuconostoc* spp. and many of the Enterobacterales strains (Table S1.3), consistent with published data of these taxa (Adeolu et al. 2016; Matsutani et al. 2011; Jeon et al. 2017), and indicated the 16S rRNA gene does not always infer species identity reliably. Phylogenetic analysis of 16S rRNA genes tended to have lower bootstrap support than the phylogenomic analysis (Fig. S1.2). Many of the species clusters identified by phylogenomics were evident in the 16S phylogeny, but some of the Enterobacteriaceae and *Gluconobacter* spp. were mis-identified as polyphyletic (Fig. S1.2).

Two complementary methods were implemented to compare congruence between the phylogenomic analysis and 16S rRNA phylogeny. First, normalized Robinson-Foulds index was used to compare dendrogram topologies, which indicated that 16S rRNA gene phylogeny had the best correspondence with Lactobacillales and weakest association with the Rhodospirillales (Fig. 1.2). Second, a Mantel test was implemented to correlate the cophenetic distances between all taxa of each dendrogram. All three bacterial orders displayed strong, statistically significant correlation, indicating that 16S rRNA gene phylogeny retains much of the overall taxonomic placement of species relationships found in the phylogenomic analysis.

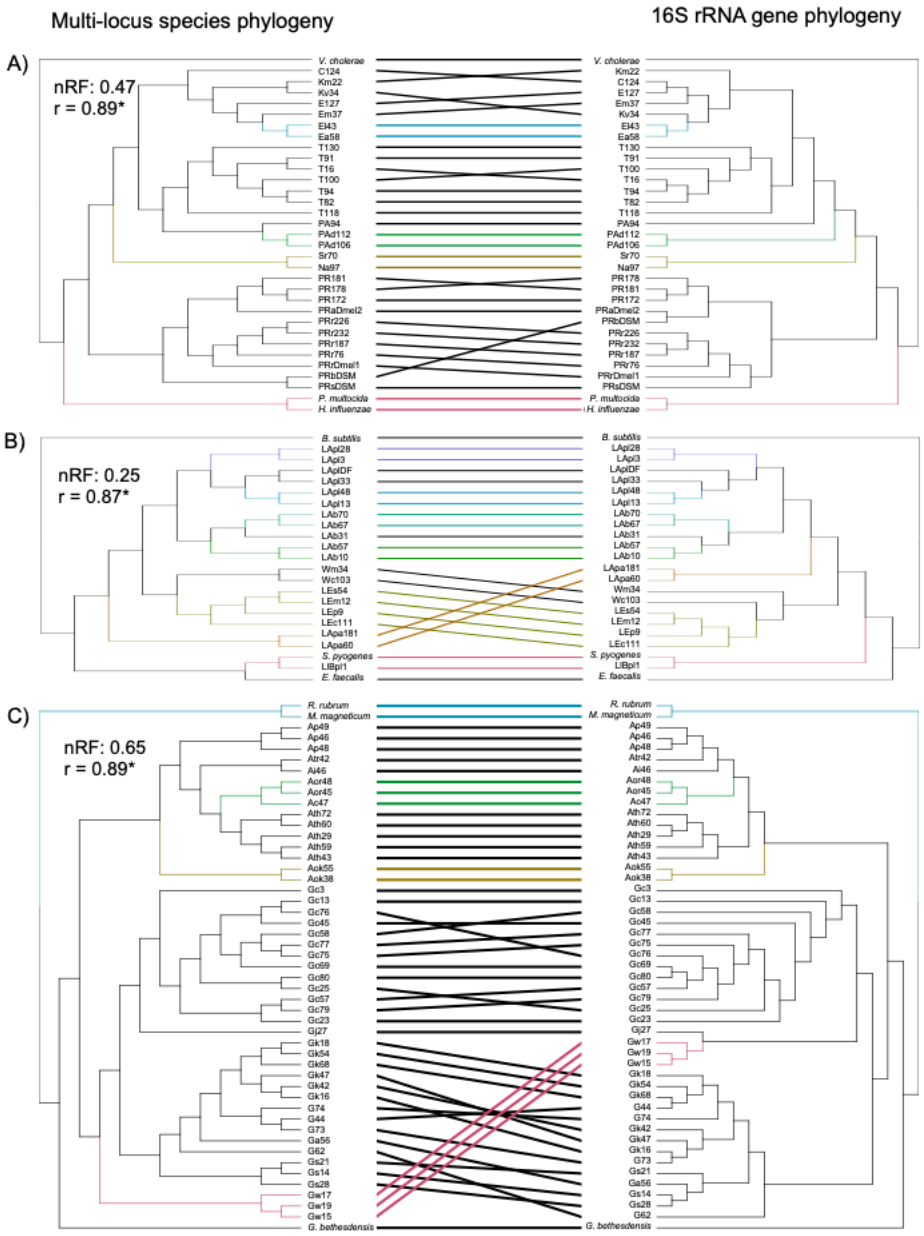

Multi-locus species phylogeny
16S rRNA gene phylogeny
A)
nRF: 0.47
r = 0.89*
V. cholerae
C124
Km22
Kv34
E127
Em37
El43
Ea58
T130
T91
T16
T100
T94
T82
T118
PA94
PAd112
PAd106
Sr70
Na97
PR181
PR178
PR172
PRaDmel2
PRr226
PRr232
PRr187
PRr76
PRrDmel1
PRbDSM
PRaDSM
P. multocida
H. influenzae
B)
nRF: 0.25
r = 0.87*
B. subtilis
LApl28
LApl3
LApIDF
LApl33
LApl48
LApl13
LAb70
LAb67
LAb31
LAb57
LAb10
Wm34
Wc103
LEs54
LEm12
LEp9
LEc111
LApa181
LApa60
S. pyogenes
LBpl1
E. faecalis
C)
nRF: 0.65
r = 0.89*
R. rubrum
M. magneticum
Ap49
Ap46
Ap48
Atr42
Ai46
Aor48
Aor45
Ac47
Ath72
Ath60
Ath29
Ath59
Ath43
Aok55
Aok38
Gc3
Gc13
Gc76
Go45
Gc58
Gc77
Gc75
Gc69
Gc80
Gc25
Gc57
Gc79
Gc23
Gj27
Gk18
Gk54
Gk68
Gk47
Gk42
Gk16
G74
G44
G73
Ga56
G62
Gs21
Gs14
Gs28
Gw17
Gw19
Gw15
G. bethesdensis

Figure 1.2. Relationship between phylogenome and 16S rRNA gene phylogeny. Tanglegrams for (A) Enterobacterales. (B) Lactobacillales. (C) Rhodospirillales. Normalized Robinson-Foulds (nRF) indices and Mantel test correlations are displayed for each order. Subtrees with the same topologies between each dendrogram are colored. * represented $p = 0.001$.

Correspondence between metabolic traits and phylogeny

Bacterial traits were grouped by the 38 RAST subcategories related to primary metabolism to infer correspondence between bacterial phylogeny and distributions of metabolic functions. Despite some variability in counts between genomes from each order, the taxa belonging to the order Enterobacterales tended to have more functions related to amino acid, carbohydrate, and vitamin metabolism than Lactobacillales and Rhodosprillales, while functions involved in lipid, nitrogen, and nucleotide metabolism generally had similar counts across all taxa (Fig. 1.3A). The expanded range of functions in the Enterobacterales is likely linked to the relatively large genome size and number of coding sequences in these bacteria (Fig. 1.1).

Principal coordinates analysis (PCoA) was applied to visualize the relationship between taxonomy and metabolic potential using relative counts to normalize the data (Fig. 1.3B). On the first axis, the three bacterial orders were distinctly separated, while on the second axis, the Enterobacterales were separated from the other two orders. PERMANOVA indicated a large effect by bacterial order on metabolic trait groupings ($F_{2,93} = 61.72$, $p = 0.001$, $R^2 = 0.57$) and a pairwise PERMANOVA analysis revealed that all three orders were significant different from each other after FDR p-value correction (Table S1.4A). In addition, each of the clusters separated by genus-level taxonomy, apart from some mixing between *Providencia* and *Tatumella* spp., further indicating metabolic differentiation by taxonomy (Fig. S3).

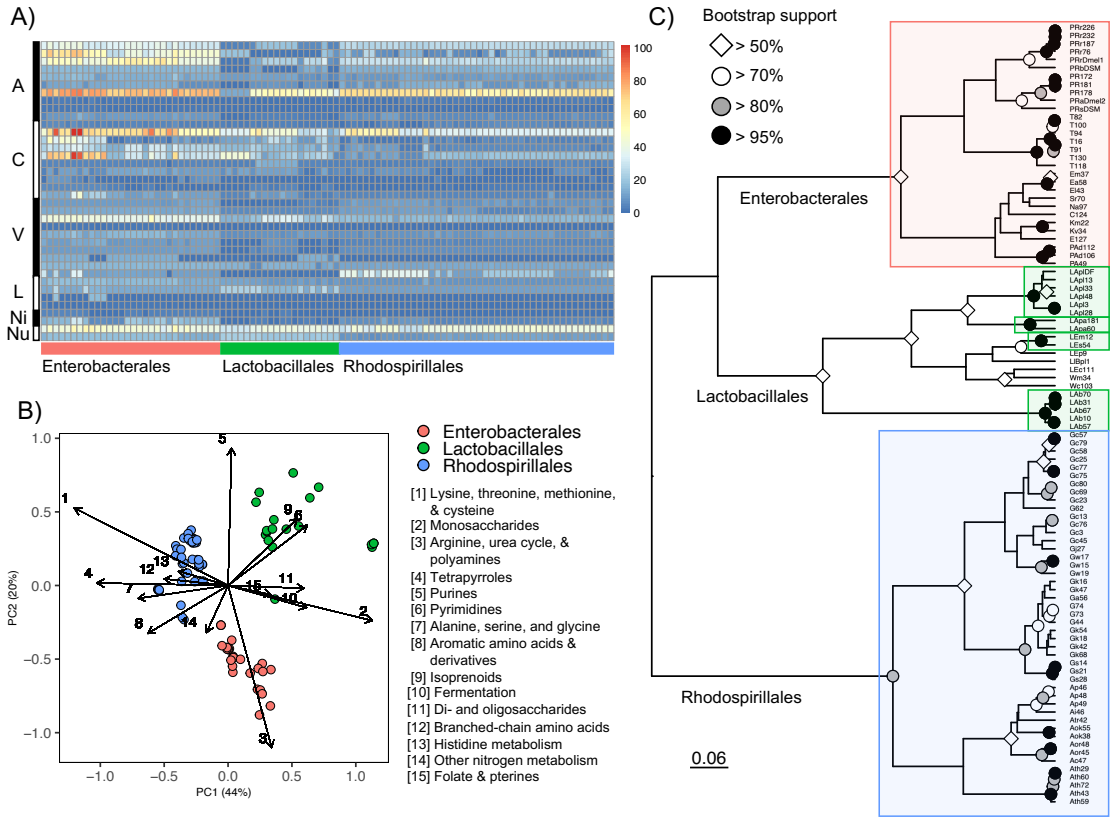


Figure 1.3. Taxonomic correspondence with encoded metabolic functions. (A) Heatmap of raw function counts in RAST subcategories displayed by bacterial order. Rows and columns are organized by alphabetical order for RAST categories and bacterial taxonomy. RAST subcategories are grouped by categories: A = amino acids and derivatives, C = Carbohydrates, V = cofactors, vitamins, prosthetic groups, and pigments, L = fatty acids, lipids, and isoprenoids, Ni = nitrogen metabolism, and Nu = nucleosides and nucleotides. (B) Principal coordinates analysis (PCoA) of relative counts for RAST subcategories with Bray-Curtis dissimilarity matrix. Arrows indicate the loading subcategories (top 15 displayed) and the percent variance explained for each axis is displayed. (C) Hierarchical cluster of relative counts for RAST subcategories. Significant clusters are boxed and colored by bacterial order.

The top 15 loadings were displayed to identify RAST subcategories that were associated with each bacterial order (Fig. 1.3B). Generally, each order was associated with different metabolic functions; the Rhodospirillales were driven by amino acid metabolism, while the Lactobacillales were associated with carbohydrate, nucleotide, and lipid metabolism and Enterobacterales were influenced by arginine, urea cycle, and polyamine metabolism. Both the Lactobacillales and Rhodospirillales were associated with functions related to vitamin and cofactor metabolism. Lastly, the Rhodospirillales may share some of the nitrogen metabolism functions with Enterobacterales, as it relates to different organic and inorganic nitrogen metabolic pathways (subcategory contains ammonia fixation, allantoin utilization, and amidase subsystems).

The metabolic functions of each strain were further clustered using an agglomerative hierarchical method (Fig. 1.3C). The three orders clustered separately with >50% bootstrap probability support. The Enterobacterales and Rhodospirillales bacteria formed two significant clusters with almost all of the genera grouped together for each order (except some of the *Gluconobacter* spp.). The Lactobacillales formed three significant clusters by *Lactobacillus* spp. (*L. brevis, L. paracasei, and L. plantarum*), as well as another *Leuconostoc* spp. cluster (Fig. 1.3C), which all had >95% bootstrap probability support. The remainder of the Lactobacillales species only had single strain representatives, likely influencing the lack of clusters.

To further understand the relationship between phylogeny and distribution of metabolism functions, the hierarchical cluster was correlated with phylogenomic and 16S rRNA gene dendrograms using normalized Robinson-Foulds index and Mantel

test. Overall topologies of dendrograms were moderately associated between phylogeny and metabolic traits, with the best congruence found when associating function with phylogenomic analysis (likely driven by the congruence of the Enterobacterales members) (Fig. S1.4). Results from Mantel test comparing cophenetic distances of each phylogeny with Bray-Curtis dissimilarities supported this finding with a 1.2x increase in correlation statistic when using the phylogenomic reconstruction compared to the 16S rRNA gene phylogeny, suggesting that the increased resolution of the species tree amplified the phylogenetic signal for overall distribution of metabolic traits. Further inspection of the tanglegrams indicated that few strains had overlapping topologies between dendrograms (i.e. displayed the same node-edge relationships), further supporting weak to moderate congruence between dendrograms found for each order (Fig. S1.4).

As a complementary analysis, orthogroups were identified between all 96 taxa to determine whether a finer resolution at the gene family incidence level would reflect the functional relationships observed based on RAST annotations. A PCoA was used to visualize taxonomic relationships between the 13,170 orthogroups with genes from at least three genomes using a Jaccard similarity coefficient. All three bacterial orders clustered away from one another, with Rhodospirillales separating on the first axis away from the Enterobacterales and Lactobacillales and all three orders distinctly separating on the second axis (Fig. S1.5). Orthogroup composition of each genome significantly differed by order (PERMANOVA: $F_{2,93} = 63.2$, $R^2 = 0.58$, $p = 0.001$, Table S1.4B). Of the total orthogroups identified, 8% were involved in metabolism-related functions (defined by RAST annotations). These 1,055 orthogroups were

extracted and subjected to the same analysis, resulting in a similar finding that all three bacterial orders are distinct on both PCoA axes (Fig. S1.5), with PERMANOVA support ($F_{2,93} = 106.5$, $R^2 = 0.7$, $p = 0.001$, Table S1.4C). In addition, a Procrustean randomization test indicated that the orientation of metabolism-related orthogroups was highly correlated with the overall relationship among all orthogroups ($m^2 = 0.008$, $r = 0.996$, $p = 0.001$).

Variation in metabolism genes of prevalent species

To extend our analysis of metabolic variation among the *Drosophila*-associated bacteria, we focused on 7 species, which we termed 'prevalent' by the criteria that they were isolated from at least 4 flies and were represented by >4 strains (Table 1.1). These taxa provide the opportunity to define the distribution of metabolic traits from a pangenomic perspective, including comparisons in orthogroup membership between species and identification of among-strain variation, i.e. enriched functions in accessory genome.

A pangenome analysis was performed using Roary to identify single orthologous genes encoding metabolic functions found within each of the seven species and to define the distribution of genes found in the metabolic pangenome. Across the 7 species, the total pangenome ranged from 287 to 538 metabolism related genes with core genome size of 264-488 and accessory genome size of strains ranging from 0-102 genes (Fig. 1.4A). A distribution index was generated to compare the relative sizes of the metabolic pangenome with values close to 1 indicating a small accessory genome with few genes per strain; values closer to 0 indicate high strain diversity with equal

numbers of genes in the core and accessory genome for a strain. A beta regression of indices for each species indicated significant differences by species (likelihood ratio test: $X^2_6 = 200.2$, $p < 2.2x10^{-16}$) with no specific statistical similarities between species with similar taxonomy, although both *P. rettgeri* and *Tatumella* sp. (Enterobacterales) had small metabolic accessory genomes (Fig. 1.4B). The relatively large accessory genome sizes of *G. cerinus, G. kondonii*, and *L. plantarum* correlated with increased residue diversity (index for measuring average strain diversity within species using the amino acid alignments from the phylogenomic analysis) (Fig. 1.4C and S1.6). Additionally, nucleotide diversity based on 16S rRNA genes among strains was also assessed and no significant relationships were found, indicating the increased resolution of the phylogenomic analysis was required to score strain diversity between species (Fig. S1.6).

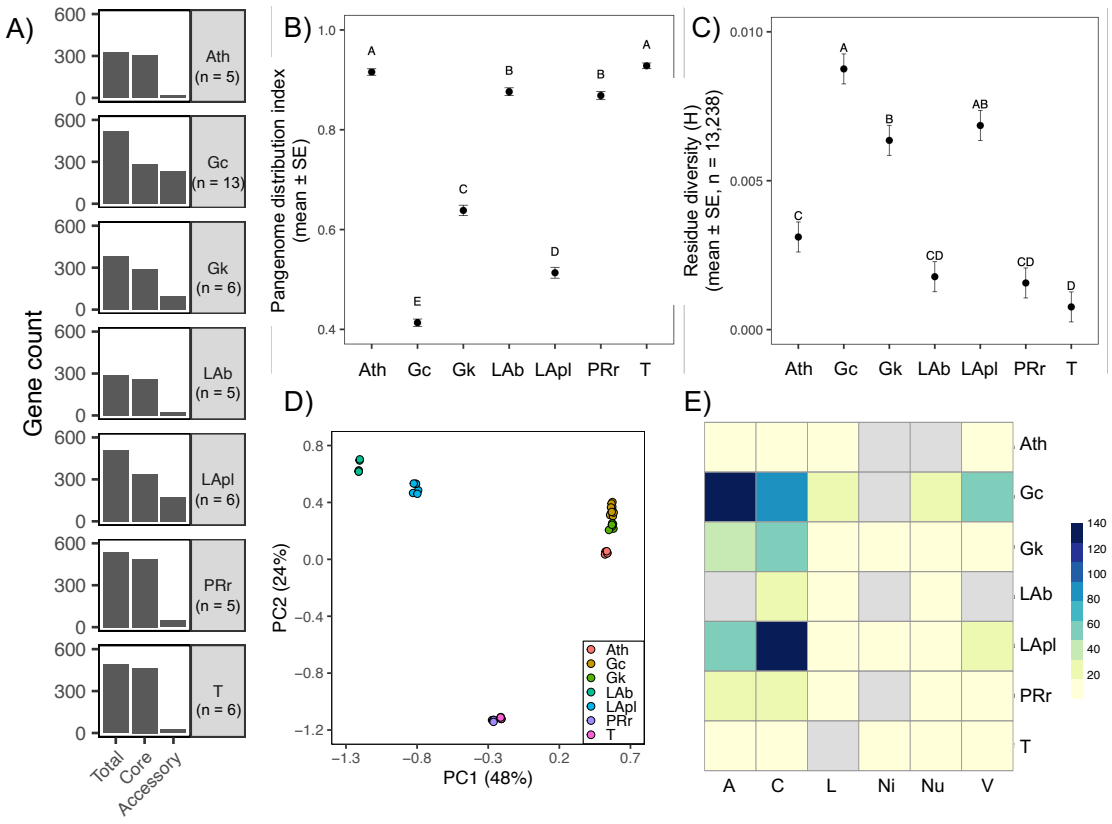

A)
Gene count
600
300
0
Ath (n = 5)
Gc (n = 13)
Gk (n = 6)
LAb (n = 5)
LApl (n = 6)
PRr (n = 5)
T (n = 6)
Total
Core
Accessory
B)
Pangenome distribution index (mean ± SE)
1.0
0.8
0.6
0.4
A
E
C
B
D
B
A
Ath Gc Gk LAb LApl PRr T
C)
Residue diversity (H) (mean ± SE, n = 13,238)
0.010
0.005
0.000
C
A
B
AB
CD
CD
D
Ath Gc Gk LAb LApl PRr T
D)
PC2 (24%)
PC1 (48%)
0.8
0.4
0.0
−0.4
−0.8
−1.2
−1.3
−0.8
−0.3
0.2
0.7
Ath
Gc
Gk
LAb
LApl
PRr
T
E)
Ath
Gc
Gk
LAb
LApl
PRr
T
A C L Ni Nu V
140
120
100
80
60
40
20

Figure 1.4. Metabolic pangenome analysis of prevalent species. (A) Distribution of metabolism genes in pangenome of each species with number of strains listed below each taxon identifier. (B) Relative pangenome distribution by species. (C) Strain diversity by species using Shannon's entropy score. (D) PCoA of orthogroup composition among species (E) Heatmap of function counts (20-140) in accessory genome by species. In panels B and C, estimated marginal means and standard error are plotted from each model with letters from *post hoc* Tukey's test representing statistical groups. Percent variation explained among significantly different orthogroups is shown for each axis in panel D. Gray cells in heatmap of panel E indicate function is absent in accessory genome. Species identifiers: Ath = *A. thailandicus*, Gc = *G. cerinus*, Gk = *G. kondonii*, LAb = *L. brevis*, LApl = *L. plantarum*, PRr = *P. rettgeri*, and T = *Tatumella* sp. RAST categories: A = amino acids and derivatives, C = Carbohydrates, V = cofactors, vitamins, prosthetic groups, and pigments, L = fatty acids, lipids, and isoprenoids, Ni = nitrogen metabolism, and Nu = nucleosides and nucleotides. Data are provided in Table S5.

The contribution of between- and within-species differences to variation in metabolic traits was investigated using two methods. In the first approach, the metabolic traits between species were compared using Fisher's exact test on orthogroup membership based on incidence (minimum threshold of three genomes represented per orthogroup). After p-value correction, 597 of the 717 orthogroups were significantly different between species. A PCoA with a Jaccard similarity coefficient was used to visualize how species separated by significant orthogroups. All seven species formed distinct taxonomic clusters (PERMANOVA: $F_{6,39} = 775.23$, $R^2 = 0.99$, $p = 0.001$, Table S1.4D) with bacteria separating by order on the first axis and Lactobacillales and Rhodospirillales separated from the Enterobacterales strains on the second axis (Fig. 1.4D). The top RAST category was assigned to each orthogroup and associated with each PCoA axis, indicating Rhodospirillales were enriched in amino acid, nucleotide, and vitamin metabolism while Enterobacterales were enriched in all metabolism categories except nucleotide metabolism (Fig. S1.7). This analysis of 7 species largely recapitulates the analysis of all strains, as displayed in Fig. 1.3B.

Further investigation into the orthogroup analysis established that the top significant orthogroups were primarily involved in carbohydrate metabolism (~50%) and that most orthogroups were present in the two *Gluconobacter* spp. while the other taxa had lower incidence rates across gene families (Table S1.5). When a given species was a member of a top orthogroup identified, all strains were found to contain at least one gene from this gene family. In addition, all of the top orthogroups were present in at least two species, and were generally not defined by higher order taxonomy (e.g. *Gluconobacter* and *Tatumella* spp. tended to have similar orthogroup functions). Of

the top gene functions identified, several were noteworthy for known effects on *Drosophila* physiology. Some of the sugar and sugar derivative dehydrogenases have been implicated as determinants of reduced lipid content in adult flies by incomplete oxidation of external carbohydrates (Chaston et al. 2014). Our analysis (Table S1.5) also identified a bacterial methionine salvage gene (5-methylthioribose kinase), which bacterial methionine metabolism lowers starvation resistance of *Drosophila* (Judd et al. 2018), and the hydroxymethylpyrimidine ABC transporter involved in the production of thiamine (vitamin B_1) (Table S1.5), an important determinant of larval development and survival on low-nutrient diets (Sannino et al. 2018). In addition, several gene functions that aid in the bacterial growth and utilization of *Drosophila* metabolic waste products (Storelli et al. 2018; Winans et al. 2017) were identified (Table S1.5). Notably, N-acetylglucosamine gene families (the monomer of chitin found in the peritrophic envelope of the insect gut as well as fungal cell walls) and xanthine degradation gene families (part of an Acetobacteraceae uric acid degredation locus, the primary nitrogen waste product of *Drosophila*) were top gene functions (Table S1.5). For the latter orthogroup, we further inspected whether the uricase gene was also present in the genomes of the prevalent strains, as it is not a function classified by the RAST subsystem annotations. This gene was present in all genomes of prevalent *Gluconobacter* and *Tatumella* spp. and is part of the orthogroup OG0001450, indicating these taxa may potentially utilize uric acid egested by *Drosophila*.

Our second analysis of variation in metabolic traits across the 7 prevalent species identified gene functions enriched in the accessory genome of each species compared

to the core genome (Table S1.6A-G). Among the different functional annotation counts in the accessory genome, amino acid metabolism in *G. cerinus* and carbohydrate metabolism in *L. plantarum* were the highest and nitrogen metabolism were low or absent in all 7 species (Fig. 1.4E). Only 7 RAST subsystems were identified as enriched in the accessory genome of four species (*G. kondonii*, both *Lactobacillus* spp., and *P. rettgeri*) after p-value correction for multiple testing (Table S1.6H). Each species included carbohydrate metabolism gene functions predicted to expand the capacity of the bacteria to utilize and ferment different carbohydrates (potentially glucose, gluconate, fructose, mannose, and trehalose) and the carboxylic acid citrate, which is important for the growth and acid resistance of *Lactobacillus* (Martin et al. 2005). Lipid/carbohydrate metabolism (related to short chain fatty acid butyric acid fermentation) and purine biosynthesis were implicated as enriched in the *P. rettgeri* accessory genome (Table S1.6F). Most of the ortholog functions in *L. plantarum* and *P. rettgeri* were exclusively found in these taxa, while the other orthologs of *G. kondonii* and *L. brevis* were found in the pangenomes of at least one other prevalent species examined (Table S1.6H), indicating that some of the accessory genome functions can be redundant among closely and distantly related taxa.

Discussion

A robust understanding of the relationship between the taxonomic identity and functional traits of microorganisms is essential for detailed analyses of the ecological and evolutionary processes that shape microbial communities. This relationship is particularly important for the microbial communities in animal guts because microbial

function can influence many host traits, but the pattern and scale of the effect of variation in taxonomic composition on microbial function are poorly understood. This study on the comparative genomics of bacteria isolated from the guts of wild *Drosophila* focused on bacterial metabolic traits, which have been implicated in the metabolic health and fitness of animal hosts (McFall-Ngai et al., 2013; Visconti et al., 2019), including *Drosophila* (Consuegra et al. 2020; Bost et al. 2018; Newell et al. 2014; Chaston et al. 2014). Two key results were obtained. First, representatives of the three dominant bacterial orders (Enterobacterales, Lactobacillales and Rhodospirillales) can be differentiated by key metabolic traits, based on annotations and homology of metabolism related genes. Second, evidence for within-species variation in metabolic function was obtained, including for functions relevant to utilization of the sugar-rich habitats and interactions with the *Drosophila* host. Here, we consider these two issues in turn.

Our finding that the variation in metabolic function partitions by the three bacterial orders of gut bacteria (Fig. 1.3) reflects the differences in lifestyles of the bacteria. For interpreting these results, it should be considered that these differences relate exclusively to the panel of genomes isolated from *Drosophila* guts, comprising members of just one, two and four families for Rhodospirillales, Lactobacillales and Enterobacterales, respectively (Table 1.1). The diversity of taxa studied are also functionally restricted by the conditions in the *Drosophila* gut, including physical instability, hypoxia (but not anoxia), low pH, and immunological defenses (Douglas 2018; Lemaitre & Miguel-Aliaga 2013). A final potential issue is that some taxa in the *Drosophila* gut microbiome may be intractable to cultivation but the magnitude of this

difficulty is likely low because the taxa in the genome panel (Table 1.1) match well to the results from cultivation-independent studies on *Drosophila* collected from the same habitats in New York State (Bost et al. 2018; Kang & Douglas 2020; Adair et al. 2018). The key lifestyle features of Acetobacteraceae (Rhodospirillales) relate to their adaptation to high sugar habitats, such as the rotting fruits utilized by *Drosophila* (Lievens et al., 2015). The distinctive metabolic features identified in this study (Fig. 1.3B) relate to aerobic fermentation of exogenous sugars via processes dependent on the tetrapyrrole derivative pyrroloquinoline quinone (Matsutani & Yakushi 2018) and the capacity to utilize simple inorganic and organic nitrogenous substrates for the synthesis of amino acids required for protein synthesis and proliferation (Sainz et al. 2017). Similarly, all but one of the Lactobacillales in this study comprised a taxonomically-restricted set of three paraphyletic genera in the family Lactobacillaceae and have the functional traits of fermentative metabolism, especially of sugars and other organic compounds, including terpenes and nucleotides (Duar et al., 2017). The Enterobacterales associated with *Drosophila* are taxonomically and functionally more diverse (Fig. 1.1D & E). The lifestyles represented by the Enterobacterales in our panel likely include both free-living bacteria associated with the food ingested by the flies and taxa that may be pathogenic to *Drosophila*, e.g. some strains of *P. rettgeri* (Adair et al. 2018; Galac & Lazzaro 2011). This metabolic diversity probably accounts for the single metabolic trait that partitions with the Enterobacterales (Fig. 1.3B). Unlike the Acetobacteraceae and Lactobacillaceae, the dynamics of Enterobacterales and other γ-Proteobacteria in the *Drosophila* gut are unknown, and these bacteria have generally not been identified as beneficial to

Drosophila metabolic health and fitness. The association of Enterobacterales with the urea cycle and polyamine synthesis raises the possibility that the association of these bacteria with *Drosophila* may be facilitated by their capacity to utilize *Drosophila* waste urea as a nitrogen source and to tolerate hostile conditions in the gut via polyamine-mediated stabilization of the genome and membranes. Microbiome-mediated polyamine production has also been implicated in microbiome effects on human health (Tofalo et al., 2019), but the role of this class of metabolites in *Drosophila*-microbe interactions has not been investigated.

The parallel analysis of within-species variation, conducted on 7 species with at least 5 sequenced genomes, provided the opportunity to assess the scale of among-strain variation in genetic and functional variation in metabolism, including metabolic traits with known effects on *Drosophila* nutritional physiology and performance (e.g. Shin et al. 2011; Judd et al. 2018; Chaston et al. 2014; Kang & Douglas 2020; Winans et al. 2017). For this analysis, we used two approaches. First, we compared between-species genetic variation (Fig. 1.4D), which was congruent with annotation-based analysis in Fig. 1.3B. Of the top gene functions found to vary by species, only a few were relevant determinants of *Drosophila* physiology and some were functionally redundant across disparate taxa. Several genes involved utilization of *Drosophila* nitrogenous waste products, and they were also identified primarily among *Glucobacter* spp., which may allow these taxa to use host nitrogenous waste for their own growth. The second analysis focused on identifying functions enriched in the accessory genome of each species. Interestingly, the majority of the genes that differed within species related to carbohydrate digestion and fermentation as well as carboxylic acid and short

chain fatty acid metabolism. The enrichment of carbohydrate metabolism genes is also supported by published pangenome analyses of *L. plantarum* and *P. rettgeri* (Galac & Lazzaro 2012; Martino et al. 2016). Taken together, the identified gene functions are suggestive of survival in sugar-rich rotting fruit environment that is enriched by the waste products of *Drosophila* larvae and possibly adults (Winans et al. 2017; Storelli et al. 2018)) (Lievens et al. 2015).

Rotting fruit provide an energy-rich but ephemeral resource colonized by numerous microorganisms. In this environment, there is strong selective pressure to utilize carbon sources due to exploitative competition and the release of toxic metabolic by-products by co-occurring microbes (e.g. citrate lyase gene functions can be involved in acid stress in *Lactobacillus* spp. (Martin et al. 2005)). Although we did not sample strains from this habitat, various studies indicate that there is frequent cycling between wild *Drosophila* and external environment (Inamine et al. 2018; Blum et al. 2013; Pais et al. 2018), and that this likely prevents genetic differentiation between strains in *Drosophila* and externally (Winans et al. 2017).

This study also raises two key methodological issues. The first relates to the utility of 16S rRNA gene sequence data for taxonomic identification and inference of functional traits. Our analysis reinforces the conclusion of many previous studies, including research on microbiomes, that 16S data can be insufficiently precise to discriminate functionally different microorganisms because functionally important sequences are gained, lost or modified by mutation more rapidly than 16S sequence change (Ellegaard & Engel 2016; Lladó Fernández et al. 2019; Koeppel & Wu 2013). 16S rRNA gene sequence evolution can also yield phylogenetic patterns that are

incongruent with patterns from phylogenomic data, as illustrated for several taxa in Fig. 1.2 as well as other bacterial orders (e.g. Maayer et al. 2019). For these reasons, inferring function from 16S gene surveys (e.g. Langille et al. 2013) is less satisfactory than genomic and metagenomic data. The second issue relates to the key limitation of genomic data, that these data provide the genetic capacity for function, and the realized capacity is dictated by gene expression, enzyme activity and pattern of flux through the metabolic network of individual microbial cells and the microbial community (Heintz-Buschart & Wilmes 2018). In microbiomes, as in other complex microbial communities, the metabolic traits of individual bacterial taxa can be strongly dependent on the identity and metabolic activity of other co-occurring microorganisms, such that the metabolic function of any taxon can be resolved most effectively by a community approach (e.g. Fischer et al. 2017; Douglas 2020; Henriques et al. 2020; McMullen et al. 2020).

We conclude by considering how this study informs our understanding of metabolic trait distribution among members of animal gut microbiomes. The metabolic network of animal gut microbiomes is influenced by diet, host, and co-occurring microorganisms. By identifying the microorganisms that mediate different functions and their evolutionary history, we can have a basis to understand and predict microbiome functions. This is the basis for rationally-designed routes to manipulate microbiomes for treatment of metabolic disease and application of probiotics. This study supports previously identified trends in complex microbiome systems, such as the human microbiome (e.g. Bauer et al. 2015), but highlights the benefits of the *Drosophila* gut microbiome system by having readily culturable taxa. The

identification of variation in metabolic functions at different phylogenetic scales in this study provides the basis for future studies to determine the ecology and evolution of microbiome functions of *Drosophila* in natural settings.

References

Adair KL, Wilson M, Bost A, Douglas AE. 2018. Microbial community assembly in wild populations of the fruit fly *Drosophila melanogaster*. ISME Journal. 12:959–972. doi: 10.1038/s41396-017-0020-x.

Adeolu M, Alnajar S, Naushad S, S. Gupta R. 2016. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. International Journal of Systematic and Evolutionary Microbiology. 66:5575–5599. doi: 10.1099/ijsem.0.001485.

Allison SD, Martiny JBH. 2008. Resistance, resilience, and redundancy in microbial communities. PNAS. 105:11512–11519. doi: 10.1073/pnas.0801925105.

Ankrah NYD, Douglas AE. 2018. Nutrient factories: metabolic function of beneficial microorganisms associated with insects. Environmental Microbiology. 20:2002–2011. doi: 10.1111/1462-2920.14097.

Baek JH, Kim KH, Moon JY, Yeo S-H, Jeon CO. 2020. *Acetobacter oryzoeni* sp. nov., isolated from Korean rice wine vinegar. International Journal of Systematic and Evolutionary Microbiology,. 70:2026–2033. doi: 10.1099/ijsem.0.004008.

Bankevich A et al. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 19:455–477. doi: 10.1089/cmb.2012.0021.

Bauer E, Laczny CC, Magnusdottir S, Wilmes P, Thiele I. 2015. Phenotypic differentiation of gastrointestinal microbes is reflected in their encoded metabolic repertoires. Microbiome. 3:1–13. doi: 10.1186/s40168-015-0121-6.

Blum, JE, Fischer CN, Miles J, Handelsman J. 2013. Frequent replenishment sustains the beneficial microbiome of *Drosophila melanogaster*. mBio. 4: e00860-13. https://doi.org/10.1128/mBio.00860-13.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 27:578–579. doi: 10.1093/bioinformatics/btq683.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120. doi: 10.1093/bioinformatics/btu170.

Bost A et al. 2018. How gut transcriptional function of *Drosophila melanogaster* varies with the presence and composition of the gut microbiota. Molecular Ecology. 27:1848–1859. doi: 10.1111/mec.14413.

Bradley PH, Pollard KS. 2017. Proteobacteria explain significant functional variability in the human gut microbiome. Microbiome. 5:1–23. doi: 10.1186/s40168-017-0244-z.

Brettin T et al. 2015. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep. 5. doi: 10.1038/srep08365.

Brockhurst MA et al. 2019. The Ecology and Evolution of Pangenomes. Current Biology. 29:R1094–R1103. doi: 10.1016/j.cub.2019.08.012.
Broderick NA, Lemaitre B. 2012. Gut-associated microbes of *Drosophila melanogaster*. Gut Microbes. 3:307–321. doi: 10.4161/gmic.19896.
Castresana J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol Biol Evol. 17:540–552. doi: 10.1093/oxfordjournals.molbev.a026334.
Chand D et al. 2017. Molecular features of bile salt hydrolases and relevance in human health. Biochimica et Biophysica Acta (BBA) - General Subjects. 1861:2981–2991. doi: 10.1016/j.bbagen.2016.09.024.
Chandler JA, Lang JM, Bhatnagar S, Eisen JA. 2011. Bacterial Communities of Diverse Drosophila Species: Ecological Context of a Host – Microbe Model System. 7. doi: 10.1371/journal.pgen.1002272.
Chaston JM, Newell PD, Douglas AE. 2014. Metagenome-wide association of microbial determinants of host phenotype in *Drosophila melanogaster*. mBio. 5:1–12. doi: 10.1128/mBio.01631-14.
Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol. 65:997–1008. doi: 10.1093/sysbio/syw037.
Choi JI, Yoon K, Subbammal Kalichamy S, Yoon S-S, Il Lee J. 2016. A natural odor attraction between lactic acid bacteria and the nematode *Caenorhabditis elegans*. The ISME Journal. 10:558–567. doi: 10.1038/ismej.2015.134.
Cold Spring Harbor. 2018. PBS (pH 7.4). Cold Spring Harbor Protocols. 2018:pdb.rec099085-pdb.rec099085. doi: 10.1101/pdb.rec099085.
Consuegra J et al. 2020. *Drosophila*-associated bacteria differentially shape the nutritional requirements of their host during juvenile growth. PLOS Biology. 18:e3000681. doi: 10.1371/journal.pbio.3000681.
Cox CR, Gilmore MS. 2007. Native microbial colonization of *Drosophila melanogaster* and its use as a model of *Enterococcus faecalis* pathogenesis. Infect Immun. 75:1565–1576. doi: 10.1128/IAI.01496-06.
Cribari-Neto F, Zeileis A. 2010. Beta Regression in R. Journal of Statistical Software. 34:1–24. doi: 10.18637/jss.v034.i02.
Douglas, AE. 2018. The *Drosophila* model for microbiome research. Lab Animal 47: 157–64. https://doi.org/10.1038/s41684-018-0065-0.
Douglas AE. 2019. Simple animal models for microbiome research. Nature Reviews Microbiology. 17:764–775. doi: 10.1038/s41579-019-0242-1.
Douglas AE. 2020. The microbial exometabolome: ecological resource and architect of microbial communities. Philosophical Transactions of the Royal Society B: Biological Sciences. 375:20190250–20190250. doi: 10.1098/rstb.2019.0250.
Douglass AP et al. 2019. Coverage-versus-length plots, a simple quality control step for *de novo* yeast genome sequence assemblies. G3. 9:879–887. doi: 10.1534/g3.118.200745.
Duar, R.M., Lin, X.B., Zheng, J., Martino, M.E., Grenier, T., Perez-Munoz, M.E., Leulier, F., Ganzle, M., and Walter, J. 2017. Lifestyles in transition: evolution

and natural history of the genus *Lactobacillus*. FEMS Microbiol Rev *41*, S27-S48.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797. doi: 10.1093/nar/gkh340.
Ellegaard KM, Engel P. 2016. Beyond 16S rRNA community profiling: intra-species diversity in the gut microbiota. Front. Microbiol. 7. doi: 10.3389/fmicb.2016.01475.
Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology. 20:238. doi: 10.1186/s13059-019-1832-y.
Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology. 16:157. doi: 10.1186/s13059-015-0721-2.
Engel P, Moran NA. 2013. The gut microbiota of insects - diversity in structure and function. FEMS Microbiology Reviews. 37:699–735. doi: 10.1111/1574-6976.12025.
Erkosar B, Storelli G, Defaye A, Leulier F. 2013. Host-intestinal microbiota mutualism: 'learning on the fly'. Cell Host and Microbe. 13:8–14. doi: 10.1016/j.chom.2012.12.004.
Fan XG, Li TG. 1997. Growth of *Helicobacter pylori* in candle jars. Journal of Medical Microbiology,. 46:354–355. doi: 10.1099/00222615-46-4-354.
Fischer CN, Trautman EP, Crawford JM, Stabb EV, Handelsman J, and Broderick NA. 2017. Metabolite exchange between microbiome members produces compounds that influence *Drosophila* behavior. eLife *6*.
Fox J, Weisberg S. 2019. *A Companion to Applied Regression*. Third edit. Thousand Oaks, CA https://socialsciences.mcmaster.ca/jfox/Books/Companion/.
Fukatsu T, Nikoh N. 1998. Two Intracellular Symbiotic Bacteria from the Mulberry Psyllid *Anomoneura mori* (Insecta, Homoptera). Appl Environ Microbiol. 64:3599–3606.
Galac MR, Lazzaro BP. 2012. Comparative genomics of bacteria in the genus *Providencia* isolated from wild *Drosophila melanogaster*. BMC Genomics. 13:612. doi: 10.1186/1471-2164-13-612.
Galac MR, Lazzaro BP. 2011. Comparative pathology of bacteria in the genus *Providencia* to a natural host, *Drosophila melanogaster*. Microbes and Infection. 13:673–683. doi: 10.1016/j.micinf.2011.02.005.
Galili T. 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics. 31:3718–3720. doi: 10.1093/bioinformatics/btv428.
Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. 2006. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics. 22:2695–2696. doi: 10.1093/bioinformatics/btl461.
Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29:1072–1075. doi: 10.1093/bioinformatics/btt086.
Hammer AJ, Walters A, Carroll C, Newell PD, Chaston JM. 2017. Draft genome

sequence of *Lactobacillus paracasei* DmW181, a bacterium isolated from wild *Drosophila*. Genome Announcements 5. https://doi.org/10.1128/genomeA.00545-17.

Heintz-Buschart A, Wilmes P. 2018. Human Gut Microbiome: Function Matters. Trends in Microbiology. 26:563–574. doi: 10.1016/j.tim.2017.11.002.

Henriques, S.F., Dhakan, D.B., Serra, L., Francisco, A.P., Carvalho-Santos, Z., Baltazar, C., Elias, A.P., Anjos, M., Zhang, T., Maddocks, O.D.K., and Ribeiro, C. (2020). Metabolic cross-feeding in imbalanced diets allows gut microbes to improve reproduction and alter host behaviour. Nature Commununications 11, 4236.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecualr Biology and Evolution. 35:518–522. doi: 10.1093/molbev/msx281.

Hooper LV, Midtvedt T, Gordon JI. 2002. How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annual Review of Nutrition. 22:283–307. doi: 10.1146/annurev.nutr.22.011602.092259.

Huang J-H, Jing X, Douglas AE. 2015. The multi-tasking gut epithelium of insects. Insect Biochemistry and Molecular Biology. 1–6. doi: 10.1016/j.ibmb.2015.05.004.

Huerta-Cepas J et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47:D309–D314. doi: 10.1093/nar/gky1085.

Hug, LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, et al. 2016. A New View of the Tree of Life. *Nature Microbiology* 1: 1–6. https://doi.org/10.1038/nmicrobiol.2016.48.

Huttenhower C et al. 2012. Structure, function and diversity of the healthy human microbiome. Nature. 486:207–214. doi: 10.1038/nature11234.

Inamine H, Ellner SP, Newell PD, Luo Y, Buchon N, Douglas AE. 2018 Spatiotemporally heterogeneous population dynamics of gut bacteria inferred from fecal time series data. MBio 9: e01453-17. https://doi.org/10.1128/mBio.01453-17.

Jeon HH et al. 2017. A proposal of *Leuconostoc mesenteroides* subsp. *jonggajibkimchii* subsp. nov. and reclassification of *Leuconostoc mesenteroides* subsp. *suionicum* (Gu et al., 2012) as *Leuconostoc suionicum* sp. nov. based on complete genome sequences. International Journal of Systematic and Evolutionary Microbiology. 67:2225–2230. doi: 10.1099/ijsem.0.001930.

Jombart T, Balloux F, Dray S. 2010. adephylo: new tools for investigating the phylogenetic signal in biological traits. Bioinformatics. 26:1907–1909. doi: 10.1093/bioinformatics/btq292.

Jones BV, Begley M, Hill C, Gahan CGM, Marchesi JR. 2008. Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. PNAS. 105:13580–13585. doi: 10.1073/pnas.0804437105.

Judd AM et al. 2018. Bacterial Methionine Metabolism Genes Influence *Drosophila melanogaster* Starvation Resistance Stabb, EV, editor. Appl Environ

Microbiol. 84:e00662-18, /aem/84/17/e00662-18.atom. doi: 10.1128/AEM.00662-18.
Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods. 14:587–589. doi: 10.1038/nmeth.4285.
Kang D, Douglas AE. 2020. Functional traits of the gut microbiome correlated with host lipid content in a natural population of *Drosophila melanogaster*. Biol. Lett. 16:20190803. doi: 10.1098/rsbl.2019.0803.
Koeppel AF, Wu M. 2013. Surprisingly extensive mixed phylogenetic and ecological signals among bacterial operational taxonomic units. Nucleic Acids Res. 41:5175–5188. doi: 10.1093/nar/gkt241.
Kumar M et al. 2012. Cholesterol-lowering probiotics as potential biotherapeutics for metabolic diseases. Experimental Diabetes Research. 2012:e902917. doi: https://doi.org/10.1155/2012/902917.
Kurtz S et al. 2004. Versatile and open software for comparing large genomes. Genome Biology. 5:R12. doi: 10.1186/gb-2004-5-2-r12.
Langille MGI. 2018. Exploring linkages between taxonomic and functional profiles of the human microbiome. mSystems. 3. doi: 10.1128/mSystems.00163-17.
Langille, MGI et al. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology 31: 814–21. https://doi.org/10.1038/nbt.2676.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357–359. doi: 10.1038/nmeth.1923.
Lee I et al. 2017. ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. International Journal of Systematic and Evolutionary Microbiology. 67:2053–2057. doi: 10.1099/ijsem.0.001872.
Lemaitre, B., and Miguel-Aliaga, I. (2013). The digestive tract of *Drosophila melanogaster*. Annual review of genetics *47*, 377-404.
Li H et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25:2078–2079. doi: 10.1093/bioinformatics/btp352.
Lievens B et al. 2015. Microbiology of sugar-rich environments: diversity, ecology and system constraints. Environmental Microbiology. 17:278–298. doi: 10.1111/1462-2920.12570.
Lladó Fernández S, Větrovský T, Baldrian P. 2019. The concept of operational taxonomic units revisited: genomes of bacteria that are regarded as closely related are often highly dissimilar. Folia Microbiol. 64:19–23. doi: 10.1007/s12223-018-0627-y.
Louca S et al. 2018. Function and functional redundancy in microbial systems. Nat Ecol Evol. 2:936–943. doi: 10.1038/s41559-018-0519-1.
Lozupone C, Stombaugh J, Gordon JI, Jansson J, Knight R. 2012. Diversity, stability and resilience of the human gut microbiota. Nature. 489:220–230.
Maayer PD, Aliyu H, Cowan DA. Reorganising the order Bacillales through phylogenomics. Systematic and Applied Microbiology 42: 178–89. https://doi.org/10.1016/j.syapm.2018.10.007.
Martin MG, Magni C, Mendoza D de, López P. 2005. CitI, a transcription factor

involved in regulation of citrate metabolism in lactic acid bacteria. Journal of Bacteriology. 187:5146–5155. doi: 10.1128/JB.187.15.5146-5155.2005.
Martino ME et al. 2016. Nomadic lifestyle of *Lactobacillus plantarum* revealed by comparative genomics of 54 strains isolated from different habitats. Environmental Microbiology. 18:4974–4989. doi: 10.1111/1462-2920.13455.
Matsutani M, Hirakawa H, Yakushi T, Matsushita K. 2011. Genome-wide phylogenetic analysis of *Gluconobacter*, *Acetobacter*, and *Gluconacetobacter*. FEMS Microbiol Lett. 315:122–128. doi: 10.1111/j.1574-6968.2010.02180.x.
Matsutani M Yakushi T. 2018. Pyrroloquinoline quinone-dependent dehydrogenases of acetic acid bacteria. Applied Microbiology and Biotechnology 102: 9531–40. https://doi.org/10.1007/s00253-018-9360-3.
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., *et al*. (2013). Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci U S A *110*, 3229-3236.
McMullen II JG, Peters-Schulze G, Cai J, Patterson AD, Douglas AE. 2020. How gut microbiome interactions affect nutritional traits of *Drosophila melanogaster*. Journal of Experimental Biology *in press*.
Newell PD et al. 2014. In vivo function and comparative genomic analyses of the *Drosophila* gut microbiota identify candidate symbiosis factors. Frontiers in Microbiology. 5:1–15. doi: 10.3389/fmicb.2014.00576.
Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 32:268–274. doi: 10.1093/molbev/msu300.
Nicholson JK et al. 2012. Host-gut microbiota metabolic interactions. Science. 336:1262–1267. doi: 10.1126/science.1223813.
Obadia B, Keebaugh ES, Yamada R, Ludington WB, Ja WW. 2018. Diet influences host–microbiota associations in *Drosophila*. Proceedings of the National Academy of Sciences 115: E4547–48. https://doi.org/10.1073/pnas.1804948115.
Oksanen J et al. 2019. vegan: Community Ecology Package. R package version 2.5-5. https://cran.r-project.org/package=vegan.
Overbeek R et al. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42:D206–D214. doi: 10.1093/nar/gkt1226.
Page AJ et al. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 31:3691–3693. doi: 10.1093/bioinformatics/btv421.
Pais IS, Valente RS, Sporniak M, Teixeira L. 2018. *Drosophila melanogaster* establishes a species-specific mutualistic interaction with stable gut-colonizing bacteria. PLOS Biology 16: e2005710. https://doi.org/10.1371/journal.pbio.2005710.
Paradis E. 2010. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics. 26:419–420. doi: 10.1093/bioinformatics/btp696.
Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and

evolutionary analyses in R. Bioinformatics. 35:526–528. doi: 10.1093/bioinformatics/bty633.
Petkau K, Fast D, Duggal A, Foley E. 2016. Comparative evaluation of the genomes of three common *Drosophila*-associated bacteria. Biology Open. 5:1305–1316. doi: 10.1242/bio.017673.
Qiao H, Keesey IW, Hansson BS, Knaden M. 2019. Gut microbiota affects development and olfactory behavior in *Drosophila melanogaster*. Journal of Experimental Biology. 222. doi: 10.1242/jeb.192500.
R Core Team. 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria https://www.r-project.org/.
Read MN, Holmes AJ. 2017. Towards an integrative understanding of diet-host-gut microbiome interactions. Frontiers in Immunology. 8:1–9. doi: 10.3389/fimmu.2017.00538.
Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution. 3:217–223. doi: 10.1111/j.2041-210X.2011.00169.x.
Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. PNAS. 106:19126–19131. doi: 10.1073/pnas.0906412106.
Ricks NJ, Carroll C, Walters A, Newell PD, Chaston JM. 2017. The genome sequence of *Weissella cibaria* DmW_103, isolated from wild *Drosophila*." Genome Announcements 5. https://doi.org/10.1128/genomeA.00512-17.
Rolhion N, Chassaing B. 2016. When pathogenic bacteria meet the intestinal microbiota. Philosophical Transactions of the Royal Society B: Biological Sciences. 371. doi: 10.1098/rstb.2015.0504.
Sainz F, Mas A, Torija MJ. 2017. Effect of ammonium and amino acids on the growth of selected strains of *Gluconobacter* and *Acetobacter*." International Journal of Food Microbiology 242: 45–52. https://doi.org/10.1016/j.ijfoodmicro.2016.11.006.
Salazar G. 2020. EcolUtils: Utilities for community ecology analysis. R package version 0.1. https://github.com/GuillemSalazar/EcolUtils.
Salvetti E, Harris HMB, Felis GE, O'Toole PW. 2018. Comparative Genomics of the Genus *Lactobacillus* Reveals Robust Phylogroups That Provide the Basis for Reclassification. Appl. Environ. Microbiol. 84. doi: 10.1128/AEM.00993-18.
Sannino DR, Dobson AJ, Edwards K, Angert ER, Buchon N. 2018. The *Drosophila melanogaster* gut microbiota provisions thiamine to its host. mBio. 9. doi: 10.1128/mBio.00155-18.
Schliep KP. 2011. phangorn: phylogenetic analysis in R. Bioinformatics. 27:592–593. doi: 10.1093/bioinformatics/btq706.
Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 30:2068–2069. doi: 10.1093/bioinformatics/btu153.
Shin SC et al. 2011. *Drosophila* microbiome modulates host developmental and metabolic homeostasis via insulin signaling. Science. 334:670–4. doi: 10.1126/science.1212782.

Sommer F, Bäckhed F. 2013. The gut microbiota-masters of host development and physiology. Nature Reviews Microbiology. 11:227–238. doi: 10.1038/nrmicro2974.
Storelli G et al. 2018. *Drosophila* perpetuates nutritional mutualism by promoting the fitness of its intestinal symbiont *Lactobacillus plantarum*. Cell Metab. 27:362-377.e8. doi: 10.1016/j.cmet.2017.11.011.
Suzuki R, Terada Y, Shimodaira H. 2019. pvclust: hierarchical clustering with p-values via multiscale bootstrap resampling. R package version 2.2-0. https://CRAN.R-project.org/package=pvclust.
Tettelin H et al. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome". PNAS. 102:13950–13955. doi: 10.1073/pnas.0506758102.
Thaiss CA, Zmora N, Levy M, Elinav E. 2016. The microbiome and innate immunity. Nature. 535:65–74. doi: 10.1038/nature18847.
Tofalo, R., Cocchi, S., and Suzzi, G. (2019). Polyamines and gut microbiota. Frontiers in Nutriton 6. https://doi.org/10.3389/fnut.2019.00016
Turkiewicz J, Jin Ryu H, Mayer EA. 2019. Gut microbes and behavior. Current Opinion in Behavioral Sciences. 28:72–77. doi: 10.1016/j.cobeha.2019.02.004.
Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics. 27:171–180. doi: 10.1111/j.1096-0031.2010.00329.x.
Van Rossum T, Ferretti P, Maistrenko OM, Bork P. 2020. Diversity within species: interpreting strains in microbiomes. Nat Rev Microbiol. doi: 10.1038/s41579-020-0368-1.
Vatanen T et al. 2019. Genomic variation and strain-specific functional adaptation in the human gut microbiome during early life. Nature Microbiology. 4:470–479. doi: 10.1038/s41564-018-0321-5.
Visconti, A., Le Roy, C.I., Rosa, F., Rossi, N., Martin, T.C., Mohney, R.P., Li, W., de Rinaldis, E., Bell, J.T., Venter, J.C., *et al*. (2019). Interplay between the human gut microbiome and host metabolism. Nat Commun *10*, 4505.
Winans NJ et al. 2017. A genomic investigation of ecological differentiation between free-living and *Drosophila*-associated bacteria. 4536–4550. doi: 10.1111/mec.14232.
Walters AW et al. 2020. The microbiota influences the *Drosophila melanogaster* life history strategy. Molecular Ecology 29: 639–53. https://doi.org/10.1111/mec.15344.
Wang Y, Kapun M, Waidele L, Kuenzel S,. Bergland AO, Staubach F. 2020. Common structuring principles of the *Drosophila melanogaster* microbiome on a continental scale and between host and substrate. Environmental Microbiology Reports 12: 220–28. https://doi.org/10.1111/1758-2229.12826.
Wright SM, Carroll C, Walters A, Newell PD, Chaston JM. 2017. Genome sequence of *Leuconostoc citreum* DmW_111, isolated from wild *Drosophila*. Genome Announcements 5. https://doi.org/10.1128/genomeA.00507-17.
Wong ACN, Vanhove AS, Watnick PI. 2016. The interplay between intestinal bacteria and host metabolism in health and disease: lessons from *Drosophila*

melanogaster. Disease Models & Mechanisms. 9. doi: 10.1242/dmm.023408.
Wong ACN, Chaston JM, Douglas AE. 2013. The inconstant gut microbiota of *Drosophila* species revealed by 16S rRNA gene analysis. The ISME Journal 7: 1922–32. https://doi.org/10.1038/ismej.2013.86.
Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution. 8:28–36. doi: 10.1111/2041-210X.12628.
Yukphan P et al. 2020. *Gluconobacter aidae* sp. nov., an acetic acid bacteria isolated from tropical fruits in Thailand. International Journal of Systematic and Evolutionary Microbiology. 70:4351–4357. doi: 10.1099/ijsem.0.004292.
Zeileis A, Hothorn T. 2002. Diagnostic checking in regression relationships. R News. 2:7–10.
Zheng H et al. 2019. Division of labor in honey bee gut microbiota for plant polysaccharide digestion. Proceedings of the National Academy of Sciences of the United States of America. 116:25909–25916. doi: 10.1073/pnas.1916224116.
Zheng J et al. 2020. A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. International Journal of Systematic and Evolutionary Microbiology,. 70:2782–2858. doi: 10.1099/ijsem.0.004107.

CHAPTER 3

HOW GUT MICROBIOME INTERACTIONS AFFECT NUTRITIONAL TRAITS OF *DROSOPHILA MELANOGASTER*[2]

Abstract

Most research on the impact of the gut microbiome on animal nutrition is designed to identify the effects of single microbial taxa and single metabolites of microbial origin, without considering the potentially complex network of interactions among co-occurring microorganisms. Here, we investigate how different microbial associations and their fermentation products affect host nutrition, using *Drosophila melanogaster* colonized with three gut microorganisms (the bacteria *Acetobacter fabarum and Lactobacillus brevis* and the yeast *Hanseniaspora uvarum*) in all seven possible combinations. Some microbial effects on host traits could be attributed to single taxa (e.g. yeast-mediated reduction of insect development time), while other effects were sex-specific and driven by among-microbe interactions (e.g. male lipid content determined by interactions between the yeast and both bacteria). Parallel analysis of nutritional indices of microbe-free flies administered different microbial fermentation products (acetic acid, acetoin, ethanol and lactic acid) revealed a single consistent effect: that the lipid content of both male and female flies is reduced by acetic acid.

[2] Presented with minor modifications from the originally published article:
McMullen, J.G., Peters-Schulze, G., Cai, J., Patterson, A.D., and Douglas A.E. (2020). How gut microbiome interactions affect nutritional traits of Drosophila melanogaster. 223(19): jeb227843.
All of the supplementary material is found in Appendix B, and data can be accessed at: https://doi.org/10.5061/dryad.ngf1vhhrj.
Peters-Schulze, G. performed the CAFE experiments and Cai, J. conducted the short-chain fatty acid quantification assay.

This effect was recapitulated in male flies colonized with both yeast and *Acetobacter*, but not for any microbial treatment in females nor in males with other microbial complements. These data suggest that the effect of microbial fermentation products on host nutritional status is strongly context-dependent, with respect to both the combination of associated microorganisms and host sex. Taken together, our findings demonstrate that among-microbe interactions can play a critically important role in determining the physiological outcome of host-microbiome interactions in *Drosophila* and, likely, in other animal hosts.

Introduction

The gut of many animals is colonized by a diverse community of microorganisms (collectively known as the microbiome) that can influence the health and fitness of the animal host. The gut microbiome contributes to various host functions, including the degradation of dietary constituents, synthesis of essential nutrients, modulation of host immunity, and protection against pathogens (Herp et al., 2019; Hooper et al., 2012; Huang et al., 2015; Karasov et al., 2011; Read and Holmes, 2017; Rolhion and Chassaing, 2016; Sommer and Bäckhed, 2013; Thaiss et al., 2016; Wong et al., 2016). Numerous studies have linked specific microorganisms and microbial metabolites with host traits (e.g. Morton et al., 2019; Tripathi et al., 2018). However, it is increasingly recognized that the traits of individual microorganisms can be influenced by interactions with other members of the microbiome (Douglas, 2020; Kundu et al., 2019; Noecker et al., 2019). Theoretical models have predicted that among-microbe interactions, which range from beneficial to antagonistic and are widespread

throughout microbial communities, can have varying levels of metabolic dependencies and cross-feeding of metabolites between taxa (Coyte et al., 2015; Freilich et al., 2011; Levy and Borenstein, 2013; Magnúsdóttir et al., 2017; Noecker et al., 2019; Zelezniak et al., 2015). These interactions can lead to the net production of metabolites that none of the individual taxa in the community are able to synthesize in isolation (Wintermute and Silver, 2010). As a result, data for associations between animals and single microbial taxa may not accurately predict how individual taxa perform in a complex community.

A valuable approach to investigate among-microbial interactions is to use *in vitro* and simplified gnotobiotic animal systems with pairs or small groups of microbial taxa. For example, some *Bacteriodes* spp. in the human gut degrade dietary polysaccharides and the simple carbohydrates liberated support the growth of co-occurring taxa and promote the release of metabolites that influence human health (Mahowald et al., 2009; Rakoff-nahoum et al., 2016; Rodriguez-Castaño et al., 2019). In the lower termite gut, folate-producing bacteria release 5-formyl-tetrahydrofolate and the spirochete *Treponema primitia* uses this metabolite as a co-factor to produce acetic acid, an energy source for the insect host (Graber and Breznak, 2005). In the honey bee gut microbiome, numerous candidate cross-feeding events have been identified between the bacteria *Bifidobacterium* spp*., Gilliamella apicola*, and *Snodgrassella alvi*, resulting in improved digestion of the pollen diet and utilization of bacterial by-products by co-occurring taxa (Kešnerová et al., 2017; Zheng et al., 2019).

The focus of this study is the impact of among-microbe interactions on the

nutritional physiology of the fruit fly *Drosophila melanogaster*, which is a fast-emerging model for gut-microbiome interactions (Broderick and Lemaitre, 2012; Douglas, 2019; Erkosar et al., 2013; Wong et al., 2016). Three taxonomic groups, the bacteria of the family Acetobacteraceae and order Lactobacillales, and the Sacchromycetales yeasts, are the dominant taxa in the *Drosophila* gut microbiome (Adair et al., 2018; Chandler et al., 2011; Chandler et al., 2012; Quan and Eisen, 2018; Wong et al., 2011). To date, most research has focused either on the bacteria or yeast partners, with evidence that co-associations involving between two and five bacterial species can recapitulate the effects of conventional multi-species communities on certain traits under laboratory conditions (Gould et al., 2018; Newell and Douglas, 2014; Rohlfs and Kürschner, 2010). For example, *D. melanogaster* reared in co-association with pairs of *Acetobacter* and *Lactobacillus* species likely display cooperative metabolism and lower fly lipid content relative to microbe-free flies (Consuegra et al., 2020; Newell and Douglas, 2014; Sommer and Newell, 2019). Indications that interactions between bacteria and yeasts may be important determinants of *Drosophila* traits come largely from a single study (Fischer et al., 2017), which demonstrated that female flies prefer to both feed and lay eggs on substrate bearing a *Saccharomyces-Acetobacter* co-culture relative to mono-cultures and an *Acetobacter* strain unable to produce acetic acid. Furthermore, research on both *Acetobacter-Lactobacillus* interactions and *Acetobacter-Saccharomyces* interactions (Fischer et al., 2017; Sommer and Newell, 2019) has identified four microbial metabolites as candidate mediators of among-microbe and microbe-host interactions: acetic acid, acetoin, ethanol, and lactic acid. Importantly, these four metabolites have

also been implicated in the regulation of *Drosophila* metabolism, immunity, and behavior (Devineni and Heberlein, 2013; Farine et al., 2017; Fry, 2014; Hang et al., 2014; Hoffmann and Parsons, 1984; Iatsenko et al., 2018; Kamareddine et al., 2018; Shin et al., 2011).

The specific goal of this study was to understand whether and how among-microbe interactions may influence *Drosophila* performance and nutrient allocation. We used three representative strains of gut microorganisms: two bacteria, *Acetobacter fabarum* and *Lactobacillus brevis,* widely used in *Drosophila* microbiome research (Dobson et al., 2015; Sommer and Newell, 2019; White et al., 2018); and the yeast, *Hanseniaspora uvarum,* a prevalent species in wild *Drosophila* populations (Chandler et al., 2012; De Camargo and Phaff, 1957) that has also been used for laboratory experiments (Hoang et al., 2015; Murgier et al., 2019; Palanca et al., 2013; Scheidler et al., 2015; Solomon et al., 2019). We administered the microorganisms combinatorially to give associations with one, two, or all three taxa. The *Drosophila* in these treatments are described as gnotobiotic, meaning that they have a defined complement of microorganisms. In parallel, we quantified for the effect of four microbial fermentation products, acetic acid, acetoin, ethanol, and lactic acid, on nutrient allocation in flies reared under microbiologically-sterile conditions (known as axenic flies) and linked these results to the fermentation product profiles in the axenic and gnotobiotic flies. This study demonstrates how non-additive among-microbe effects can influence *Drosophila* performance and nutrition, and it establishes the contribution of individual microbial metabolites in the observed among-microbe interactions.

Materials and methods

Insects and microorganisms

The three microorganisms were: *Acetobacter fabarum* DsW54 (ACE), *Lactobacillus brevis* DmCS_003 (LAC) and *Hanseniaspora uvarum* FYH04C (YST). Details of their provenance and culture protocols are provided in Table 2.1.

The stock culture of *D. melanogaster* strain Canton S (*Wolbachia*-free) was maintained at 25°C with 50% relative humidity in a 12 h:12 h light:dark cycle on a yeast-glucose (Y-G) diet comprising 10% inactive brewer's yeast (MP Biomedicals), 10% glucose (Sigma), and 1.2% *Drosophila* type II agar (Apex) with preservatives, 0.04% phosphoric acid (Fisher Scientific) and 0.42% propionic acid (Fisher Scientific).

Axenic flies were prepared by a standard protocol for egg dechorionation (Koyle et al., 2016). Briefly, flies from the stock culture were allowed to oviposit for 16-18 hours on grape juice agar plates (Y-G diet with 1% *Drosophila* agar and approx. 15% Welch's grape juice concentrate), and the deposited eggs were thrice washed in 0.6% hypochlorite solution for 5 min, followed by three rinses in sterile deionized water. Approximately 60 eggs were aseptically transferred to 50 ml sterile conical Falcon tubes (Globe Scientific, Inc.) containing 7.5 ml autoclaved Y-G diet without preservatives, and they were raised under standard culture conditions to generate axenic flies.

Table 2.1. Microbial strains

Microbial strain		Origin	% glycerol for stock at -80°C	Culture media at 30°C	Incubation time		Calibration equations[1]
	Abbreviation				Solid medium for colony growth	Broth for mid-logarithmic culture	
Acetobacter fabarum DsW54	ACE	Gut of wild *Drosophila suzukii* (Winans et al., 2017)	20	mMRS[2], ambient oxygen	2 days	16 h with agitation	$C = 2x10^9 \bullet A - 1x10^8$
Lactobacillus brevis DmCS_003	LAC	Gut of laboratory culture of *Drosophila melanogaster* (Newell et al., 2014)	20	mMRS[2], reduced oxygen[3]	2 days	35 h without agitation	$C = 4x10^9 \bullet A + 2x10^8$
Hanseniaspora uvarum FYH04C	YST	Gut of wild *Drosophila melanogaster* from Ithaca, NY (2014)[4]	10	YPD[5], ambient oxygen	1 day	13 h with agitation	$C = 9x10^7 \bullet A + 3x10^7$

[1] Obtained by parallel quantification of OD and CFUs for 5 replicate tubes, each for 5 time points over 20 to 45 h growth (depending on the strain) of 5 ml broth culture with shaking at 200 rpm for *A. fabarum* and *H. uvarum* and no agitation for *L. brevis*. All calibration curves yielded $R^2 > 0.8$. C = CFU concentrations [CFU ml^{-1}]. A = optical density.
[2] modified De Man, Rogosa, and Sharpe medium: 1.25% bacto-proteose peptone, 0.75% yeast extract, 2% glucose, 0.5% sodium acetate, 0.2% dipotassium hydrogen phosphate, 0.2% tirammonium citrate, 0.02% magnesium sulphate heptahydrate, 0.005% manganese sulfate tetrahydrate, and 1.2% agar (for plates only). All ingredients are from Sigma, except bacto-proteose peptone from Becton Dickinson.
[3] high CO_2 conditions in a brewer's jar with a lit candle to consume most of the available oxygen.
[4]The *H. uvarum* isolate was identified with molecular characterization of the ITS sequence with Sanger sequencing (primers ITS1 (forward: 5'-CCGTAGGTGAACCTGCGG-3') and ITS4 (reverse: 5'-TCCTCCGCTTATTGATATGC-3') from White et al. (1990)). NCBI accession: MT217045. Note: the isolate originated from a female fly, and although male flies appeared to be *D. melanogaster*, female flies are indistinguishable from *D. simulans.*
[5]yeast-peptone-dextrose medium: 1% yeast extract, 2% bacto-peptone, 2% glucose, and 1.5% agar (for plates only). All ingredients are from Sigma, except bacto-peptone from Becton Dickinson.

To produce gnotobiotic insects colonized with one, two, or all three microbial strains, microorganisms were administered aseptically to Falcon tubes containing dechorionated eggs at a density of 5 x 10^6 cells (50 µl suspension of microorganism at 1 x 10^8 cells ml^{-1}) in phosphate-buffered saline (PBS) (Cold Spring Harbor, 2018). For treatments containing multiple microbial species, the taxa were added in equal proportion to give the same total concentration. To prepare the microorganisms, a loop of frozen glycerol stock was streaked onto solid medium and, following 1-2 days of growth, a single colony was picked and grown in 5 ml broth to mid-logarithmic phase (see Table 2.1 for details). Cell densities were determined by optical density at 600 nm (OD) on 96-well plates (Globe Scientific, Inc.) using an xMark Spectrophotometer (BioRad) and diluted to the required cell number based on calibration curves of OD against number of colony-forming units (CFUs) constructed for each strain (Table 2.1).

Experimental design

The eight treatments comprised axenic insects and seven gnotobiotic treatments of insects administered every combination of mono-, di-, and tri-association with ACE, LAC, and YST, with six replicates (i.e. Falcon tubes) per treatment. The insects were reared from dechorionated eggs under the standard culture conditions and scored twice daily (two- and ten-hours post dawn) for number of pupae and number of empty puparia to obtain time to pupation and eclosion, respectively. On the first day post-eclosion (dpe), the flies in each Falcon tube were aseptically transferred to a fresh Falcon tube containing autoclaved Y-G diet. At 5 dpe, the flies were anesthetized

using CO_2, separated by sex by visualizing morphological differences in genitalia, counted, and analyzed with one sample of each sex per Falcon tube for microbial load (two flies per sample) and nutritional indices (five flies per sample). In addition, ca. 20 mg (10-20 flies per sample) of each sex from two Falcon tubes per treatment were used for quantification of short chain fatty acids (SCFAs); the number of flies available for quantification varied with treatment. The weight of each sample used for nutritional indices and SCFA content was determined with a Mettler-Toledo balance, to an accuracy of 0.1 mg. All samples were flash frozen in liquid nitrogen and stored at -80°C. This experiment was replicated three times independently to generate a total of 18 replicates of males and females for each treatment.

Microbial load in flies

The abundance of each microbial strain per fly of each sex at 5 dpe was scored as number of colony-forming units (CFUs). Two male flies and two female flies from each Falcon tube were surface sterilized in 500 µl 0.3% hypochlorite solution in a sterile 1.5 ml centrifuge tube with gentle shaking by hand, followed by two rinses in sterile PBS. The flies were transferred to a sterile 2 ml screw-cap microcentrifuge tube containing 100 µl sterile lysis matrix D beads (MP Biomedicals) and 200 µl sterile PBS, homogenized using a FastPrep-24 instrument at 4.0 m/s for 30 s, diluted with 800 µl sterile PBS, and then inoculated in duplicate onto the appropriate medium (Table 2.1) using a WASP-2 spiral plater (Microbiology International). A Protocol3 instrument (Microbiology International) was utilized to enumerate CFUs after 2-3 days of growth at 30ºC, and the data were normalized to CFUs per fly. The detection

limit was 10 CFU fly^{-1}; this value was added for statistical analysis to eliminate zeros from microbial treatments in the dataset.

To facilitate the automated counting, samples from di- and tri-associations were plated onto media supplemented with antimicrobial(s) that selectively suppressed one of the microorganisms: 100 μg kanamycin sulfate ml^{-1} (Sigma) to suppress ACE, 10 μg ampicillin sodium salt ml^{-1} (Sigma) to suppress LAC, 500 μg methylparaben ml^{-1} (Apex) under high CO_2 on mMRS medium to suppress YST (Fig. S2.1A-D). Preliminary experiments confirmed that each antimicrobial did not suppress growth of the non-target microorganisms (Fig. S2.1E), apart from methylparaben which had a negative effect on ACE growth and was not used in plates selecting for ACE growth (Fig. S2.1B). For ACE and YST co-associations, colonies were differentiated by morphology: ACE has small orange/tan colonies and YST has large off-white colonies.

Negative controls comprised axenic flies plated onto mMRS and YPD agar without antimicrobials and incubated at 30°C for one week. Individual replicates were discarded if microbial contamination was present on plates for the entire experiment. Microbial biomass (dry weight) was estimated following the method of Norland et al. (1987), using estimates of biovolume (ACE = 1.1 μm^3, LAC = 1.7 μm^3 , YST = 38.7 μm^3) derived from published data (Cleenwerck et al., 2008; Gries and Ly, 2009; Cadez and Smith, 2011).

Nutritional indices

The lipid, carbohydrate, and protein contents of the flies were quantified by the

procedure of Newell and Douglas (2014) as indices of fly nutritional status. Briefly, each fly sample was homogenized in 125 µl ice-cold TET buffer (10 mM Tris pH 8, 1 mM EDTA, 0.1% Triton X-100) and 100 µl lysis matrix D beads, using a FastPrep-24 at 4 m/s for 45 s. For protein analysis, 10 µl was removed from the homogenate and diluted 1:6 for males and 1:7 for females in ice-cold TET buffer. The remaining homogenate was incubated at 72°C for 30 min to inactivate endogenous enzymes. Samples were stored at -80°C prior to quantification.

Protein content was determined using the DC Protein Assay Kit (Bio-Rad) according to manufacturer's protocol, with a standard curve using 0-1.4 mg bovine serum albumin ml^{-1}. For lipid content, the triglyceride (TAG) content was determined by incubating 5 µl heat-inactivated homogenate with 37.5 µl lipase (20 U ml^{-1} in 20 mM potassium phosphate, 20 mM EDTA, and 20 mM magnesium chloride, pH 7.5, Sigma L9518), followed by assay of liberated glycerol, using the Free Glycerol Reagent (Sigma-Aldrich F6428) following manufacturer's protocol with 0-3 mg triolein equivalent glycerol ml^{-1} for the standards. Absorbance attributed to endogenous free glycerol was subtracted from TAG for quantification. Glucose, trehalose, and glycogen were assayed with the glucose (GO) kit (Sigma GAGO20) following manufacturer's protocol, with 0-0.8 mg glucose ml^{-1}, 0-0.2 mg trehalose ml^{-1}, and 0-0.2 mg glycogen ml^{-1} as standards. For trehalose and glycogen determination, 2.5 µl trehalase (1 U ml^{-1}, Sigma T0167) and 5 µl amyloglucosidase (2 U ml^{-1} in 5 mM acetic acid, 5 mM sodium acetate, Sigma A7420), respectively, were added to samples and incubated for 1 h at 37 °C prior to quantification of glucose. For trehalose measurements, 2 µl 1 mM EDTA and 2 µl 0.2 M sodium citrate dihydrate were added

to samples and incubated for 10 min at 37 °C prior to the addition of the enzyme. All samples and standards were assayed in duplicate, and the absorbance values of the two technical replicates were averaged. Final absolute quantifications for all nutritional indices were normalized to fly weight.

Short chain fatty acid (SCFA) content

SCFAs were assayed using the propyl esterification derivatization method of Cai et al. (2017) on ca. 20 mg whole bodies of pooled flies for each sex, with quantification by an Agilent 7890A gas chromatograph coupled with an Agilent 5975 mass spectrometer (GC-MS, Agilent Technologies, Santa Clara, CA) at the Penn State Metabolomics Facility, University Park, PA. A targeted analysis was performed to quantify the following 12 SCFAs: acetate, propionate, butyrate, isobutyrate, isovalerate, valerate, 2-methylbutyrate, 2-methylpentanoate, 3-methylpentanoate, 2-methylhexanoate, 4-methylvalerate and heptanoate. Each fly sample was homogenized at 6,500 rpm for 60 s (Precellys, Bertin Technologies, Rockville, MD) in 1 ml 5 mM NaOH, with 10 µg caproic acid-6,6,6-d3 ml^{-1} (internal standard) using 1.0 mm diameter zirconia/silica beads (BioSpec, Bartlesville, OK), and centrifuged at 13,200 g for 20 min at 4 °C. 500 µl of sample supernatant was added to 500 µl 1-propanol:pyridine (v/v = 3:2), followed by an addition of 100 µl propyl chloroformate (esterification reagent). Then, samples were vortexed for 1 min, with a subsequent 1 h incubation at 60 °C. The derivatized samples were extracted with a two-step hexane extraction (300 µl + 200 µl) following the procedure of Zheng et al. (2013), yielding 500 µl, which was transferred to a glass autosampler vial. A standard curve was

generated for each analyte to quantify biological concentration (metabolite amount normalized to fly fresh weight [μmol g^{-1}]) of SCFAs. The data was range scaled to normalize SCFA titers across three experimental replicates; values were centered to the mean metabolite concentration and scaled by the range of metabolite quantified in each experimental replicate (van de Berg et al., 2006).

Dietary administration of fermentation products to flies

The effect of microbial fermentation products on fly nutritional indices was assayed using a modification of the CAFÉ method (Ja et al., 2007) to enable parallel quantification of food consumption in adult flies and remove effects of larval manipulation to the diet. The feeding chamber comprised a sterile 50 ml Falcon tube, containing a filter paper (1" x 2") wetted with 0.5 ml sterile deionized water to maintain humidity. Calibrated capillary tubes (1-5 μl, Drummond Scientific Co.) were aseptically filled to 5 μl with a chemically-defined liquid food. The diet followed the protocol of Piper et al. (2014) with the following modifications. First, the agar was omitted. Second, glucose was provided as sole sugar source, as in Y-G diet used in this study, but at 5% (w/v) because 10% (the concentration in Y-G diet) was viscous and supported very low feeding rates in the CAFÉ system (unpub. data). Third, the cholesterol was dissolved in 2.5:1 mmol ratio of cyclodextrin to cholesterol (replacing ethanol) for improved solubilization (Christian et al., 1997). Finally, acetic acid buffer was replaced by citric acid buffer, with pH adjusted to 4.8 using 10 M NaOH. Three capillaries were inserted into the lid of each Falcon tube. Pilot experiments confirmed that food consumption by the insects did not differ significantly between the liquid

holidic diet and liquid meridic diet comprising 5% glucose and 5% yeast extract (Fig. S2.2A) with high insect survival (> 80%; Fig. S2.2B) and small differences in nutrient allocation (Fig. S2.2C-E). We also compared fly traits on liquid and solid media (Fig. S2.2C-F), finding that flies on the liquid diets had lower or similar survival, TAG, and glucose content compared to the solid diets, while weight per fly also did not vary significantly, apart from elevated values for females on the solid meridic diet.

The experimental design of the metabolite administration assay tested one metabolite: acetic acid, acetoin, ethanol, and lactic acid, at three treatment levels: 0, 0.15, and 0.3 M based on concentrations used in Kim et al. (2018) for acetic acid effects on *Drosophila*, which was adopted for all four metabolites. The concentrations generally reflect published data on microbial production of these metabolites (Adler et al., 2014; Aranda-Díaz et al., 2020; Barata et al., 2012; Consuegra et al., 2020; Hall et al., 2018). Exceptionally, acetoin was administered at higher concentrations than reported in *Drosophila* cultures to enable comparisons to the other metabolites in the experiment (Adler et al., 2014; Barata et al., 2011). Three independent experiments were conducted for each metabolite, with male and female flies in separate replicate chambers for each experiment. All the chambers of a given diet were stored together in a sterile air-tight Snapware 23-cup container (Corelle Brands, Illinois, USA). To control for diet evaporation, one fly-free chamber was included in each container. The experiments used axenic flies that had been raised to three dpe, with six flies per chamber, incubated at 25°C with 12:12 h light:dark cycle. Each chamber was scored daily for four days for the number of live insects and the volume of food in each capillary tube (quantified from the liquid height to an accuracy of 0.1 µl). Capillary

tubes containing fresh diet were replaced daily. At day-4 of the experiment, all living flies from each feeding chamber were weighed, flash frozen in liquid nitrogen, and stored at -80°C prior to quantification of the protein, glucose, and TAG content as above, except protein samples were diluted 1:3 for males and 1:4 for females.

To calculate food consumption, the change in height in each of the three capillaries over the 24 h test period was determined, and then summed to obtain the total amount consumed. The change in height of the liquid column in the three capillaries in the negative control was also scored and subtracted from the experimental values to control for evaporation. To determine diet consumption per fly per day, the value per chamber was divided by the number of live flies at the end of each day (i.e. it is assumed that flies which died in each 24 h period did not feed over that period).

Statistics

All statistics were performed using R version 3.6.1 (R Core Team, 2018) with $\alpha = 0.05$ as the cut-off for statistical significance of model predictors. An analysis of variance (ANOVA) was implemented to examine microbial abundance, relative insect numbers, fly weight, nutritional indices, feeding rates, and SCFA content. All models used the 'lmer' function in the lme4 package (Bates et al., 2015) for mixed-effect models, expect for relative insect numbers as no random effect was required. Residuals from each model were visually assessed for normality and homoscedasticity. See supplemental tables for specific information on predictors and random effects included in models. Differences between fixed effects for the ANOVA models were determined by type II Wald F test with Kenward-Roger degrees of

freedom approximation using the 'Anova' function in the car package (Fox and Weisberg, 2019); type III method was used for feeding rates due to repeated measures. Either a *post hoc* Tukey's or Dunnett's test was implemented to discriminate between effect of predictors on response variables with the emmeans package (Lenth, 2019). The contribution of random effects in the model was tested by an analysis of deviance using the 'Anova' function in the car package. The Multi-Model Inference (MuMIn) package (Bartoń, 2019) was used to calculate the marginal and conditional R^2 values to assess model fit of fixed and random effects. For SCFA content analysis, effect sizes (ω^2) for each predictor and interactions were calculated from ANOVA table results provided by lmerTest (Kuznetsova et al., 2017) to determine degree of associations for fixed effects. Treatments were considered to have an effect on metabolite concentrations when $\omega^2 \geq 0.01$ (Cohen, 1988; Kirk, 1996). For the analysis of flies administered fermentation products, the best linear unbiased predictions were estimated for all response variables to obtain between-trial effect sizes with the 'ranef' function for the controls (without added metabolite) and were assessed for between-trial variation when significant results were found for analysis of deviance. Specifically, we simulated the posterior distributions (n = 10,000) from each model using the 'REsim' function in the merTools package (Knowles and Frederick, 2019).

Insect development time summary statistics (Kaplan-Meir method) were calculated using the 'survfit' function in the survival package (Therneau, 2015). A Cox mixed-effect model was used to assess the impact of administered microorganisms on pupation and eclosion rates with the coxme package (Therneau, 2019). Microbial treatments were coded as a categorical fixed effect predictor, and replicate (i.e. each

Falcon tube) was nested within each of the three experiments as a categorical random effect. Pairwise comparisons between all treatments were assessed by *post hoc* Tukey test using the multcomp package (Hothorn et al., 2008). An analysis of deviance was performed as ANOVA models.

The two-sample Kolmogorov-Smirnov test was used to assess differences in the cumulative distribution function of each microbial treatment against the axenic fly treatment; a Bonferroni correction was used for multiple tests (i.e. $\alpha = 0.05/7$). The skewness and kurtosis of each treatment were calculated with the DescTools package (Signorell et al., 2019); 95% confidence intervals (bias corrected and accelerated [BCa] method) were generated with 10,000 bootstrap replicates to enable comparisons among treatments. D'Agostino and Anscombe tests were performed to test for skewness and kurtosis, respectively, for each treatment using the moments package (Komsta and Novomestky, 2015); a Bonferroni correction was used for multiple tests (i.e. $\alpha = 0.05/8$).

To determine whether the elevated insect mortality in the treatments including YST reduced development time by selective death of slowly developing individuals, a re-sampling approach of the axenic eclosion data was implemented to impute the "missing" flies of the YST treatment using the 'sample' function (replace = TRUE). The missing flies comprised 347 insects (29% of 1195 axenic flies observed) that were randomly selected from the axenic eclosion distribution for each simulation (n = 5,000). Cox mixed-effect model was performed for each simulation to assess the proportion of significant results and changes in the hazard ratio. In addition, skewness and kurtosis were calculated for each simulation.

Nutritional indices were analyzed using several multivariate methods. First, principal component analysis (PCA) was implemented with a correlation matrix to visualize nutritional status of each fly replicate for each sex separately using the vegan package (Oksanen et al., 2019). TAG, glucose, and trehalose content of each sex were square root-transformed to reduce right-skew. Principal components were correlated with microbial abundance using 'envfit' function, and the resulting vectors were plotted onto PCA plots. Second, a permutational multivariate ANOVA (PERMANOVA) was performed using autoscaled data to identify the influence of microbiota on nutritional status for each sex. Presence of microbes was coded as three binary fixed effect predictors with dummy codes using the 'adonis' function in vegan with Euclidean distances (999 permutations). Then, ordination plots of female and male nutritional status were correlated with a Procrustean analysis and significance was assessed using a Procrustean randomization test ('protest' function in vegan with 999 permutations). Finally, a structural equation model (SEM) was used to identify how microbiota influenced each nutritional index with the piecewiseSEM 2.0.2 package (Lefcheck, 2016). Mixed effect linear models were constructed for the piecewise SEM analysis to determine the relationship for microbial abundance of each taxon and each nutritional index (see Fig. S2.4A for links tested). Indirect effects in SEMs were assessed by multiplying significant standardized coefficients of relationship between microbial abundances and the effect of individual taxa on a nutritional index.

Fly survival was investigated using a mixed effect logistic regression with the lme4 package using the 'glmer' function. Metabolite concentration and sex were included as

categorical fixed effects. Total food consumed per fly was included as continuous covariate in the CAFÉ experiments other than the method development experiments (feeding rates were not obtained for flies on solid media). Experimental replicate was included as categorical random effect with replicate (i.e. each Falcon tube) nested in experimental replicate for survival analysis to account for overdispersion by using replicate as an observation level random effect. A Wald's χ^2 test was performed to examine effect of predictors using the 'Anova' function. A *post hoc* Dunnett's test was used to compare each metabolite concentration to the control treatment for both sexes. An analysis of deviance was performed as ANOVA models.

A moderation analysis was performed to assess how each microorganism influenced the association between SCFA content and a given nutritional index. A mixed-effect multiple linear regression was performed for each sex separately with microbes included as three separate fixed effects with binary dummy codes and raw SCFA content as a continuous predictor with 'lmer' function. Experimental replicate was included as a categorical random effect. An ANOVA effect test table was generated as for SCFA content analysis. A Bonferroni correction was implemented to control for multiple tests (i.e. $\alpha = 0.05/30$). A follow-up analysis was performed to examine the effect of ACE and YST presence on acetic acid titer and TAG content in male flies, and LAC predictor was removed from the analysis as it had no strong moderation effect on acetic acid titer. The model was performed as before, however a categorical random effect of treatment was nested within experimental replicate. The function 'emtrends' was used to assess the moderation effect of ACE presence on simple slopes for treatments with and without YST. An analysis of deviance and R^2 values were

performed as ANOVA models.

Results

Microbial abundance

We, first, quantified the microbial abundance in the flies at 5 days post eclosion (dpe) with the three microorganisms (Table 2.1) in each of the 7 microbial treatments. The majority of flies sampled contained at least 10 cells of the microbial taxon that had been administered and no other microorganisms, except for the di-association treatment with ACE and LAC where LAC was below the detection limit, i.e. $\leq$10 CFU fly^{-1}, for all replicates (Fig. 2.1A). No microorganisms were detected in the axenic flies. The among-treatment variation differed significantly by sex for total microbial abundance, with female flies harboring more microbial cells than males in all treatments except the ACE/YST di-association (Table S2.1A). The effects of co-association on abundance of individual taxa varied among the microorganisms and were predominantly negative (Fig. 2.1A & B). The abundance of both LAC and YST was reduced by co-association with ACE, and both ACE and LAC were also significantly suppressed by YST. These negative effects were also evident in the tri-association, although YST ameliorated the negative effect of ACE on LAC. Just one positive interaction was identified: ACE populations were increased in the di-association with LAC (Fig. 2.1A), as reported previously (Newell and Douglas, 2014). Although ACE tended to be the most abundant taxon in co-associations, YST (which is 17-26 times the estimated biomass $cell^{-1}$ of ACE and LAC) attained a comparable or greater biomass than the bacteria in co-associations (Fig. 2.1C). In addition, the

inclusion of total estimated microbial biomass increased model explanatory power from 40% in total abundance model to 63% (Table S2.1A). This result reinforces the finding of Keebaugh et al. (2018) that microbial biomass indices improve the statistical power of analyses of microbial abundance in *Drosophila*, even though different conversion factors were adopted (using empirical data for *Escherichia coli* and *S. cerevisiae* for Keebaugh et al., and use of an allometric equation based on biovolume estimates in the literature for this study).

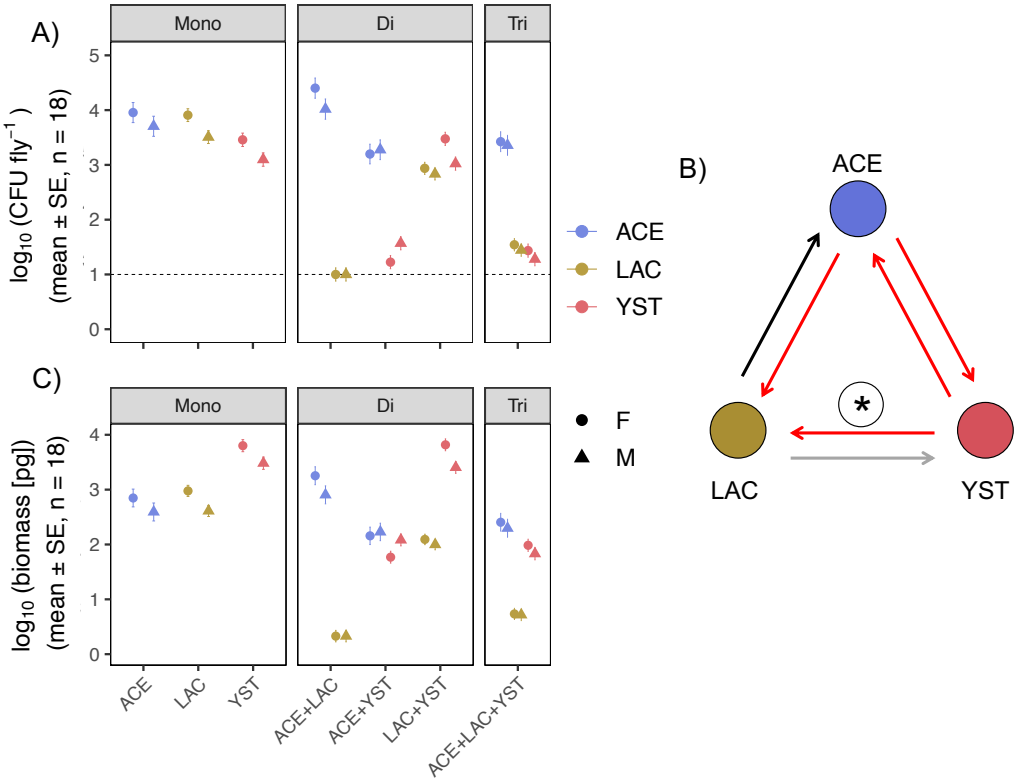


Figure 2.1. Abundance of microorganisms in flies. (A) Number of colony forming units (CFUs) per fly, (B) pairwise effects of co-association on microbial abundance, and (C) estimated biomass of microorganisms in female (F) and male (M) flies. The black line indicates the limit of detection per fly in panel (A). The estimated marginal mean and standard error from ANOVA analyses are plotted. Results from statistical analyses are provided in Table S2.1A. Among-microbe effects are summarized in panel (B); arrows indicate co-association effects (black = positive, red = negative, and grey = null) and the yeast amelioration of the negative effect of ACE on LAC in the tri-association is shown by encircled-*. Mono = mono-association. Di = di-association. Tri = tri-association.

Survival and development time of Drosophila

Our first analysis of *Drosophila* administered the 7 microbial treatments (every combination of ACE, LAC, and YST), together with axenic insects as a control, focused on insect performance. YST, irrespective of co-associating bacteria, significantly decreased development time to both pupation and eclosion compared to bacteria-only and axenic treatments (Fig. 2.2A & B and Table 2.2). ACE significantly decreased development time relative to the axenic treatment. LAC-alone did not change development time relative to the axenic treatment, nor influence the effect of ACE in the di-association (Fig. 2.2A & B and Table 2.2). These results are broadly consistent with published evidence that axenic cultivation extends larval development time of *Drosophila* (e.g. Murgier et al. 2019; Newell & Douglas, 2014; Shin et al. 2011), although this developmental delay can be much reduced or undetectable on high nutrient diets, e.g. Storelli et al. 2011; Tefit & Leulier, 2017; Wong et al. 2014.

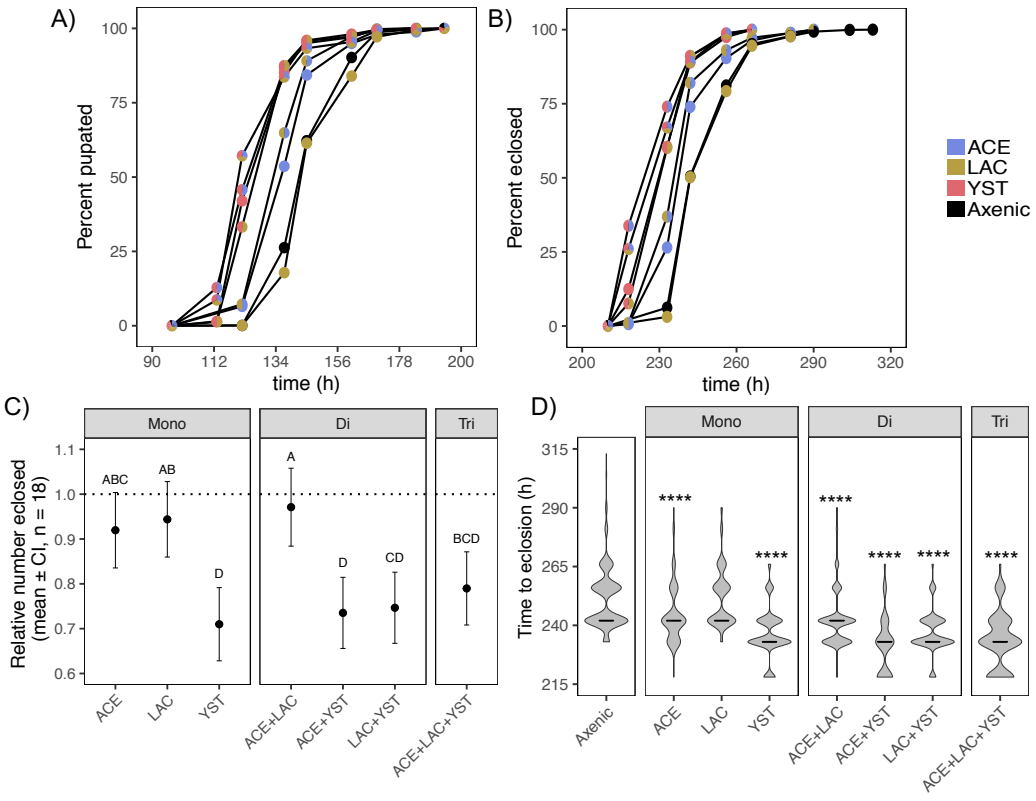

A)
Percent pupated
time (h)
B)
Percent eclosed
time (h)
ACE
LAC
YST
Axenic
C)
Relative number eclosed
(mean ± CI, n = 18)
Mono
Di
Tri
ABC
AB
D
A
D
CD
BCD
ACE
LAC
YST
ACE+LAC
ACE+YST
LAC+YST
ACE+LAC+YST
D)
Time to eclosion (h)
****
Axenic

Figure 2.2. Development time of insects colonized with different microorganisms. (A) time to pupation and (B) eclosion, (C) number of insects eclosed relative to axenic treatment, and (D) violin plots for time to eclosion. Kaplan-Meier results are plotted for panels (A) and (B), showing mean pupation and eclosion rates at the observed times. In panel C, the estimated marginal mean and 95% confidence intervals from ANOVA analysis are plotted, along with letter rankings from *post hoc* Tukey tests. The dashed line is the average value for axenic flies; confidence intervals that do not overlap with this line indicate a significant difference from axenic insects. In panel (D), probability density function and median time (black bar) for the time to eclosion are shown. Kolmogorov-Smirnov tests comparing development time of each microbial treatment to axenic insects, with $p < 0.0001$ indicated by asterisks (statistical analyses provided in Table S2.1C). Statistics for: (A) Cox mixed-effect model treatment effect: integrated $\chi^2_9 = 2491.67$, $p < 0.0001$ with full summary provided in Table S2.1B. (B) Cox mixed-effect model treatment effect: integrated $\chi^2_9 = 2187.88$, $p < 0.0001$ with full summary displayed in Table S1B. (C) ANOVA treatment effect: $F_{6,110} = 6.88$, $p = 3.20 \times 10^{-6}$, $R^2 = 0.27$.

Table 2.2. Summary statistics for average and median development time

Time to pupation				
Treatment	RMPT[a] ± SE[b]	PT50[c] (CI[d])	Tukey rank[e]	Total number
Axenic	150 ± 0.344	145 (145,145)	C	1165
ACE	143 ± 0.388	137 (137,137)	B	951
LAC	152 ± 0.394	145 (145,145)	C	992
YST	133 ± 0.393	137 (137,137)	A	819
ACE+LAC	141 ± 0.321	137 (137,137)	B	951
ACE+YST	131 ± 0.425	137 (137,137)	A	912
LAC+YST	134 ± 0.361	137 (137,137)	A	909
ACE+LAC+YST	131 ± 0.453	122 (122,122)	A	943
Time to eclosion				
Treatment	RMET[f] ± SE[b]	ET50[g] (CI[d])	Tukey rank[e]	Total number
Axenic	251 ± 0.330	242 (242,256)	C	1195
ACE	245 ± 0.371	242 (242, 242)	B	959
LAC	251 ± 0.364	242 (242,256)	C	974
YST	236 ± 0.373	233 (233, 233)	A	802
ACE+LAC	242 ± 0.350	242 (242,242)	B	951
ACE+YST	232 ± 0.409	233 (233, 233)	A	881
LAC+YST	237 ± 0.304	233 (233, 233)	A	894
ACE+LAC+YST	234 ± 0.395	233 (233, 233)	A	897

[a]restricted mean pupation time
[b]standard error
[c]median pupation time
[d]95% confidence interval
[e]letter ranking generated from Cox mixed-effect *post hoc* Tukey test
[f]restricted mean eclosion time
[g]median eclosion time

The total number of pupae and flies was reduced by 15-29% in YST-bearing *Drosophila* compared to axenic insects (Fig. S2.3A & 2.2C, respectively). This result suggests that the presence of YST reduces the larval population prior to late third instar when larvae wander to pupate. None of the bacterial associations (mono- or di-association) significantly affected the number of insects pupated or eclosed relative to

the axenic treatment and the bacteria also did not alter the effect of YST in co-associations (Fig. 2.2C & S2.3A-B and Table S2.1C).

The finding that YST reduced both development time and numbers of surviving insects raised the possibility that the treatment may have disproportionately increased mortality of slowly developing insects, thereby artifactually inflating the development rate. This hypothesis was supported by finding that the presence of YST decreased skewness for time to eclosion compared to the axenic or bacterial treatments (Fig. 2.2D). To address this issue, we conducted 5,000 simulations in which time to eclosion for insects bearing YST was supplemented with values randomly drawn from the dataset for axenic insects. The YST median development time was shifted to the axenic time in these simulations (Fig. S2.3C), indicating that the elevated developmental rate of insects with YST could be explained by disproportionate mortality of slowly developing individuals.

Nutritional indices

The weight and four of the five nutritional indices (glucose, glycogen, trehalose, triglyceride (TAG)) of the flies at 5 dpe varied significantly with the interaction term between microbial treatment and sex; the interaction term for protein content was not significant, although the main effects sex and microbial treatment were significant (Fig. 2.3 and Table S2.1D).

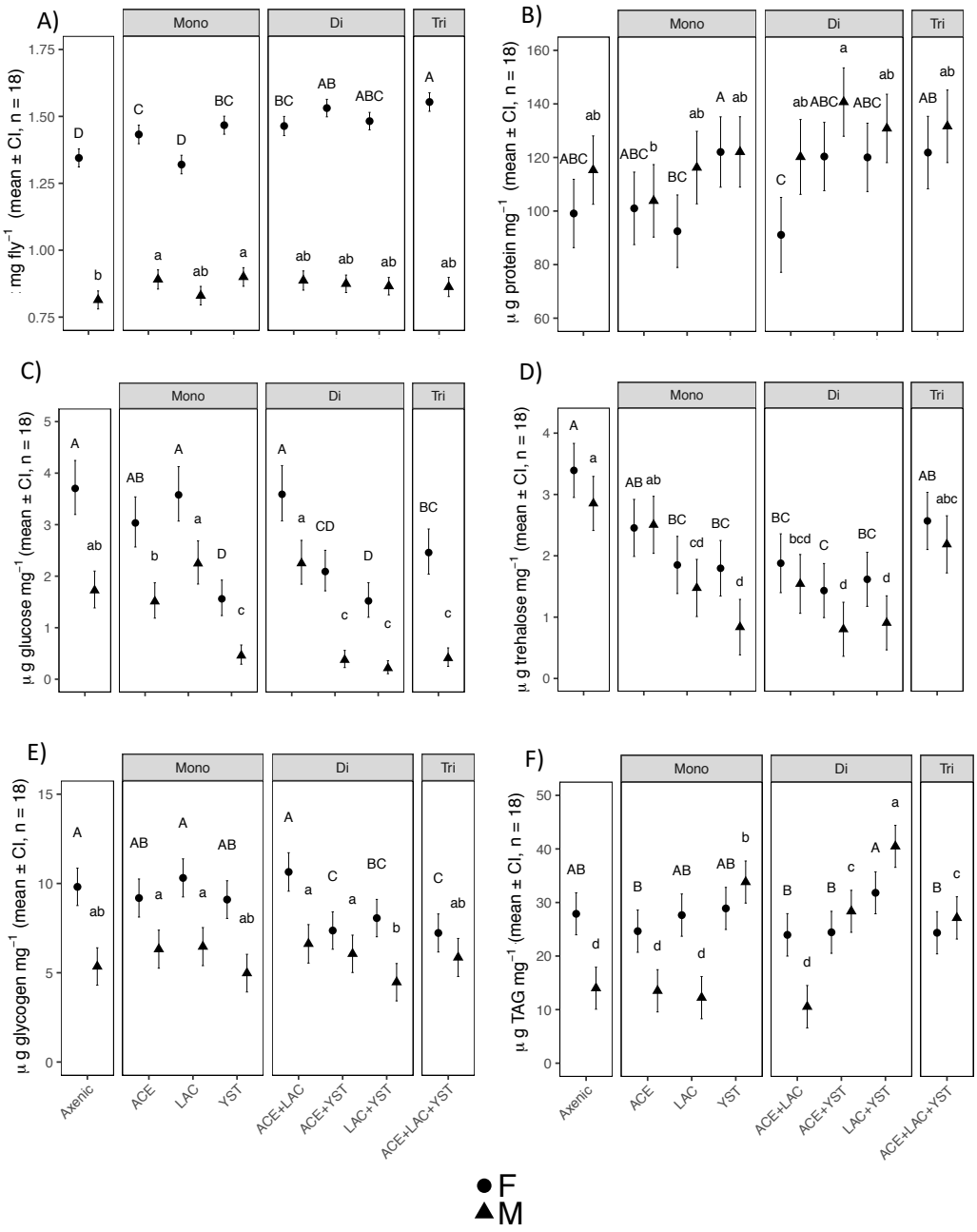

A)
B)
C)
D)
E)
F)
Mono
Di
Tri
mg fly⁻¹ (mean ± CI, n = 18)
µg protein mg⁻¹ (mean ± CI, n = 18)
µg glucose mg⁻¹ (mean ± CI, n = 18)
µg trehalose mg⁻¹ (mean ± CI, n = 18)
µg glycogen mg⁻¹ (mean ± CI, n = 18)
µg TAG mg⁻¹ (mean ± CI, n = 18)
Axenic
ACE
LAC
YST
ACE+LAC
ACE+YST
LAC+YST
ACE+LAC+YST
●F
▲M

Figure 2.3. Individual nutritional indices in flies colonized with different microorganisms. (A) Average fresh weight. (B) Protein content. (C) Glucose content. (D) Trehalose content. (E) Glycogen content. (F) Triglyceride (TAG) content. The estimated marginal mean and 95% confidence interval are plotted from each ANOVA model. Letters indicate *post hoc* Tukey test results with female and male specific comparisons indicated by capital and lowercase letters, respectively. Full description of statistical tests are in Table S2.1D.

Female flies bearing microbes weighed more on average than axenic flies (except LAC-associated flies) and bacterial co-associations with YST significantly increased female fly weight, while male flies showed little variation in weight across all treatments (Fig. 2.3A). Further inspection of the data revealed that males had a significantly lower TAG content than females in both axenic flies and flies bearing bacteria, as has also been observed for *Drosophila* in routine culture (Jehrke et al., 2018; Wong et al., 2014), while the stereotypical difference in TAG content between the sexes was ameliorated in the presence of YST (Fig. 2.3F).

To investigate the overall nutritional status of the flies further, we applied principal component analysis (PCA) (Fig. 2.4A & 2.4B). For both sexes, ordination plots significantly correlated with each other as determined by Procrustes analysis (m^2 = 0.681, r = 0.565, p = 0.001). YST was a strong separator on the first PC axis (PC1) for both males and females, but other effects differed between the sexes. YST-containing treatments were associated with investment in protein content for females, but in TAG content for males, while bacteria-only and axenic treatments increased carbohydrate content (glucose, trehalose and glycogen) for both sexes.

Multivariate correlation of PC-1 and PC-2 with microbial abundance detected significant effects of all three microbial taxa on the nutritional status of the insects, apart from LAC in males; and that the effect size of YST was substantially greater than the bacteria (Table 2.3). The PERMANOVA full factorial analysis (Table 2.3) further showed that the effect of YST was significantly influenced by co-association with either ACE or LAC in both sexes. A significant interaction between the two bacteria, ACE and LAC, was also evident for females.

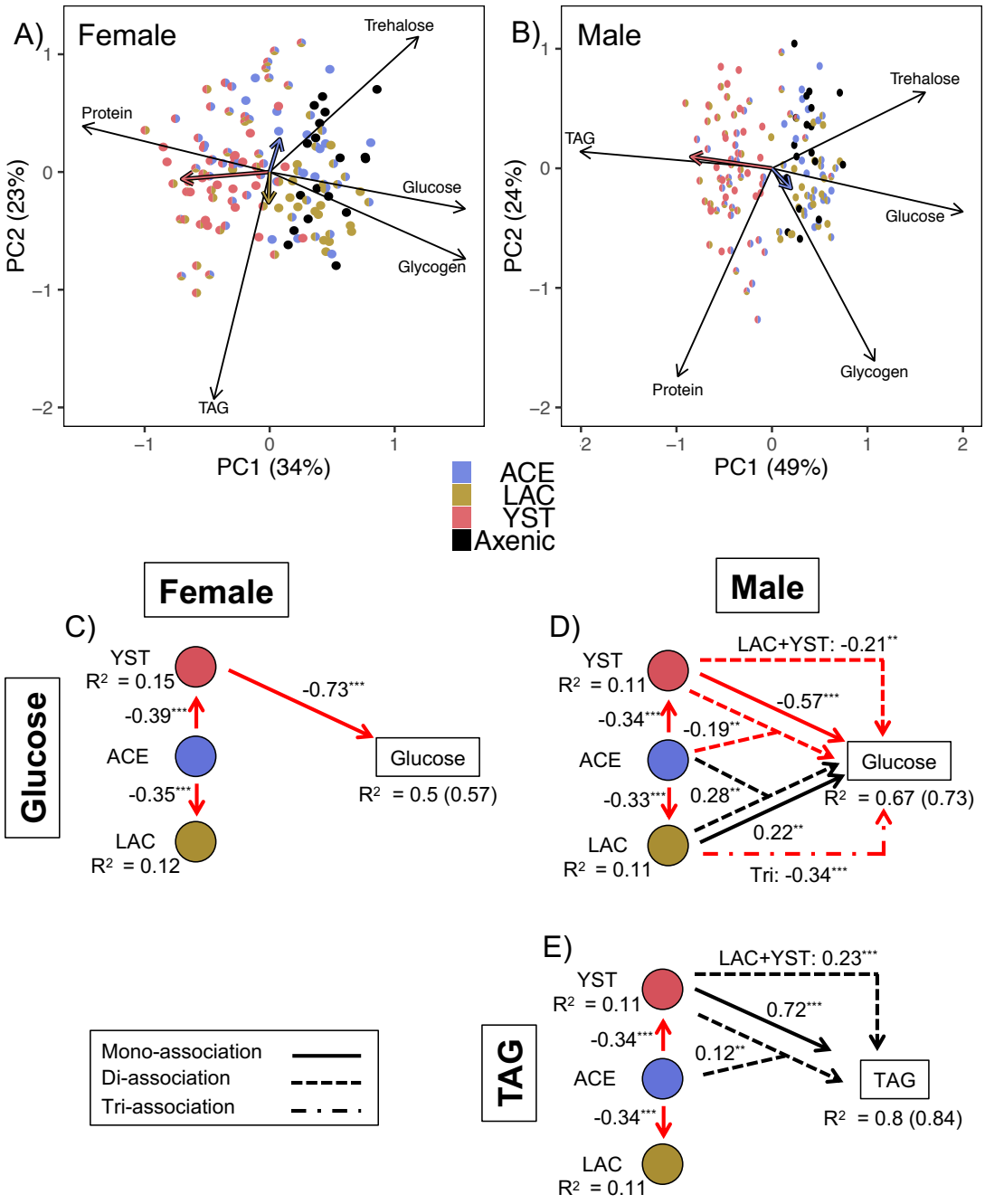

A)
Female
Trehalose
Protein
Glucose
Glycogen
TAG
PC2 (23%)
PC1 (34%)
B)
Male
Trehalose
TAG
Glucose
Protein
Glycogen
PC2 (24%)
PC1 (49%)
ACE
LAC
YST
Axenic
Female
C)
Glucose
YST
R² = 0.15
-0.73***
-0.39***
ACE
-0.35***
LAC
R² = 0.12
Glucose
R² = 0.5 (0.57)
Male
D)
LAC+YST: -0.21**
YST
R² = 0.11
-0.57***
-0.34***
-0.19**
ACE
Glucose
-0.33***
0.28**
R² = 0.67 (0.73)
0.22**
LAC
R² = 0.11
Tri: -0.34***
E)
LAC+YST: 0.23***
YST
R² = 0.11
0.72***
-0.34***
0.12**
ACE
TAG
-0.34***
R² = 0.8 (0.84)
LAC
R² = 0.11
Mono-association
Di-association
Tri-association

Figure 2.4. Nutritional status of flies colonized with different microorganisms. Principal component analyses (PCA) for (A) females and (B) males. Black arrows indicate loading indices and colored vectors correspond to correlation of microbial abundances (log_{10}-transformed CFU fly^{-1} + 1) with PCA axes. Significant results from structural equation models are shown for (C) female glucose and for (D) male glucose and (E) TAG content. Red and black arrows indicate negative and positive associations, respectively. The standardized coefficient for each significant association and marginal R^2 values for all response variables are shown, with conditional R^2 values shown in parentheses if needed. Full structural equation model test results are provided in Table S2.1E and Fig S2.4A.

Table 2.3. Effect of microbial presence and abundance on fly nutritional status.

PERMANOVA	Female		Male	
	Effect test	Effect size (R^2)	Effect test	Effect size (R^2)
ACE	**$F_{1,126}$ = 8.672, p = 0.001**	**0.04453**	**$F_{1,126}$ = 5.432, p = 0.002**	**0.02260**
LAC	$F_{1,126}$ = 1.404, p = 0.216	0.00721	$F_{1,126}$ = 0.813, p = 0.505	0.00338
YST	**$F_{1,126}$ = 44.474, p = 0.001**	**0.22835**	**$F_{1,126}$ = 90.885, p = 0.001**	**0.37811**
ACE * LAC	**$F_{1,126}$ = 3.670, p = 0.006**	**0.01885**	$F_{1,126}$ = 2.608, p = 0.055	0.01085
ACE * YST	**$F_{1,126}$ = 4.320, p = 0.002**	**0.02218**	**$F_{1,126}$ = 3.079, p = 0.029**	**0.01281**
LAC * YST	**$F_{1,126}$ = 5.989, p = 0.002**	**0.03075**	**$F_{1,126}$ = 9.782, p = 0.001**	**0.04069**
ACE * LAC * YST	$F_{1,126}$ = 0.234, p = 0.941	0.00120	$F_{1,126}$ = 1.770, p = 0.148	0.00736
Multivariate correlation	Effect size (R^2)	P-value	Effect size (R^2)	P-value
Log_{10} (ACE+1)	**0.084**	**0.008**	**0.068**	**0.005**
Log_{10} (LAC+1)	**0.068**	**0.006**	0.008	0.602
Log_{10} (YST+1)	**0.496**	**0.001**	**0.727**	**0.001**

P-values below the threshold are bolded.

We then implemented structural equation modeling to investigate the contributions of the different microbial taxa and their abundance on each nutritional index (Fig. S2.4A-B). The most pronounced effects of microbial composition were obtained for fly glucose and TAG content (marginal $R^2 > 0.5$ in at least one sex), while the explanatory power for protein content was particularly weak (marginal R^2 0.08-0.2) (Fig. S2.4B). The model outputs for glucose and TAG content are shown in Fig. 2.4. For female flies, YST negatively impacted glucose content, while ACE indirectly promoted glucose content by lowering the YST population (standardized coefficient =

0.28) (Fig. 2.4C). The relationships were more complex in males. YST-only reduced the glucose content but increased the TAG content, and both of these effects were dampened by co-association with ACE or LAC. In addition, LAC promoted male glucose content, both alone and in co-association with ACE (Fig. 2.4D & E).

Nutritional indices of flies administered microbial fermentation metabolites

We hypothesized that the effects of microorganisms on the nutritional indices of the flies were mediated, at least partly, by by-products of microbial metabolism that alter nutrient allocation in the fly (see Introduction). We tested the hypothesis by feeding adult flies on chemically-defined liquid diet, administered via capillary tubes and supplemented individually with acetic acid, acetoin, ethanol or lactic acid. The flies feeding from the capillary tubes increased in daily consumption per fly over the full four days of the experiment for female flies, and over the first two days followed by stable feeding rates between days-3 and -4 for male flies (Fig. S2.5A & B). On average 85% and 82% of female and male flies, respectively, survived during the 4-day study period with significant differences by concentration for all metabolites, except acetoin, and a sex-specific difference in survival for lactic acid (Fig. S2.5C and Table S2.1F). Generally, the indices scored did not vary significantly between the controls (i.e. flies on fermentation metabolite-free diets) for the three independent experiments per fermentation metabolite or between the experiments on different fermentation metabolites. (Exceptionally, among-experiment variation was obtained for food consumption in the acetoin-control and glucose content in the ethanol-control (Fig. S2.5C); and the mean glucose content of male flies in the acetoin control was

significantly greater than the lactic acid control (Table S2.1F).)

The effect of metabolite concentration (0, 0.15, and 0.3 M) and fly sex were evaluated for TAG, glucose and protein content, with fly weight and food consumption included in each model as covariates to discriminate between the effects of each metabolite and associated covariates on nutritional status. Acetic acid significantly reduced both TAG and protein content of the flies, with significant effects for both concentrations tested (0.15 and 0.3 M) for females and for one concentration for males (Fig. 2.5 and Table S2.1F). This result could not be attributed to effects of acetic acid on food consumption (the volume of diet ingested, used as a covariate in the analysis was not significant, Table S2.1F) but TAG content was positively associated with fly weight in females for both experiments (Fig. 2.5 and Table S2.1F). The other three metabolites had no significant effects, apart from a change in glucose content of females administered 0.3 M acetoin and lactic acid and a small reduction in female protein content at 0.3 M lactic acid (Table S2.1F).

To investigate the link between acetic acid and microbial effects on fly TAG and protein content, we quantified the acetic acid content of the flies colonized with different microbial taxa. This analysis was extended to other SCFAs, as a check for the possible contribution of other SCFAs to this effect.

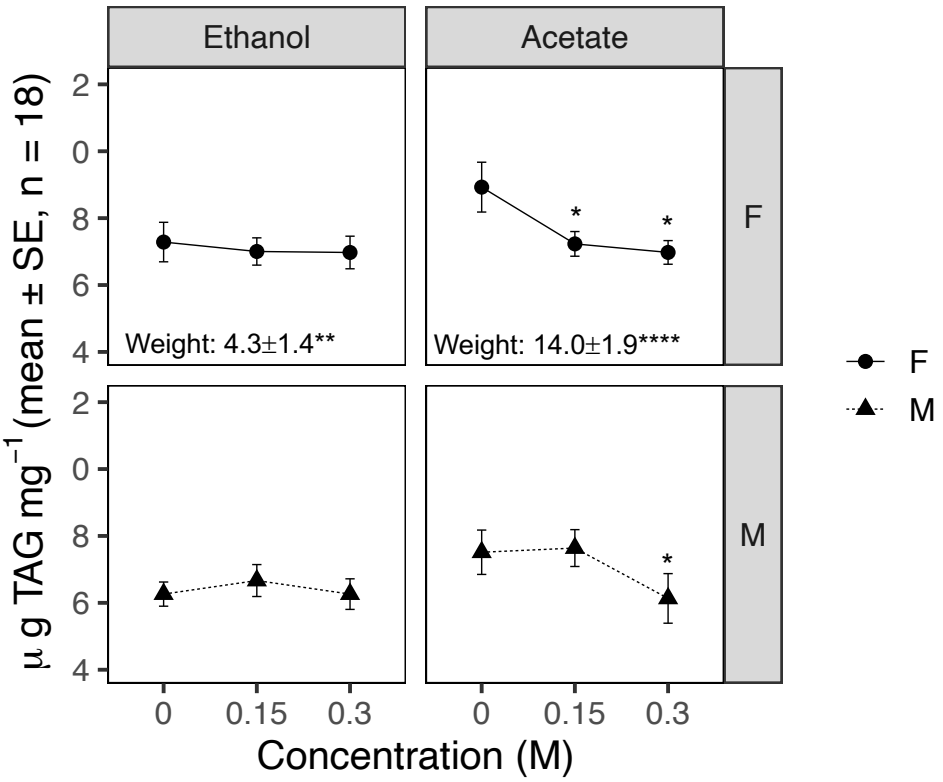


Figure 2.5. TAG content of axenic flies administered microbial fermentation compounds. The effect of ethanol and acetic acid on TAG content is indicated with raw mean and standard error. * indicates p-value < 0.05 for *post hoc* Dunnett's test from each ANOVA model. The covariate slopes and SEs for weight per fly from each ANOVA model are included when significant and displayed under each metabolite label; the volume of diet consumed by the flies was non-significant for all experiments. Statistical analyses are provided in Table S2.1F and ANOVA model plots are included in extended Fig. S5.2.

SCFA content and association with host nutrient allocation

Of the 12 SCFAs tested (see methods), just three compounds were detected in the adult flies: acetic acid, butyric acid and propionic acid, which are also the three dominant SCFAs found in the mammalian intestine (Cummings et al., 2004). The normalized content of all three SCFAs in *Drosophila* (Fig. 2.6A) was significantly increased in flies bearing YST with differences by sex in butyric acid (Fig. 2.6A and Table S2.1G). For acetic acid, this effect was increased by the YST/ACE interaction and also varied with both LAC and sex (Table S2.1G). ACE also significantly affected propionic and butyric acid content, although the magnitude of this effect varied with sex and LAC (Fig. 2.6A and Table S2.1G).

A moderation analysis was performed to investigate how the microorganisms influenced the association between SCFA content and each nutritional index. A significant relationship was obtained between one SCFA, acetic acid, and one nutritional index, TAG, in male flies only (Table S2.1H). Specifically, the association between TAG and acetic acid content varied significantly with both the presence of ACE and YST (Table 2.4), with a correlated increase in acetic acid titer and decrease in TAG content of flies bearing ACE and YST (Fig. 2.6B). This finding suggests that a YST factor may modulate the effect of acetic acid on TAG content observed in Fig. 2.5, while an ACE factor is capable of reducing the YST factor to decrease TAG content as acetic acid titer increases (Fig. 2.6C).

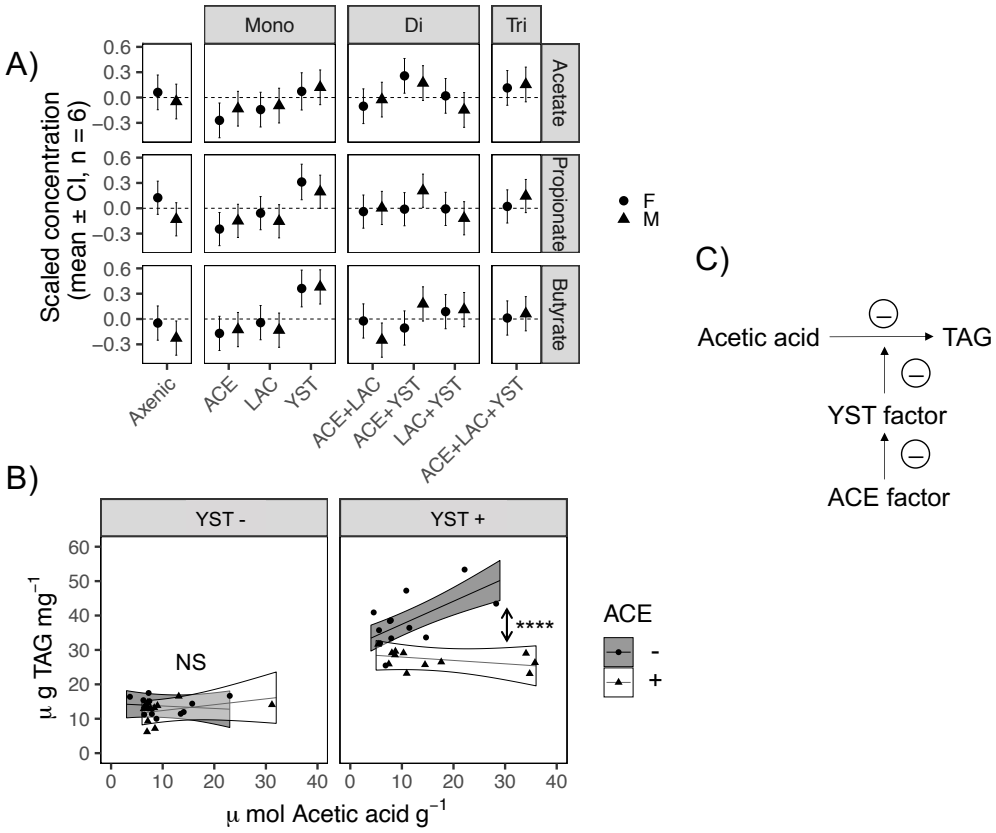

A)
Mono
Di
Tri
Acetate
Propionate
Butyrate
Scaled concentration
(mean ± CI, n = 6)
0.6
0.3
0.0
−0.3
Axenic
ACE
LAC
YST
ACE+LAC
ACE+YST
LAC+YST
ACE+LAC+YST
F
M
B)
YST -
YST +
μ g TAG mg−1
NS
****
ACE
-
+
μ mol Acetic acid g−1
C)
Acetic acid
TAG
YST factor
ACE factor

Figure 2.6. The SCFA content of *Drosophila* colonized with different microorganisms and the association between the SCFA acetic acid and TAG content. (A) SCFA profile of flies, (B) moderation analysis for microbial effect on TAG content by acetic acid titer, and (C) proposed ACE and YST factors influencing the relationship between acetic acid and TAG (see Discussion for details). For panel (A), the dotted horizontal line indicates the mean concentration across all treatments for each SCFA. The estimated marginal mean and 95% confidence interval from ANOVA analyses are plotted. Confidence intervals that do not overlap with zero indicate a significant effect of a microbial treatment on SCFA concentration. Statistical analyses are provided in Table S2.1G. Panel (B) shows the linear prediction for male fly TAG content by acetic acid titer for the presence (+) and absence (-) of ACE and YST in all combinations from the moderation analysis with the results averaged over LAC treatments. The ribbon around the line represents the confidence interval with symbols displaying individual replicates sampled. A *post hoc* linear contrast was performed comparing the effect of the presence of ACE on the regression slopes with and without YST. Asterisks indicates statistical significance ($p < 0.0001$); NS = not significant.

Table 2.4. Moderation effect of acetic acid titer and microorganism presence on TAG content in male flies.

Predictors	Male	
	ANOVA effect test	***Regression coefficient[a]***
ACE	$F_{1,17.12} = 20.49$***	-3.7 ± 3.0
YST	$F_{1,17.45} = 180.63$***	16.3 ± 3.0***
[Acetic acid]	$F_{1,31.53} = 5.19$*	-0.1 ± 0.2
ACE*[Acetic acid]	$F_{1,32.67} = 7.27$*	0.2 ± 0.2
YST*[Acetic acid]	$F_{1,31.65} = 1.59$	0.7 ± 0.2**
ACE*YST	$F_{1,17.48} = 10.39$**	1.9 ± 4.3
ACE*YST*[Acetic acid]	$F_{1,29.94} = 12.53$**	-1.0 ± 0.3***
Analysis of Deviance		
Experimental replicate (treatment)	$X^2_2 = 6.52$*	
R^2		
Marginal	0.877	
Conditional	0.945	

Effect test tables for all moderation analyses are in Table S2.1H. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[a]The coefficient estimate (β) and standard error are reported to show the effect of factors on TAG content. The estimate is based on the presence of each microbe.

Discussion

This study on the interaction between gut microorganisms and their *Drosophila* host yielded two key results. First, the impact of individual bacterial and yeast taxa on the performance and nutritional status of *Drosophila* is strongly influenced by both the presence of other microorganisms and host sex. Second, the relationship between a key microbial fermentation product, acetic acid, and fly lipid content is strongly dependent on both microbiota composition and sex. Here we address the processes that may contribute to the interactive effects of the bacteria and yeasts, focusing the latter part of the Discussion specifically on the role of acetic acid. Our interpretation of the results is informed by two important aspects of *Drosophila*-microbiome system. The first is that viable microbial cells are shed via the feces onto the diet (Fink et al.,

2013; Inamine, et al., 2018), and these shed cells are both available for re-ingestion (i.e. fecal-oral cycling) and can proliferate in the food, thereby altering the nutritional composition of the food consumed by the insects (Broderick and Lemaitre, 2012; Huang and Douglas, 2015; Martino et al., 2018; Wong et al., 2015). In this study, we focused on microbial populations in the insects, while recognizing that some microbial effects on the nutritional status of *Drosophila* may be mediated by diet-associated microorganisms. Secondly, various studies indicate that the effect of gut microorganisms on the nutritional status of *Drosophila* can be mediated by post-ingestive processes, including nutrient assimilation across the gut wall and nutrient allocation by the insect (e.g. Shin et al., 2011; Storelli et al., 2011; Kamareddine et al., 2018; Yamada et al., 2015), and this is the primary focus of this study. Nevertheless, we recognize that the nutritional physiology of *Drosophila* can be influenced by microbial effects on feeding rates (e.g. Huang and Douglas, 2015; Wong et al., 2014; Wong et al., 2017), and that this important facet of the *Drosophila*-microbiome interactions remains to be investigated systematically.

Many of the among-microbe interactions and sex-specific effects of the microorganisms involved the YST treatment (an isolate of *Hanseniaspora uvarum* from wild *Drosophila*). Previous studies on gut bacteria in *Drosophila* and other animals have shown that most among-bacteria interactions reduce bacterial abundance, i.e. they are negative interactions (Coyte and Rakoff-Nahoum, 2019; Faust et al., 2018; Gould et al., 2018; Newell and Douglas, 2014; Venturelli et al., 2018). This study extends this generality to a yeast, including evidence for both mutual reduction in numbers of live YST and ACE in flies colonized with both taxa and a

negative effect of YST on LAC abundance (Fig. 2.1). The basis for these effects in the *Drosophila* system remains to be determined but may include competition for nutrients in the food and *Drosophila* gut, as well as toxicity of certain metabolic by-products released from the different microorganisms (Dashko et al., 2014; Hibbing et al., 2010; Mitri and Foster, 2013). In some instances, however, the cross-feeding of metabolic by-products between microbial taxa (Fischer et al., 2017) can promote microbial abundance in *Drosophila*. For example, the LAC-dependent promotion of ACE abundance (Consuegra et al., 2020; Newell and Douglas, 2014; this study) is likely mediated by LAC-derived fermentation metabolites (e.g. lactic acid) (Consuegra et al., 2020; Sommer and Newell, 2019). In addition, the effect of YST on LAC abundance may be due to YST-derived small metabolites as observed in other yeast-*Lactobacillus* systems (Ponomarova et al., 2017).

The effect of YST on *Drosophila* performance was highly significant and independent of the presence of bacteria. The pattern of reduced survival but increased developmental rate to adulthood of *Drosophila* with YST obtained in our study has also been reported by Murgier et al. (2019), who compared pre-adult performance of *Drosophila* associated with a different strain of the same yeast species (*Hanseniaspora uvarum*) to a treatment comprising dried *Saccharomyces cerevisiae* routinely used to maintain *Drosophila* cultures. Our dataset is compatible with the interpretation that YST killed slowly-developing larvae (Fig. 2.2D and S2.3C). Recognizing that yeasts release a wide range of fermentation metabolites and other compounds (Arguello et al., 2013; Bueno et al., 2019; Halbfeld et al., 2014; Jolly et al., 2014; Krause et al., 2018; Rossouw et al., 2008), we hypothesize that one or more of the YST products

accumulate in the diet to levels that are toxic for the slower-developing larval insects. YST may, in this way, exert very strong selection for rapid *Drosophila* development time. A high priority for future research is to identify and investigate the mode of action of the putative insecticidal products of YST.

Our analysis of the nutritional indices of 5-day-old adult flies revealed markedly different effects of YST on nutrient allocation in male and female *Drosophila*, with significantly increased protein content of females and TAG (lipid) content of males (Fig. 2.3 & 2.4). The female response is fully in keeping with published evidence that females derive protein from dietary yeast for ovary maturation and egg production, underpinned at the molecular level by heightened expression of yolk protein genes for vitellogenesis (Bownes et al., 1988; Roy et al., 2018; Terashima and Bownes, 2004). The male response to YST comprised the loss of the stereotypical lower lipid content in male than female flies that is exhibited by *Drosophila* in a range of rearing conditions (Jehrke et al., 2018; Schwasinger-Schmidt et al., 2012; Wong et al., 2014), and also by axenic flies and *Drosophila* bearing bacteria (Fig. 2.3A). The effect of YST on TAG in male flies matches the effect of eliminating the function of the *Drosophila* lipase gene *brummer* in neurons and somatic cells of the testes (Wat et al., 2020). The evidence that *brummer* expression in these organs plays an important role in global lipid homeostasis (Wat et al., 2020) raises the possibility that the nutritional consequences of YST in male flies involves system-level change in the regulation of lipid metabolism via a *brummer*-dependent process. In addition, changes in feeding behavior and dietary nutrient composition may provide insight into the microbial-mediated effects on nutrient allocation.

Whatever the mechanistic basis of YST-mediated increase in TAG of male flies, this effect is ameliorated by co-colonization with ACE (Fig. 2.3F & 2.4E). Importantly, aspects of the interactive effect of YST and ACE on male TAG content did not match well to published evidence that various *Acetobacter* strains reduce fly TAG content (Chaston et al., 2014; Newell and Douglas, 2014; Shin et al., 2011). Contrary to these previous studies, flies colonized with ACE in this study did not have significantly lower TAG content than axenic flies. The likely basis for this discrepancy is a difference in experimental design. *Acetobacter* growing in the food consume dietary glucose, thereby reducing the availability of glucose substrate for TAG synthesis by *Drosophila* (Huang and Douglas, 2015). Dietary glucose would be less-depleted for the 5-day-old flies analyzed in this study than in previous work because we transferred the newly-eclosed flies to fresh diet, while published studies reared the *Drosophila* with the *Acetobacter* inoculum from egg without refreshing the diet at adulthood. A further discrepancy is the finding from our structural equation modeling that the effects of microorganisms on fly TAG content could not be predicted from microbial population size, contrary to the published finding of a negative correlation between fly TAG content and *Acetobacter* load (Chaston et al., 2014). This difference likely arises from the inclusion of YST in our analyses.

Insight into the basis of the interactive effect of ACE and YST on fly TAG content comes from consideration of the response of adult *Drosophila* to microbial fermentation products. Our demonstration that dietary acetic acid significantly reduces TAG content of axenic flies in both sexes is consistent with previous studies, which have, additionally, shown that acetic acid-mediated TAG reduction is mediated by

enhanced antimicrobial peptide production in the gut and systemic insulin signaling (Kamareddine et al., 2018; Shin et al., 2011).

Our study demonstrates, further, that the relationship between acetic acid titer and *Drosophila* TAG content is strongly influenced by the presence and composition of the gut microbiota. The first issue is the microbial source of acetic acid. To date, acetic acid in the *Drosophila* system has been identified as the product of aerobic fermentation by Acetobacteraceae, a metabolic trait that is displayed when the ethanol substrate is cross-fed from yeasts or heterofermentative lactobacilli (Fischer et al., 2017; Shin et al., 2011; Sommer and Newell, 2019). However, this interpretation may be incomplete because, although (as predicted) fly acetic acid levels are greatest in the YST/ACE treatments, they are also significantly elevated in YST-only flies compared to axenic flies. We hypothesize that YST may be a net producer of acetic acid, as demonstrated for various yeasts (Jolly et al., 2014), including a different strain of the same species, *Hanseniaspora uvarum* (Bueno et al., 2019). A further potential source of metabolic complexity is the production of acetic acid by various heterofermentative lactobacilli (Adler et al., 2013; Oude Elferink et al., 2001), although we obtained no indication of this effect from the acetic acid content of flies bearing the heterofermentative LAC used in this study.

Taken together, these considerations lead to the apparently paradoxical conclusion that male flies bearing YST have elevated TAG levels, despite their high acetic acid titer. This paradox can be resolved by invoking a YST-derived factor that suppresses the metabolic response of male flies to acetic acid, and the reversal of this effect by an ACE-factor that suppresses the YST-factor (Fig. 2.6C). Priorities for future research

are twofold: to establish the identity of the putative YST- and ACE-factors; and to investigate how the putative YST-factor may interact with the IMD and insulin signaling pathways (Kamareddine et al., 2018; Shin et al. 2011) that mediate TAG-reduction by acetic acid and the *brummer*-mediated regulation of TAG levels in males.

Our analysis of the SCFA profiles detected propionic acid and butyric acid in *Drosophila*, with elevated titers in YST-bearing flies. Although we obtained no significant relationship between these SCFAs and *Drosophila* nutritional indices, these microbial-derived SCFAs are important effectors of gut microbiome-host interactions in mammals (Den Besten et al., 2013; Gentile and Weir, 2018), including transgenerational effects on energy homeostasis (Kimura et al., 2020). In addition, propionic acid has been identified as an appetite stimulant for *Drosophila* larvae under nutrient stress (Depetris-Chauvin et al., 2017). Considering the evolutionary conservation of many aspects of gut microbiome interactions and metabolism across the animal kingdom (Douglas, 2019; Musselman and Kühnlein, 2018), we cannot exclude the possibility that these SCFAs may influence *Drosophila* metabolism at different developmental stages, on different diets or over longer timescales than used in this study.

We conclude by considering how this study contributes to our understanding of the central role of gut microorganisms and the microbial metabolite acetic acid as determinants of *Drosophila* lipid content. Although most research on this topic has been conducted on *Drosophila* associated with a single bacterial strain (e.g. Chaston et al., 2014; Ma et al., 2019; Shin et al., 2011), there is growing interest in the effects of among-microbe interactions on various fly traits (Aranda-Díaz et al., 2020; Consuegra

et al., 2020; Fischer et al., 2017; Gould et al., 2018; Judd et al., 2018; Sommer and Newell, 2019). Here, we demonstrate that the impact of a microbial community on metabolism-related traits, especially lipid content, cannot be predicted reliably from the study of mono-associations because the effect of individual microorganisms is strongly influenced by other microorganisms, especially yeast-bacterial interactions, and by host sex. There is growing evidence that this complexity is not peculiar to *Drosophila* but applies to other animals, including humans (Bolnick et al., 2014; Haro et al., 2016; Markle et al., 2013; Weger et al., 2019). Because *Drosophila* is superbly amenable to large experiments, including combinatorial designs of microbial treatments (Gould et al., 2018; this study), it is an excellent system to investigate the fundamental processes underlying gut microbiome-host interactions.

References

Adair, K. L., Wilson, M., Bost, A. and Douglas, A. E. (2018). Microbial community assembly in wild populations of the fruit fly *Drosophila melanogaster*. *ISME J.* 12, 959–972.

Adler, P., Bolten, C. J., Dohnt, K., Hansen, C. E. and Wittmann, C. (2013). Core fluxome and metafluxome of lactic acid bacteria under simulated cocoa pulp fermentation conditions. *Appl. Environ. Microbiol.* 79, 5670–5681.

Adler, P., Frey, L. J., Berger, A., Bolten, C. J., Hansen, C. E. and Wittmann, C. (2014). The key to acetate: metabolic fluxes of acetic acid bacteria under cocoa pulp fermentation-simulating conditions. *Appl. Environ. Microbiol.* 80, 4702-4716.

Aranda-Díaz, A., Obadia, B., Dodge, R., Thomsen, T., Hallberg, Z. F., Güvener, Z. T., Ludington, W. B. and Huang, K. C. (2020). Bacterial interspecies interactions modulate pH-mediated antibiotic tolerance. *eLife* 9, e51493.

Arguello, J. R., Sellanes, C., Lou, Y. R. and Raguso, R. A. (2013). Can yeast (*S. cerevisiae*) metabolic volatiles provide polymorphic signaling? *PLoS One* 8, e70219.

Barata, A., Campo, E., Malfeito-Ferreira, M., Loureiro, V. and Ferreira, V. (2011). Analytical and sensorial characteriation of the aroma of wines produced with sour rotten grapes using GC-O and GC-MS: identifiation of key aroma compounds. *J. Agric. Food Chem.* 59, 2543-2553.

Barata, A., Santos, S. C., Malfeito-Ferreira, M. and Loureiro, V. (2012). New insights into the ecological interaction between grape berry microorganisms and *Drosophila* flies during development of sour rot. *Microb. Ecol.* 64, 416-430.

Bartoń, K. (2019). MuMIn: Multi-Model Inference, R package version 1.43.6.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* 67, 1–48.

Bolnick, D. I., Snowberg, L. K., Hirsch, P. E., Lauber, C. L., Org, E., Parks, B., Lusis, A. J., Knight, R., Caporaso, J. G. and Svanbäck, R. (2014). Individual diet has sex-dependent effects on vertebrate gut microbiota. *Nat. Commun.* 5, 4500.

Bownes, M., Scott, A. and Shirras, A. (1988). Dietary components modulate yolk protein gene transcription in *Drosophila melanogaster*. *Development* 103, 119–128.

Broderick, N. A. and Lemaitre, B. (2012). Gut-associated microbes of *Drosophila melanogaster*. *Gut Microbes* 3, 307–321.

Bueno, E., Martin, K. R., Raguso, R. A., McMullen II, J. G., Hesler, S. P., Loeb, G. M. and Douglas, A. E. (2019). Response of wild spotted wing *Drosophila* (*Drosophila suzukii*) to microbial volatiles. *J. Chem. Ecol.* Epub 26 December.

Cadez, N. and Smith, M. T. (2011). *Hanseniaspora* Zikes (1912). In *The Yeasts* (ed. Kurtzman, C. P., Fell, J. W. and Boekhout, T.), pp. 421–434. Amsterdam, Netherlands: Elsevier B.V.

Cai, J., Zhang, J., Tian, Y., Zhang, L., Hatzakis, E., Krausz, K. W., Smith, P. B., Gonzalez, F. J. and Patterson, A. D. (2017). Orthogonal comparison of GC − MS and 1 H NMR spectroscopy for short chain-fatty acid quantitation. *Anal. Chem.*

89, 7900-7906.
Chandler, J. A., Lang, J. M., Bhatnagar, S. and Eisen, J. A. (2011). Bacterial communities of diverse *Drosophila* species : Ecological context of a host-microbe model system. *PLoS Genet.* 7, e1002272.
Chandler, J. A., Eisen, J. A. and Kopp, A. (2012). Yeast communities of diverse *Drosophila* species : Comparison of two symbiont groups in the same hosts. *Appl. Environ. Microbiol.* 78, 7327–7336.
Chaston, J. M., Newell, P. D. and Douglas, A. E. (2014). Metagenome-wide association of microbial determinants of host phenotype in *Drosophila melanogaster*. *MBio* 5, 1–12.
Christian, A. E., Haynes, M. P., Phillips, M. C. and Rothblat, G. H. (1997). Use of cyclodextrins for manipulating cellular cholesterol content. *J. Lipid Res.* 38, 2264–2272.
Cleenwerck, I., Gonzalez, Á., Camu, N., Engelbeen, K., De Vos, P. and De Vuyst, L. (2008). *Acetobacter fabarum* sp. nov., an acetic acid bacterium from a Ghanaian cocoa bean heap fermentation. *Int. J. Syst. Evol. Microbiol.* 58, 2180–2185.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Second Edi. Hillsdale, NJ: Lawrence Earlbaum Associates.
Cold Spring Harbor (2018). PBS (pH 7.4). *Cold Spring Harb. Protoc.*
Consuegra, J., Grenier, T., Akherraz, H., Rahioui, I., Gervais, H., da Silva, P. and Leulier, F. (2020). Metabolic cooperation among commensal bacteria supports *Drosophila* juvenile growth under nutrient stress. *ISCIENCE In press.*
Coyte, K. Z. and Rakoff-Nahoum, S. (2019). Understanding competition and cooperation within the mammalian gut microbiome. *Curr. Biol.* 29, R538–R544.
Coyte, K. Z., Schluter, J. and Foster, K. R. (2015). The ecology of the microbiome: Networks, competition, and stability. *Science*. 350, 663–666.
Cummings, J.H., Rombeau, J.L. and Sakata, T. (2004). *Physiological and clinical aspects of short-chain fatty acids*. First edi. Cambridge, UK: Cambridge University Press.
Dashko, S., Zhou, N., Compagno, C. and Piškur, J. (2014). Why, when, and how did yeast evolve alcoholic fermentation? *FEMS Yeast Res.* 14, 826–832.
De Camargo, R. and Phaff, H. J. (1957). Yeasts occurring in *Drosophila* flies and in fermenting tomato fruits in Northern California. *J. Food Sci.* 22, 367–372.
Den Besten, G., Van Eunen, K., Groen, A. K., Venema, K., Reijngoud, D. J. and Bakker, B. M. (2013). The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J. Lipid Res.* 54, 2325–2340.
Depetris-Chauvin, A., Galagovsky, D., Chevalier, C., Maniere, G. and Grosjean, Y. (2017). Olfactory detection of a bacterial short-chain fatty acid acts as an orexigenic signal in *Drosophila melanogaster* larvae. *Sci. Rep.* 7, 1–14.
Devineni, A. V. and Heberlein, U. (2013). The evolution of *Drosophila melanogaster* as a model for alcohol research . *Annu. Rev. Neurosci.* 36, 121–138.
Dobson, A. J., Chaston, J. M., Newell, P. D., Donahue, L., Hermann, S. L., Sannino, D. R., Westmiller, S., Wong, A. C. N., Clark, A. G., Lazzaro, B. P., et al. (2015). Host genetic determinants of microbiota-dependent nutrition revealed by genome-wide analysis of *Drosophila melanogaster*. *Nat. Commun.* 6, 6312.

Douglas, A. E. (2019). Simple animal models for microbiome research. *Nat. Rev. Microbiol.* 17, 764–775.
Douglas, A. E. (2020). The microbial exometabolome: ecological resource and architect of microbial communities. *Philos. Trans. R. Soc. B Biol. Sci.* 375, 20190250.
Erkosar, B., Storelli, G., Defaye, A. and Leulier, F. (2013). Host-intestinal microbiota mutualism: "learning on the fly." *Cell Host Microbe* 13, 8-14.
Farine, J., Habbachi, W., Cortot, J., Roche, S. and Ferveur, J. (2017). Maternally-transmitted microbiota affects odor emission and preference in *Drosophila* larva. *Sci. Rep.* 7, 1–10.
Faust, K., Bauchinger, F., Laroche, B., de Buyl, S., Lahti, L., Washburne, A. D., Gonze, D. and Widder, S. (2018). Signatures of ecological processes in microbial community time series. *Microbiome* 6, 1–13.
Fink, C., Staubach, F., Kuenzel, S., Baines, J.F. and Roeder, T. (2013). Noninvasive analysis of microbiome dynamics in the fruit fly *Drosophila melanogaster*. *Appl. Environ. Microbiol.* 79, 6984-6988.
Fischer, C., Trautman, E. P., Crawford, J. M., Stabb, E. V., Handelsman, J. and Broderick, N. A. (2017). Metabolite exchange between microbiome members produces compounds that influence *Drosophila* behavior. *Elife* 6, e18855.
Fox, J. and Weisberg, S. (2019). *An {R} Companion to Applied Regression*. Third edit. Thousand Oaks, CA.
Freilich, S., Zarecki, R., Eilam, O., Segal, E. S., Henry, C. S., Kupiec, M., Gophna, U., Sharan, R. and Ruppin, E. (2011). Competitive and cooperative metabolic interactions in bacterial communities. *Nat. Commun.* 2, 589.
Fry, J. D. (2014). Mechanisms of naturally evolved ethanol resistance in *Drosophila melanogaster*. *J. Exp. Biol.* 217, 3996–4003.
Gentile, C. L. and Weir, T. L. (2018). The gut microbiota at the intersection of diet and human health. *Science.* 362, 776–780.
Gould, A. L., Zhang, V., Lamberti, L., Jones, E. W., Obadia, B., Korasidis, N., Gavryushkin, A., Carlson, J. M., Beerenwinkel, N. and Ludington, W. B. (2018). Microbiome interactions shape host fitness. *Proc. Natl. Acad. Sci. U. S. A.* 115, E11951–E11960.
Graber, J. R. and Breznak, J. A. (2005). Folate cross-feeding supports symbiotic homoacetogenic spirochetes. *Appl. Environ. Microbiol.* 71, 1883–1889.
Gries, O. and Ly, T. (2009). Volume Three, The Firmicutes. In *Bergey's Manula of Systematic Bacteriology* (ed. Vos, P. De, Garrity, G. M., Jones, D., Krieg, N. R., Ludwig, W., Rainey, F. A., Schleifer, K.-H. and Whitman, W. B.), pp. 203–208. Heidelberg, Germany: Springer Berlin Heidelberg.
Halbfeld, C., Ebert, B. E. and Blank, L. M. (2014). Multi-capillary column-ion mobility spectrometry of volatile metabolites emitted by *Saccharomyces cerevisiae*. *Metabolites* 4, 751–774.
Hall, M. E., Loeb, G. M., Cadle-Davidson, L., Evans, K. J., and Wilcox, W. F. (2018). Grape sour rot: a four-way interaction involving the host, yeast, acetic acid bacteria, and insects. *Phytolpathol.* 108, 1429-1442.
Hang, S., Purdy, A. E., Robins, W. P., Wang, Z., Mandal, M., Chang, S., Mekalanos,

J. J. and Watnick, P. I. (2014). The acetate switch of an intestinal pathogen disrupts host insulin signaling and lipid metabolism. *Cell Host Microbe* 16, 592–604.

Haro, C., Rangel-Zúñiga, O. A., Alcalá-Díaz, J. F., Gómez-Delgado, F., Pérez-Martínez, P., Delgado-Lista, J., Quintana-Navarro, G. M., Landa, B. B., Navas-Cortés, J. A., Tena-Sempere, M., et al. (2016). Intestinal microbiota is influenced by gender and body mass index. *PLoS One* 11, 1–16.

Herp, S., Brugiroux, S., Garzetti, D., Ring, D., Jochum, L. M., Beutler, M., Eberl, C., Hussain, S., Walter, S., Gerlach, R. G., et al. (2019). *Mucispirillum schaedleri* antagonizes *Salmonella* virulence to protect mice against colitis. *Cell Host Microbe* 25, 681-694.e8.

Hibbing, M. E., Fuqua, C., Parsek, M. R. and Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nat. Rev. Microbiol.* 8, 15–25.

Hoang, D., Kopp, A. and Chandler, J. A. (2015). Interactions between *Drosophila* and its natural yeast symbionts—Is *Saccharomyces cerevisiae* a good model for studying the fly-yeast relationship? *PeerJ* 3, e1116.

Hoffmann, A. A. and Parsons, P. A. (1984). Olfactory response and resource utilization in *Drosophila*: interspecific comparisons. *Biol. J. Linn. Soc.* 22, 43–53.

Hooper, L. V., Littman, D. R. and Macpherson, A. J. (2012). Interactions between the microbiota and the immune system. *Science.* 336, 1268–1273.

Hothorn, T., Bretz, F. and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical J.* 50, 346–363.

Huang, J. H. and Douglas, A. E. (2015). Consumption of dietary sugar by gut bacteria determines *Drosophila* lipid content. *Biol. Lett.* 11, 20150469.

Huang, J. H., Jing, X. and Douglas, A. E. (2015). The multi-tasking gut epithelium of insects. *Insect Biochem. Mol. Biol.* 67, 1–6.

Iatsenko, I., Boquete, J. P. and Lemaitre, B. (2018). Microbiota-derived lactate activates production of reactive oxygen species by the intestinal NADPH oxidase Nox and shortens *Drosophila* lifespan. *Immunity* 49, 929-942.

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

Ja, W. W., Carvalho, G. B., Mak, E. M., de la Rosa, N. N., Fang, A. Y., Liong, J. C., Brummel, T. and Benzer, S. (2007). Prandiology of *Drosophila* and the CAFE assay. *Proc. Natl. Acad. Sci. U. S. A.* 104, 8253–8256.

Jehrke, L., Stewart, F. A., Droste, A. and Beller, M. (2018). The impact of genome variation and diet on the metabolic phenotype and microbiome composition of *Drosophila melanogaster*. *Sci. Rep.* 8, 1–15.

Jolly, N. P., Varela, C. and Pretorius, I. S. (2014). Not your ordinary yeast: Non-*Saccharomyces* yeasts in wine production uncovered. *FEMS Yeast Res.* 14, 215–237.

Judd, A. M., Matthews, M., K., Hughes, R., Veloz, M., Sexton, C. E. and Chaston, J. M. (2018) Bacterial methionine metabolism genes influence *Drosophila*

melanogaster starvation resistance. *Appl. Environ. Microbiol.* 84, e00662-18.
Kamareddine, L., Robins, W. P., Berkey, C. D., Mekalanos, J. J. and Watnick, P. I. (2018). The *Drosophila* immune deficiency pathway modulates enteroendocrine function and host metabolism. *Cell Metab.* 28, 449-462.e5.
Karasov, W. H., Martínez del Rio, C. and Caviedes-Vidal, E. (2011). Ecological physiology of diet and digestive systems. *Annu. Rev. Physiol.* 73, 69–93.
Keebaugh, E. S., Yamada, R., Obadia, B., Ludington, W. B. and Ja W. W. (2018). Microbial quantity impacts *Drosophila* nutrition, development, and lifespan. *iScience* 4, 247-259.
Kešnerová, L., Mars, R. A. T., Ellegaard, K. M., Troilo, M., Sauer, U. and Engel, P. (2017). Disentangling metabolic functions of bacteria in the honey bee gut. *PLoS Biol.* 15, 1–28.
Kim, G., Huang, J. H., McMullen II, J. G., Newell, P. D. and Douglas A. E. (2018). Physiolgoical responses of insects to microbial fermentation products: Insights from the interactions between *Drosophila* and acetic acid. *J. Insect Physiol.* 106, 13-19.
Kimura, I., Miyamoto, J., Ohue-Kitano, R., Watanabe, K., Yamada, T., Onuki, M., Aoki, R., Isobe, Y., Kashihara, D., Inoue, D., et al. (2020). Maternal gut microbiota in pregnancy influences offspring metabolic phenotype in mice. *Science.* 367, eaaw8429.
Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educ. Psychol. Meas.* 56, 746–759.
Knowles, J. E. and Frederick, C. (2019). merTools: Tools for analyzing mixed effect regression models. R package version 0.5.0.
Komsta, L. and Novomestky, F. (2015). moments: Moments, cumulants, skewness, kurtosis and related tests. R package version 0.14.
Koyle, M. L., Veloz, M., Judd, A. M., Wong, A. C. N., Newell, P. D., Douglas, A. E. and Chaston, J. M. (2016). Rearing the fruit fly *Drosophila melanogaster* under axenic and gnotobiotic conditions. *J. Vis. Exp.,* e54219.
Krause, D. J., Kominek, J., Opulente, D. A., Shen, X. X., Zhou, X., Langdon, Q. K., DeVirgilio, J., Hulfachor, A. B., Kurtzman, C. P., Rokas, A., et al. (2018). Functional and evolutionary characterization of a secondary metabolite gene cluster in budding yeasts. *Proc. Natl. Acad. Sci. U. S. A.* 115, 11030–11035.
Kundu, P., Manna, B., Majumder, S. and Ghosh, A. (2019). Species-wide metabolic interaction network for understanding natural lignocellulose digestion in termite gut microbiota. *Sci. Rep.* 9, 1–13.
Kuznetsova, A., Brockhoff, P. B. and Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *J. Stat. Softw.* 82, 1-26.
Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods Ecol. Evol.* 7, 573–579.
Lenth, R. (2019). emmeans: Estimated marginal means, aka least-squares means. R package version 1.4.
Levy, R. and Borenstein, E. (2013). Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12804–12809.

Ma, D., Bou-Sleiman, M., Joncour, P., Indelicato, C. E., Frochaux, M., Braman, V., Litovchenko, M., Storelli, G., Deplancke, B. and Leulier, F. (2019). Commensal gut bacteria buffer the impact of host genetic variants on *Drosophila* developmental traits under nutritional stress. *iScience* 19, 436–447.

Magnúsdóttir, S., Heinken, A., Kutt, L., Ravcheev, D. A., Bauer, E., Noronha, A., Greenhalgh, K., Jäger, C., Baginska, J., Wilmes, P., et al. (2017). Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* 35, 81–89.

Mahowald, M. A., Rey, F. E., Seedorf, H., Turnbaugh, P. J., Fulton, R. S., Wollam, A., Shah, N., Wang, C., Magrini, V., Wilson, R. K., et al. (2009). Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proc. Natl. Acad. Sci.* 106, 5859–5864.

Markle, J. G. M., Frank, D. N., Mortin-Toth, S., Robertson, C. E., Feazel, L. M., Rolle-Kampczyk, U., Von Bergen, M., McCoy, K. D., Macpherson, A. J. and Danska, J. S. (2013). Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science.* 339, 1084–1088.

Martino, M.E., Joncour, P., Leenay, R., Gervais, H., Shah, M., Hughes, S., Gillet, B., Beisel, C. and Leulier F. (2018). Bacterial adaptation to the host's diet is a key evolutionary force shaping *Drosophila-Lactobacillus* symbiosis. *Cell Host Microbe* 24, 109-119.

Mitri, S. and Foster, K.R. (2013). The genotypic view of social interactions in microbial communities. *Annu. Rev. Genet.* 47, 247–273.

Morton, J. T., Aksenov, A. A., Nothias, L. F., Foulds, J. R., Quinn, R. A., Badri, M. H., Swenson, T. L., Van Goethem, M. W., Northen, T. R., Vazquez-Baeza, Y., et al. (2019). Learning representations of microbe–metabolite interactions. *Nat. Methods* 16,.

Murgier, J., Everaerts, C., Farine, J. P. and Ferveur, J. F. (2019). Live yeast in juvenile diet induces species-specific effects on *Drosophila* adult behaviour and fitness. *Sci. Rep.* 9, 1–12.

Musselman, L. P. and Kühnlein, R. P. (2018). *Drosophila* as a model to study obesity and metabolic disease. *J. Exp. Biol.* 221, jeb163881.

Newell, P. D., Chaston, J. M., Wang, Y., Winans, N. J., Sannino, D. R., Wong, A. C., Dobson, A. J., Kagle, J., and Douglas, A. E. (2014). *In vivo* function and comparative genomic analyses of the *Drosophila* gut microbiota identify candidate symbiosis factors. *Front. Microbiol. 5*, 576.

Newell, P. D. and Douglas, A. E. (2014). Interspecies interactions determine the impact of the gut microbiota on nutrient allocation in *Drosophila melanogaster*. *Appl. Environ. Microbiol.* 80, 788–796.

Noecker, C., Chiu, H.-C., McNally, C. P. and Borenstein, E. (2019). Defining and evaluating microbial contributions to metabolite variation in microbiome-metabolome association studies. *mSystems* 4, e00579-19.

Norland, S., Heldal, M. and Tumyr, O. (1987). On the relation between dry matter and volume of bacteria. *Microb. Ecol.* 13, 95–101.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., et al. (2019). vegan:

Community Ecology Package. R package version 2.5-5.
Oude Elferink, S. J. W. H., Krooneman, J., Jan, C., Spoelstra, S. F., Faber, F., Elferink, S. O., Krooneman, J. and Spoelstra, S. F. (2001). Anaerobic conversion of lactic acid to acetic acid and 1,2-propanediol by *Lactobacillus buchneri*. *Appl. Environ. Microbiol.* 67, 125–132.
Palanca, L., Gaskett, A. C., Günther, C. S., Newcomb, R. D. and Goddard, M. R. (2013). Quantifying variation in the ability of yeasts to attract *Drosophila melanogaster*. *PLoS One* 8, e75332.
Piper, M. D. W., Blanc, E., Leitão-Gonçalves, R., Yang, M., He, X., Linford, N. J., Hoddinott, M. P., Hopfen, C., Soultoukis, G. A., Niemeyer, C., et al. (2014). A holidic medium for *Drosophila melanogaster*. *Nat. Methods* 11, 100–105.
Ponomarova, O., Gabrielli, N., Sévin, D. C., Mülleder, M., Zirngibl, K., Bulyha, K., Andrejev, S., Kafkia, E., Typas, A., Sauer, U., et al. (2017). Yeast creates a niche for symbiotic lactic acid bacteria through nitrogen overflow. *Cell Syst.* 5, 345-357.e6.
Quan, A. S. and Eisen, M. B. (2018). The ecology of the *Drosophila*-yeast mutualism in wineries. *PLoS One* 13, 1–25.
R Core Team (2018). R: A language and environment for statistical computing.
Rakoff-Nahoum, S., Foster, K. R. and Comstock, L. E. (2016). The evolution of coooperation within the gut microbiota. *Nature* 533, 255–259.
Read, M. N. and Holmes, A. J. (2017). Towards an integrative understanding of diet-host-gut microbiome interactions. *Front. Immunol.* 8, 1–9.
Rodriguez-Castaño, G. P., Dorris, M. R., Liu, X., Bolling, B. W., Acosta-Gonzalez, A. and Rey, F. E. (2019). *Bacteroides thetaiotaomicron* starch utilization promotes quercetin degradation and butyrate production by *Eubacterium ramulus*. *Front. Microbiol.* 10, 1–8.
Rohlfs, M. and Kürschner, L. (2010). Saprophagous insect larvae, *Drosophila melanogaster*, profit from increased species richness in beneficial microbes. *J. Appl. Entomol.* 134, 667–671.
Rolhion, N. and Chassaing, B. (2016). When pathogenic bacteria meet the intestinal microbiota. *Philos. Trans. R. Soc. B Biol. Sci.* 371, 20150504.
Rossouw, D., Næs, T. and Bauer, F. F. (2008). Linking gene regulation and the exo-metabolome: A comparative transcriptomics approach to identify genes that impact on the production of volatile aroma compounds in yeast. *BMC Genomics* 9, 1–18.
Roy, S., Saha, T. T., Zou, Z. and Raikhel, A. S. (2018). Regulatory pathways controlling female insect reproduction. *Annu. Rev. Entomol.* 63, 489–511.
Scheidler, N. H., Liu, C., Hamby, K. A., Zalom, F. G. and Syed, Z. (2015). Volatile codes: Correlation of olfactory signals and reception in *Drosophila*-yeast chemical communication. *Sci. Rep.* 5, 1–13.
Schwasinger-Schmidt, T. E., Kachman, S. D. and Harshman, L. G. (2012). Evolution of starvation resistance in *Drosophila melanogaster*: measurement of direct and correlated responses to artificial selection. *J. Evol. Biol.* 25, 378–387.
Shin, S. C., Kim, S.-H., You, H., Kim, B., Kim, A. C., Lee, K.-A., Yoon, J.-H., Ryu, J.-H. and Lee, W.-J. (2011). *Drosophila* microbiome modulates host

developmental and metabolic homeostasis via insulin signaling. *Science.* 334, 670–674.
Signorell, A., Aho, K., Alfons, A., Anderegg, N., Aragon, T., Arppe, A., Baddeley, A., Barton, K. and Bolker, B. (2019). DescTools: Tools for descriptive statistics. R package version 0.99.28.
Solomon, G. M., Dodangoda, H., McCarthy-Walker, T. T., Ntim-Gyakari, R. R. and Newell, P. D. (2019). The microbiota of *Drosophila suzukii* influences the larval development of *Drosophila melanogaster*. *PeerJ* 19, e8097.
Sommer, F. and Bäckhed, F. (2013). The gut microbiota - masters of host development and physiology. *Nat. Rev. Microbiol.* 11, 227–238.
Sommer, A. J. and Newell, P. D. (2019). Metabolic basis for mutualism between gut bacteria and its impact on the *Drosophila melanogaster* host. *Appl. Environ. Microbiol.* 85, 1–15.
Storelli, G., Defaye, A., Erkosar, B., Hols, P., Royet, J. and Leulier, F. (2011). *Lactobacillus plantarum* promotes *Drosophila* systemic growth by modulating hormonal signals through TOR-dependent nutrient signaling. *Cell Metab.* 14, 403-414.
Tefit, M. A. and Leulier, F. (2017). *Lactobacillus plantarum* favors the early emergence of fit and fertile adult *Drosophila* under chronic undernutrition. *J. exp. Biol.* 220, 900-907.
Terashima, J. and Bownes, M. (2004). Translating available food into the number of eggs laid by *Drosophila melanogaster*. *Genetics* 167, 1711–1719.
Thaiss, C. A., Zmora, N., Levy, M. and Elinav, E. (2016). The microbiome and innate immunity. *Nature* 535, 65–74.
Therneau, T. M. (2015). A package for survival analysis in S. version 2.38.
Therneau, T. M. (2019). coxme: Mixed effects Cox models. R package version 2.2-14.
Tripathi, A., Marotz, C., Gonzalez, A., Vázquez-Baeza, Y., Song, S. J., Bouslimani, A., McDonald, D., Zhu, Q., Sanders, J. G., Smarr, L., et al. (2018). Are microbiome studies ready for hypothesis-driven research? *Curr. Opin. Microbiol.* 44, 61–69.
van de Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K. and van der Werf, M. J. (2006). Centering, scaling, and transformations: imporving the biological informaiton content of metabolomics data. *BMC Genomics* 7, 142.
Venturelli, O. S., Carr, A. V, Fisher, G., Hsu, R. H., Lau, R., Bowen, B. P., Hromada, S., Northen, T. and Arkin, A. P. (2018). Deciphering microbial interactions in synthetic human gut microbiome communities. *Mol. Syst. Biol.* 14, 1–19.
Wat, L. W., Chao, C., Bartlett, R., Buchanan, J. L., Millington, J. W., Chih, H. J., Chowdhury, Z. S., Biswas, P., Huang, V., Shin, L. J., et al. (2020). A role for triglyceride lipase brummer in the regulation of sex differences in *Drosophila* fat storage and breakdown. *PLoS Biol.* 18, e3000595.
Weger, B. D., Gobet, C., Yeung, J., Martin, E., Jimenez, S., Betrisey, B., Foata, F., Berger, B., Balvay, A., Foussier, A., et al. (2019). The mouse microbiome is required for sex-specific diurnal rhythms of gene expression and metabolism. *Cell Metab.* 29, 362-382.
White, M. K., Matthews, M. K., Hughes, R. C., Sommer, A. J., Griffitts, J. S., Newell,

P. D. and Chaston, J. M. (2018). A metagenome-wide association study and arrayed mutant library confirm *Acetobacter* lipopolysaccharide genes are necessary for association with *Drosophila melanogaster*. *G3 Genes, Genomes, Genet.* 8, 1119–1127.

White, T.J., Bruns, T., Lee, S., and Taylor, J.W. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols: A Guide to Methods and Applications* (ed. Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J.), pp. 315–322. New York, NY: Academic Press Inc.

Winans, N. J., Walter, A., Chouaia, B., Chaston, J. M., Douglas, A. E., and Newell, P. D. (2017). A genomic investigation of ecological differentiation between free-living and *Drosophila*-associated bacteria. *Mol Ecol. 26*, 4536-4550.

Wintermute, E. H. and Silver, P. A. (2010). Emergent cooperation in microbial metabolism. *Mol. Syst. Biol.* 6, 1–7.

Wong, A. C. N., Ng, P. and Douglas, A. E. (2011). Low-diversity bacterial community in the gut of the fruitfly *Drosophila melanogaster*. *Environ. Microbiol.* 13, 1889–1900.

Wong, A. C. N., Dobson, A. J. and Douglas, A. E. (2014). Gut microbiota dictates the metabolic response of *Drosophila* to diet. *J. Exp. Biol.* 217, 1894–1901.

Wong, A. C. N., Luo, Y., Jing, X., Franzenburg, S., Bost, A. and Douglas, A. E. (2015). The host as driver of the microbiota in the gut and external environment of *Drosophila melanogaster*. *Appl. Environ. Microbiol.* 81, 6232-6240.

Wong, A. C. N., Vanhove, A. S. and Watnick, P. I. (2016). The interplay between intestinal bacteria and host metabolism in health and disease: lessons from *Drosophila melanogaster*. *Dis. Model. Mech.* 9, 271-281.

Wong, A.C.N., Wang, Q-P., Morimoto, J., Senior, A.M., ihoreau, M., Neeley, G.G., Simpson, S.J. and Ponton, F. (2017). Gut microbiota modifies olfactory-guided microbial preferences and foraging decisions in *Drosophila*. *Curr. Biol.* 27, 2397-2406.

Yamada, R., Deshpande, S.A., Bruce, K.D., Mak, E.M. and Ja, W.W. (2015). Microbes promote amino acid harvest to resuce undernutrition in *Drosophila*. *Cell Reports* 10, 865-872.

Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D. R., Bork, P. and Patil, K. R. (2015). Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* 112, 6449-6454.

Zheng, X., Qiu, Y., Zhong, W., Baxter, S., Su, M., Li, Q., Xie, G., Ore, B. M., Qiao, S., Spencer, M. D., et al. (2013). A targeted metabolomic protocol for short-chain fatty acids and branched-chain amino acids. *Metabolomics* 9, 818–827.

Zheng, H., Perreau, J., Elijah Powell, J., Han, B., Zhang, Z., Kwong, W. K., Tringe, S. G. and Moran, N. A. (2019). Division of labor in honey bee gut microbiota for plant polysaccharide digestion. *Proc. Natl. Acad. Sci. U. S. A.* 116, 25909–25916.

CHAPTER 4

IMPACT OF MICROBIOME COMPOSITION ON THE METABOLOME OF *DROSOPHILA MELANOGASTER*[3]

Abstract

Despite the abundant evidence that the gut microbiome influences the metabolic health of the animal host, understanding of how the microbiome composition influences host metabolism is fragmented and incomplete. To obtain a global overview of the metabolic consequences of the microbiome, we determined the metabolome of *Drosophila melanogaster* experimentally associated with the bacteria *Acetobacter fabarum* and *Lactobacillus brevis* and the yeast *Hanseniaspora uvarum* in communities of one, two, or three taxa. Of 153 metabolites identified, the abundance of 50 (34%) varied significantly with microbial treatment and sex. Titers of various nucleotides and amino acids were elevated in females with di-associations that included *Acetobacter*, while central carbon intermediates were enriched in males with di- and tri-associations. However, the most robust correspondence between metabolites and two key indices of metabolic health, lipid and glucose content, related to a microbial metabolite, phenyllactic acid, which was correlated negatively with glucose (both sexes) and positively with lipid content (males). This study reveals that, although microbiome composition and nutritional effects are similar for males and

[3] Article in preparation for journal submission by McMullen, J.G., Koo, I., Vijay, A.M., Patterson, A.D., and Douglas, A.E.
Supplemental material are provided in Appendix C.
Koo, I. processed the metabolomics data to select peak areas and metabolite identifications and Vijay, A.M. performed metabolite extraction.

females, the underlying metabolic basis differs between the sexes, and it identifies phenyllactic acid as a candidate microbial determinant of the metabolic health of *Drosophila*.

Introduction

The microorganisms that colonize the animal gut, collectively referred to as the microbiome, have diverse effects on the physiology and health of the animal host, with impacts on nutrition, growth, immune function, and behavior (McFall-Ngai et al., 2013; Morais et al., 2020; Read and Holmes, 2017; Rolhion and Chassaing, 2016; Sommer and Bäckhed, 2013; Thaiss et al., 2016). Most host-microbiome interactions are metabolic in nature and microbiome effects on the animal metabolome can be categorized as either direct effects of microbial-produced or -processed nutrients or indirect effects of microbial-derived effectors on host signaling pathways involved in metabolic regulation and nutrient allocation (Ankrah and Douglas, 2018; Douglas, 2014; Engel and Moran, 2013; Hooper et al., 2002). Although it is widely accepted that the microbiome influences the animal metabolome (Nicholson et al., 2012), much of our understanding of host-microbiome metabolic interactions comes from investigations using metagenomic approaches e.g. Armour et al. (2019; Fromont et al. (2019; Kau et al. (2011) or hosts experimentally colonized with a single microbial partner (Consuegra et al., 2020b; Judd et al., 2018; Kešnerová et al., 2017). There is, however, increasing interest in how among-microbe interactions influence animal physiology (Consuegra et al., 2020a; Coyte and Rakoff-Nahoum, 2019; Figueiredo and Kramer, 2020; Gould et al., 2018; Granato et al., 2019), but current information is

fragmentary.

Simple animal models can be powerful experimental systems to investigate the effects of among-microbe interactions on the host metabolic function (Douglas, 2019). Specifically, *Drosophila melanogaster* is tractable for metabolomics-based studies (Cox et al., 2017; Jin et al., 2020; Li and Tennessen, 2017; Musselman and Kühnlein, 2018; Zhou et al., 2019) and a fast-emerging model for animal-microbiome interactions (Broderick and Lemaitre, 2012; Douglas, 2019; Erkosar et al., 2013; Wong et al., 2016). The dominant microorganisms associated with the *Drosophila* gut are bacteria in the family Acetobacteraceae and order Lactobacillales and the Saccharomycetales yeasts (Adair and Douglas, 2017; Chandler et al., 2012, 2011; Quan and Eisen, 2018; Wong et al., 2013). Currently, several lines of evidence support direct effects of the microorganisms on *Drosophila* nutrition through competition for dietary glucose to reduce lipid storage (Chaston et al., 2014; Huang and Douglas, 2015; Sommer and Newell, 2019) and indirect effects of microbial metabolism on host nutritional physiology through insulin-like signaling via acetic acid and branched chain amino acids (Dobson et al., 2016; Shin et al., 2011; Storelli et al., 2011). Host nutrient allocation patterns are also influenced by among-microbe interactions, primarily between *Acetobacter-Lactobacillus* (Aranda-Díaz et al., 2020; Consuegra et al., 2020a; Gould et al., 2018; Newell and Douglas, 2014; Sommer and Newell, 2019) and *Acetobacter*-yeast (Fischer et al., 2017; McMullen et al., 2020). In particular, the interaction between *Acetobacter*-yeast increases *Drosophila* acetic acid titers and lowers lipid content (McMullen et al., 2020); however, the underlying metabolic processes have not yet been resolved.

The goal of this study was to investigate how co-colonization of *Drosophila* with different gut microorganisms influenced *Drosophila* metabolic function, as revealed by the fly metabolome. For this study, we used a combinatorial experimental approach and associated *Drosophila* of each sex with one, two, or three representative taxa of the three dominant microbiome groups (the bacteria *Acetobacter fabarum* and *Lactobacillus brevis* and the yeast *Hanseniaspora uvarum*). The metabolomics data presented were combined with data from McMullen et al. (2020) to link metabolite titers with microbial abundance and indices of fly nutritional status. This study identified significant differences in metabolites from several pathways with respect to both microbial complement and host sex. The pattern of variation points to microbiome effects on the function of central carbon, amino acid, and nucleotide metabolism, which may be influenced through microbiome-mediated modulation of metabolic signaling pathways.

Materials and methods

Experimental design

D. melanogaster were experimentally associated with three microorganisms, *Acetobacter fabarum* (ACE), *Lactobacillus brevis* (LAC), and *Hanseniaspora uvarum* (YST), in all possible combinations of mono-, di-, and tri-associations with an axenic (microbe-free) control. Full description of the experimental design is available in McMullen et al. (2020). Briefly, *Drosophila* eggs were dechorionated in bleach following Koyle et al. (2016) and microorganisms were added at a dose of 5×10^6 cells with co-associations combined in equal proportions onto a diet of 10% inactive

brewer's yeast (MP Biomedicals), 10% glucose (Sigma), and 1.2% *Drosophila* type II agar (Apex). Insects were allowed to develop until one day post eclosion (dpe), at which time adults were transferred to fresh, sterile diet using aseptic technique. At 5 dpe, flies were harvested, sorted by sex, and weighed. Samples were pooled for microbial abundance (2 flies per sex), nutritional indices (5 flies per sex), and global metabolomics screen (ca. 20 mg flies per sex). The first two measures were previously reported in McMullen et al. (2020). Microbial abundance was surveyed by colony forming units (CFU) using a WASP-2 spiral platter and counted with a Protocol 3 instrument (Microbiology International). For nutritional indices, glucose, protein, and triglyceride (TAG) content were measured using spectrophotometric enzyme assays (glucose: Sigma GAGO20, protein: BioRad DC kit, and TAG: L9518 & F6428); data was autoscaled by experiment replicate for correlations with individual metabolites.

Global metabolite profiling assay

Pools of ca. 20 mg of flies for each sex were flash frozen in liquid nitrogen and stored at -80°C until processing for global metabolomics screen at the Penn State Metabolomics Facility, University Park, PA, USA following Blow et al. (2020). Metabolites were extracted from samples in 1 mL ice-cold 3:3:2 acetonitrile:isopropyl alcohol:water containing the internal standards alanine, pyruvate, isoleucine, glutamate, uracil, and tyrosine and 0.1 mm zirconia/silica beads (BioSpec. Products, Bartlesville, OK, USA) by homogenizing samples for two 20 sec cycles at 6,500 rpm with a Precellys 24 homogenizer (Bertin Technologies, Rockville, MD, USA) and were shaken for an additional 6 min at 4°C. Samples were placed on ice in between

cycles. Homogenates were centrifuged for 10 min at 20,187 g and 4°C. The supernatant was transferred, dried using a SpeedVac (Thermo Fisher Scientific), and washed in 500 μL 1:1 acetonitrile:water. Following the previous conditions, the supernatant was centrifuged, transferred, dried, and resuspend in 100 μL 3% methanol:97% water containing 1 μM internal standard chlorpropamide (Santa Cruz Biotech). A blank sample was included as a control. Liquid chromatography-mass spectrometry (LC-MS) was performed using a Dionex Ultimate 3000 quaternary high-performance liquid chromatography (HPLC) system coupled with an Exactive Plus Orbitrap mass spectrometer controlled by Xcalibur 2.2 software (Thermo Fisher Scientific) following the conditions described by Cai et al. (2018).

Data generated from LC-MS were processed using MS-DIAL (Tsugawa et al., 2015). The raw instrument files to mzML format using Proteowizard (Chambers et al., 2012). MS1 tolerance was set to 0.001 Da. Linear weighted moving average was applied (smoothing level = 3) to chromatograms. Minimum peak height and width were set to 1,000 and 5, respectively, to select peaks. An identification threshold of 85% with retention time tolerance of 0.5 min and accurate mass of 0.001 was implemented to identify peaks from an internal curated library of metabolites. Alignment thresholds of retention time and MS1 were set to 0.35 min and 0.0015. Peak areas were normalized by subtracting the blank peak area, dividing by the internal standard chlorpropamide peak area and sample mass, and multiplying by the average peak area of the internal standard to retain original scaling. Data were further range scaled (van den Berg et al., 2006) to control for variation between experimental

replicates, which was used for all statistical analyses unless specified. Metabolite classification was determined using Chemical Translation Service (Wohlgemuth et al., 2010) and ClassyFire (Djoumbou Feunang et al., 2016).

Statistics

All statistical analyses were performed using R version 4.0.2 (R Core Team, 2018) with an α of 0.05. For all analyses, female and male flies were analyzed separately, following our previous demonstration that male and female flies differ in their metabolic response to similar microbiome composition (McMullen et al., 2020). An analysis of variance (ANOVA) was applied to assess the effect of microbial treatment on range scaled metabolite concentration, richness, and diversity with the car package version 3.0.10 (Fox and Weisberg, 2019). Residuals from models were visually assessed for normality and homoscedasticity. A Benjamini-Hochberg false discovery rate p-value correction method (FDR) was applied for multiple comparisons with a threshold of 0.2 (Efron, 2007), when required. Richness and diversity were calculated by number of metabolites and inverse Simpson index with vegan package version 2.5.6 (Oksanen et al., 2019) using the normalized dataset prior to range scaling.

Several multivariate statistics were implemented to visualize and associate metabolome profiles with microorganisms. First, a principal components analysis was used to identify the first two components of the metabolome data using a correlation matrix with the 'rda' function from vegan. Each component was then correlated with microbial abundance (log_{10}-transformed) using the vegan 'envfit' function. Second, permutational multivariate analysis of variance (PERMANOVA) was implemented

with a full factorial analysis (three-way interaction for presence of each microorganism) using the 'adonis' function with 999 permutations. Lastly, a partial least squares-discriminant analysis (PLS-DA) was implemented to categorize metabolites by microbial treatment on the first two components (with 1,000 permutations to assess model significance) using the ropls package version 1.20.0 (Thévenot et al., 2015). The top metabolites that varied with the fly microbiome were identified by selecting variable importance in projection (VIP) scores >1.

Metabolites determined from PLS-DA were correlated with nutritional indices using a Spearman's rank correlation method with FDR correction. A heatmap was produced using pheatmap package version 1.0.12 (Kolde, 2019) with an agglomerative hierarchical cluster applied to mean response of metabolite concentration by treatment with Euclidean distance and Ward's linkage. *Post hoc* Tukey's test with the emmeans package 1.5.1 (Lenth, 2019) or a full factorial ANOVA was implemented to identify differences among treatments for each metabolite.

Results

Identification of microbial-responsive metabolites

Our first approach to assess the effect of the microbiome on the *Drosophila* metabolome was to identify metabolites that vary with microbial treatment. In total, 153 metabolites were identified from the global metabolomics screen. Of these, 50 metabolites (34%) were identified as microbial-responsive after p-value correction (Table S3.1A), including 9% (13 metabolites) identified in both female and male flies, 16% (24) specific to females, and 9% (13) specific to males (Table S3.1B). The

majority of metabolites found in female and male flies were amino acids, carbohydrates, and nucleotides (Table S3.1B). In female flies, richness and diversity of microbial-responsive metabolites differed significantly by treatment in a full factorial ANOVA with the presence of ACE and YST increasing number of metabolites (Table 3.1). The interaction of ACE with either LAC or YST decreased or increased metabolite diversity (Table 3.1), suggesting that the interactions among-microbes influenced the richness and relative abundance of metabolome profiles. Although the concentration of 26 metabolites varied with microbial treatment in males (see above), no significant differences in richness or diversity of metabolites were obtained, and suggests that male flies maintain metabolite diversity by shifting metabolite concentrations in response to microorganisms present (Table 3.1).

Table 3.1. Effect of microbiome on metabolite richness and diversity by treatment.

Microbial treatments	Female		Male	
	Richness	Inverse Simpson	Richness	Inverse Simpson
(a) Estimated marginal means (standard error)				
Axenic	31.2 (0.7)	4.8 (0.5)	23.7 (0.6)	3.0 (0.3)
ACE	33.5 (0.7)	4.4 (0.5)	23.2 (0.6)	2.5 (0.3)
LAC	32.8 (0.7)	5.2 (0.5)	22.5 (0.6)	3.0 (0.3)
YST	33.0 (0.7)	4.9 (0.5)	22.5 (0.6)	2.6 (0.3)
ACE+LAC	32.8 (0.7)	2.4 (0.5)	24.0 (0.6)	3.0 (0.3)
ACE+YST	34.0 (0.7)	5.7 (0.5)	22.2 (0.6)	3.0 (0.3)
LAC+YST	33.2 (0.7)	5.3 (0.5)	24.2 (0.6)	3.3 (0.3)
ACE+LAC+YST	34.7 (0.7)	5.3 (0.5)	23.0 (0.6)	2.9 (0.3)
(b) Full factorial ANOVA F statistics (df = 1,40)				
ACE	6.2*	3.3	0.1	0.5
LAC	0.9	1.2	1.4	1.7
YST	5.4*	11.0**	0.7	0.4
ACE*LAC	0.9	5.9*	0.4	0.1
ACE*YST	0.0	9.2**	1.9	0.4
LAC*YST	0.0	1.4	2.4	0.0
ACE*LAC*YST	2.1	1.3	2.4	2.5

* $p < 0.05$; ** $p < 0.01$.

Patterns of microbiome-dependent variation in the Drosophila *metabolome*

To visualize how the metabolome varies with the composition of the microbiome, a principal component analysis (PCA) was performed for each sex. For both males and females, mono-associations (i.e. with a single microbial taxon) were separated from co-association treatments (Fig. 3.1). The axenic flies overlapped with the mono-association treatments, although female replicates were more variable than male samples (Fig. 3.1). On the second axis, the co-association treatments in female and male flies largely separated by the presence of ACE. A multivariate correlation between principal components and microbial abundance indicated that metabolome profiles of both fly sexes were associated with ACE and LAC abundance in male flies

(Fig. 3.1 & Table S3.2A). PERMANOVA results indicated that the presence of all three microbes influenced the *Drosophila* metabolome by three-way interaction in females and all two-way interactions in males (Table S3.2B). For female flies, all four co-association treatments were clustered separately, while male flies indicated that all three di-associations were distinct from each other and the tri-association overlapped with the ACE+LAC treatment (Fig. 3.1).

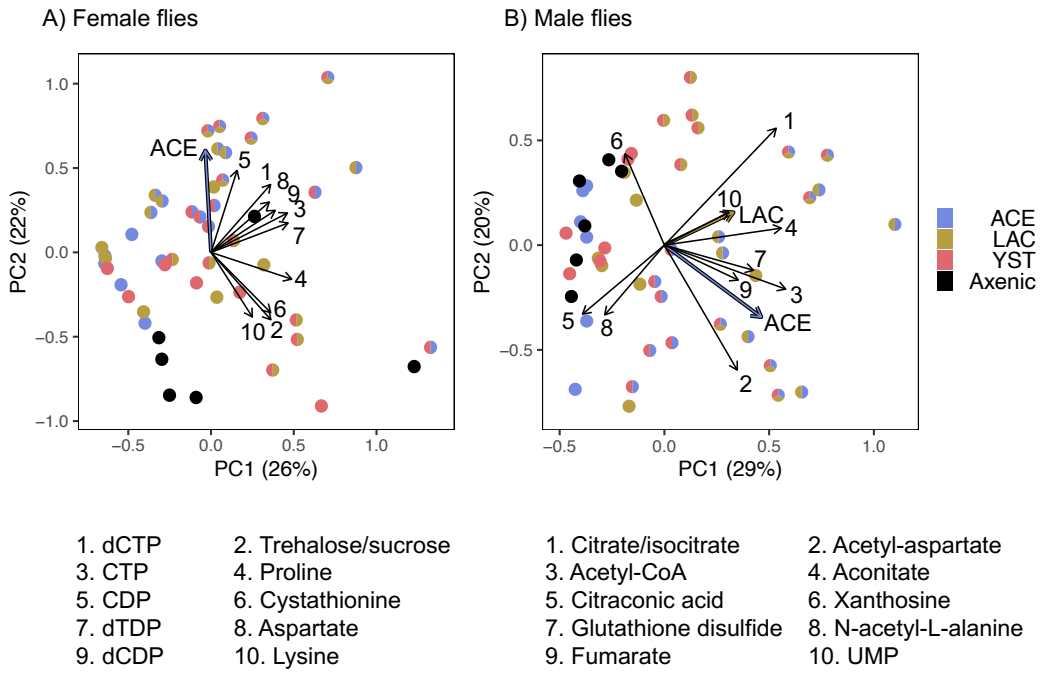


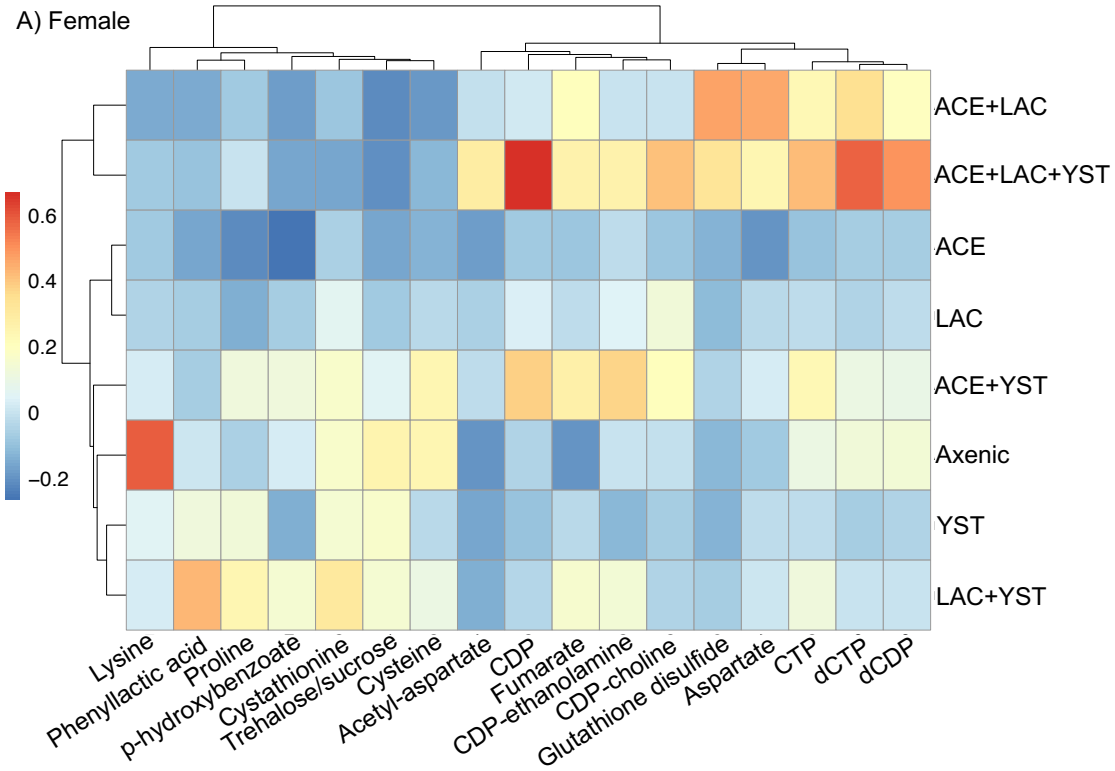
Figure 3.1. Effect of microorganisms on the metabolome of *Drosophila*. PCA of microbial-responsive metabolites in (A) female and (B) male flies. Symbols are colored according to microbial treatment. Top 10 metabolite loadings are displayed for each sex (black arrows). The colored arrows correspond to significant multivariate correlations with microbial abundance (log_{10}-transformed [CFU fly^{-1} +1]). Percent variation explained on each axis is shown.

The top 10 loadings were shown to highlight metabolites associated with the microbial treatments. All 10 metabolites differed between female and male flies, indicative of distinctly different metabolic patterns for the two sexes (Fig. 3.1). Metabolites identified in female flies consisted of pyrimidine related molecules, amino acids (including derivatives), and carbohydrates, which had higher titers in the co-association treatments (Fig. 3.1A). Male flies were associated with TCA cycle intermediates, a purine nucleoside, a pyrimidine nucleotide, and amino acid derivatives, which were mostly positively correlated with ACE co-association treatments (Fig. 3.1B).

Our subsequent analysis focused on the variation in metabolites with microbial treatments. A partial least squares-discriminant analysis (PLS-DA) was used and 17 and 8 metabolites were identified in female and male flies, respectively (Fig. S3.1 & Table S3.3A), including four metabolites in both sexes: acetyl-aspartate, fumarate, glutathione disulfide and phenyllactic acid (Table S3.3B). Most of the metabolites identified in the PCA (Fig. 3.1) were supported with the PLS-DA (Fig. S3.1).

A hierarchical cluster was then used to identify patterns among treatments and metabolites. In both sexes, the ACE+LAC di-association and the tri-association clustered together while the pattern for the rest of the treatments differed between males and females (Fig. 3.2). For female flies, treatments containing both ACE and LAC had elevated titers of pyrimidine nucleotides, some amino acids (including derivatives), and the TCA cycle intermediate, fumarate. Higher concentrations for amino acids and derivatives, ubiquinone intermediate (p-hydroxybenzoate), and a carbohydrate were found in bacteria-YST di-associations, YST-alone, or axenic fly

treatments (Fig 3.2A). Contrastingly, TCA cycle intermediates and amino acid derivatives were enriched in male flies, particularly for the ACE+LAC treatments. The remainder of the male microbial treatments generally had low titers of all metabolites, except the LAC+YST di-association which had the highest concentrations observed for phenyllactic acid and the TCA cycle intermediate, citrate/isocitrate (Fig 3.2B). Axenic flies of both sexes had relatively low to moderate levels for most metabolites, apart from lysine in female flies. Although lysine catabolism is hypothesized to be in part microbial (St. Clair et al., 2017), the process driving sex-specific response observed is unclear.


A) Female
ACE+LAC
ACE+LAC+YST
ACE
LAC
ACE+YST
Axenic
YST
LAC+YST
0.6
0.4
0.2
0
−0.2
Lysine
Phenyllactic acid
Proline
p-hydroxybenzoate
Cystathionine
Trehalose/sucrose
Cysteine
Acetyl-aspartate
CDP
Fumarate
CDP-ethanolamine
CDP-choline
Glutathione disulfide
Aspartate
CTP
dCTP
dCDP

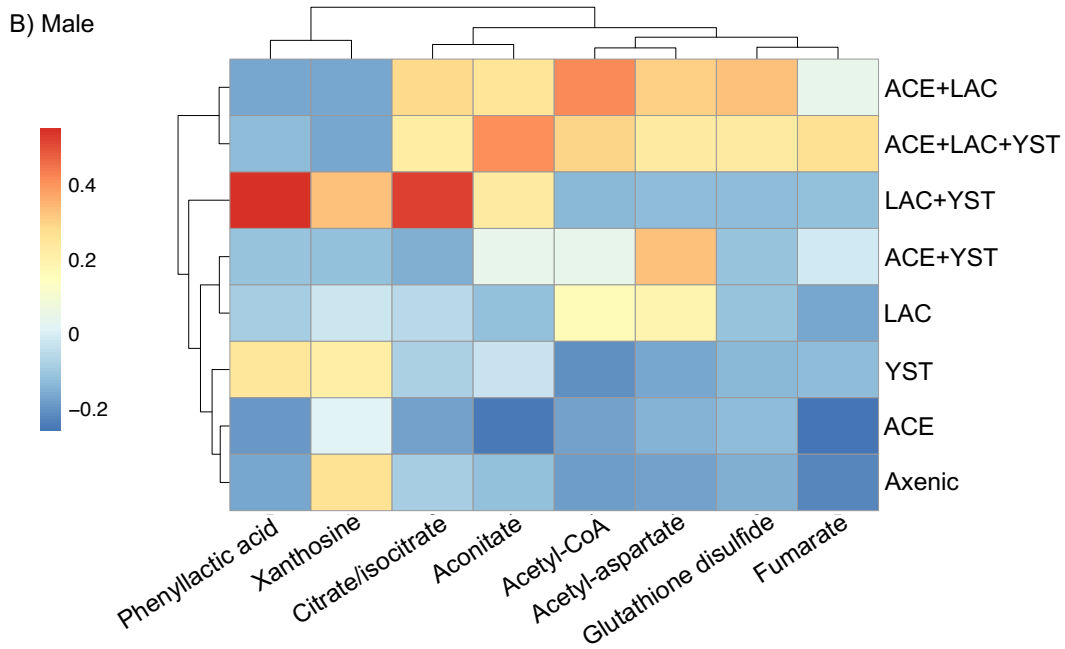

B) Male
ACE+LAC
ACE+LAC+YST
LAC+YST
ACE+YST
LAC
YST
ACE
Axenic
0.4
0.2
0
−0.2
Phenyllactic acid
Xanthosine
Citrate/isocitrate
Aconitate
Acetyl-CoA
Acetyl-aspartate
Glutathione disulfide
Fumarate

Figure 3.2. Variation in metabolite profile by treatment. Heatmap with hierarchical cluster dendrograms for (A) female and (B) male flies. Values reported are the range scaled metabolite concentrations. For among-treatment differences, see Fig. S3.2 for individual plots of each metabolite for both sexes.

Correlation between individual metabolites and host nutritional status

Individual metabolites (identified from the PLS-DA VIP scores) were correlated with nutritional indices (glucose, protein, and TAG content) to identify metabolites associated with *Drosophila* nutrient allocation patterns (Table S3.4). A total of six metabolites were identified (Table 3.2), of which three are TCA cycle intermediates associated with glucose (aconitate, fumarate) or TAG content of males (acetyl CoA). A nucleoside (xanthosine) had a significant relationship with male protein content. In addition, one amino acid, proline, was correlated with glucose in females, and the amino acid derivative, phenyllatic acid, was negatively associated with glucose in both sexes and positively associated with TAG in males.

Table 3.2. Spearman rank correlation between nutritional indices and individual metabolites.

Nutritional index	Metabolite	Rho	q value
(a) Female flies			
Glucose	Phenyllactic acid	-0.65	8.8×10^{-5}
Glucose	Proline	-0.48	0.02
(b) Male flies			
Glucose	Phenyllactic acid	-0.63	3.3×10^{-5}
Glucose	Aconitate	-0.38	0.06
Glucose	Fumarate	-0.34	0.09
Protein	Xanthosine	-0.46	0.01
TAG	Phenyllactic acid	0.70	1.3×10^{-6}
TAG	Acetyl-CoA	-0.33	0.11

Discussion

Variation in the metabolome with gut microbiome composition provides insight into

how microorganisms influence metabolic function, and in turn, nutrient allocation of the animal host. Here, we demonstrated that gut microorganisms in *Drosophila* can have a significant effect on metabolites involved in various metabolic pathways (e.g. central carbon, amino acid, nucleotide metabolism), but the pattern of these effects differs between female and male insects. In addition, several individual metabolites correlated with fly glucose, protein, and lipid content, offering a basis for further investigation of the metabolic basis of nutrient allocation.

Our interpretation of the metabolomics data is based on the expectation that the majority of the metabolites observed are reflective of host cellular metabolism because the gut microbiome contributes < 1% of the total fly biomass (Keebaugh et al., 2018; McMullen et al., 2020). This differs from many mammalian metabolomics studies, which focus on gut tissues and body fluids (e.g. urine, blood), in which microbial-derived metabolites are abundant (Vernocchi et al., 2016). Nevertheless, one metabolite highlighted by our statistical analysis, phenyllactic acid, is likely microbial in origin. *Drosophila* has no known genetic capacity to synthesize phenyllactic acid (Kanehisa and Goto, 2000), but both lactic acid bacteria and various yeasts synthesize this compound (Ruggirello et al., 2019; Storelli et al., 2018; Svanström et al., 2013). A further important issue for interpreting the metabolomic data relates to the many metabolites identified that did not differ significantly with microbial treatments. Additional metabolites of interest may be recognized in an expanded analysis that includes a greater diversity of microbial treatments or larger sample sizes. In addition, metabolomics has the limitation that it offers a steady-state view of metabolic function, and cannot detect biologically-important differences in metabolic flux that

retain metabolite pools of uniform size.

The very substantial differences in the composition of microbial-responsive metabolites between female and male flies parallels previously identified sex differences found in nutrient allocation and molecular function (Bost et al., 2018; Jehrke et al., 2018; McMullen et al., 2020; Ridley et al., 2013; Wong et al., 2014). These several effects are likely due to sex-specific strategies for metabolic investment in reproduction (Camus et al., 2019) with female investment in lipid and protein synthesis for vitellogenesis and egg production (Bownes et al., 1988; Roy et al., 2018; Terashima and Bownes, 2004) and costs in male flies associated with courtship and general vigor (Harvanek et al., 2017; Jensen et al., 2015; McKean and Nunney, 2001). For male flies, results indicated a shift in the metabolic function from mono-association to co-associations with maintenance of metabolite diversity across treatments, while increasing microbiome community complexity for female flies resulted in either increased (with ACE+YST treatments) or decreased (with ACE+LAC di-association) metabolite diversity. Many of the top metabolites were related to products of central carbon, amino acid, and nucleotide metabolism, which indicates changes in metabolic pathways central to energy production and biosynthesis and likely TOR/insulin-like signaling (Mattila and Hietakangas, 2017; Wilinski et al., 2019).

Overlaying the sex-specific effects of the microbial treatments is the common feature that the di-association with ACE and LAC and the tri-association had the greatest effect on enriching metabolite titers in each sex. Functional interpretation of differences in metabolite titers is constrained without complementary flux data

because, for example, high titers can be indicative of increased metabolic importance or reduced utilization, while low titers can be explained as reduced production or increased turnover. Nevertheless, the overall pattern of metabolic effects of ACE and LAC is consistent with increased central carbon metabolism in two respects. First, the interpretation that increased titers of TCA cycle intermediates indicates increased TCA activity and is supported by the elevated glutathione disulfide titer, associated with oxidative stress as observed in the mouse model (Gansemer et al., 2020); and, second, the increased acetyl-aspartate concentration, fits to prior evidence linking this metabolite with energy metabolism in *Drosophila* (Wilinski et al., 2019) and the use of low acetyl-aspartate titer as a biomarker for obesity in humans (Coplan et al., 2014; Kaur et al., 2017). Reinforcing these lines of evidence, the titers of several TCA cycle intermediates were negatively correlated with glucose and TAG content in males. In addition, female flies displayed a negative correlation between proline titers and glucose content, which may indicate changes in energy metabolism due to proline utilization as a respiratory fuel (Goncalves et al., 2014).

Two additional metabolites were correlated with nutritional indices and may serve as biomarkers for *Drosophila* nutritional status. First, phenyllactic acid was positively correlated with lipid content in males and negatively correlated with glucose content in both sexes. The male-specific association of phenyllactic acid with lipid content may be a general animal feature, because this microbial metabolite is also elevated in urine of hyperlipidemic male rats (Wu et al., 2014). Second, xanthosine, which is part of purine degradation to uric acid, a major nitrogenous waste product of *Drosophila* (Dow and Davies, 2003), was negatively associated with protein content of males, and

this correlation may indicate male-specific patterns of nitrogen metabolism that are influenced by the microbiome.

We conclude with considering how these data contribute to our understanding of how the microbiome shapes animal metabolism. From an evolutionary standpoint, many of the interactions between animals and their gut microbiomes, including microbial effects on host metabolic signaling (Gérard and Vidal, 2019; Shin et al., 2011; Storelli et al., 2011), are highly conserved across the animal kingdom (Douglas, 2019; Musselman and Kühnlein, 2018). The evolutionary drivers are not well-understood but can involve both microbial manipulation of host nutritional status and life history traits, as well as host utilization of microbial metabolic activity as a cue for adaptive trait responses. In particular, an important outstanding question is the effects of microorganisms on host signaling pathways, which are likely adaptive for the animal host by allowing for shifts in metabolic activity in response to their environment. The use of *Drosophila*, as it is amenable to large experiments and combinatorial designs for microbiome treatments, e.g. Gould et al. (2018; McMullen et al. (2020) and this study, offers the opportunity to interrogate the processes determining the effects of microbiome composition on animal metabolic function.

References

Adair, K.L., Douglas, A.E., 2017. Making a microbiome: the many determinants of host-associated microbial community composition. Current Opinion in Microbiology 35, 23–29. https://doi.org/10.1016/j.mib.2016.11.002

Ankrah, N.Y.D., Douglas, A.E., 2018. Nutrient factories: metabolic function of beneficial microorganisms associated with insects. Environmental Microbiology 20, 2002–2011. https://doi.org/10.1111/1462-2920.14097

Aranda-Díaz, A., Obadia, B., Dodge, R., Thomsen, T., Hallberg, Z.F., Güvener, Z.T., Ludington, W.B., Huang, K.C., 2020. Bacterial interspecies interactions modulate pH-mediated antibiotic tolerance. eLife 9, e51493. https://doi.org/10.7554/eLife.51493

Armour, C.R., Nayfach, S., Pollard, K.S., Sharpton, T.J., 2019. A Metagenomic Meta-analysis Reveals Functional Signatures of Health and Disease in the Human Gut Microbiome. mSystems 4. https://doi.org/10.1128/mSystems.00332-18

Blow, F., Ankrah, N.Y.D., Clark, N., Koo, I., Allman, E.L., Liu, Q., Anitha, M., Patterson, A.D., Douglas, A.E., 2020. Impact of Facultative Bacteria on the Metabolic Function of an Obligate Insect-Bacterial Symbiosis. mBio 11. https://doi.org/10.1128/mBio.00402-20

Bost, A., Franzenburg, S., Adair, K.L., Martinson, V.G., Loeb, G., Douglas, A.E., 2018. How gut transcriptional function of Drosophila melanogaster varies with the presence and composition of the gut microbiota. Molecular Ecology 27, 1848–1859. https://doi.org/10.1111/mec.14413

Bownes, M., Scott, A., Shirras, A., 1988. Dietary components modulate yolk protein gene transcription in Drosophila melanogaster. Development 103, 119–128.

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

Cai, J., Nichols, R.G., Koo, I., Kalikow, Z.A., Zhang, L., Tian, Y., Zhang, J., Smith, P.B., Patterson, A.D., 2018. Multiplatform Physiologic and Metabolic Phenotyping Reveals Microbial Toxicity. mSystems 3. https://doi.org/10.1128/mSystems.00123-18

Camus, M.F., Piper, M.D., Reuter, M., 2019. Sex-specific transcriptomic responses to changes in the nutritional environment. eLife 8, e47262. https://doi.org/10.7554/eLife.47262

Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T.A., Brusniak, M.-Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S.L., Nuwaysir, L.M., Lefebvre, B., Kuhlmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, T., Holly, K., Eckels, J., Deutsch, E.W., Moritz, R.L., Katz, J.E., Agus, D.B., MacCoss, M., Tabb, D.L., Mallick, P., 2012. A cross-platform toolkit for mass spectrometry and proteomics. Nature Biotechnology 30, 918–920. https://doi.org/10.1038/nbt.2377

Chandler, J.A., Eisen, J.A., Kopp, A., 2012. Yeast Communities of Diverse

Drosophila Species : Comparison of Two Symbiont Groups in the Same Hosts 78, 7327–7336. https://doi.org/10.1128/AEM.01741-12
Chandler, J.A., Lang, J.M., Bhatnagar, S., Eisen, J.A., 2011. Bacterial Communities of Diverse Drosophila Species : Ecological Context of a Host – Microbe Model System 7. https://doi.org/10.1371/journal.pgen.1002272
Chaston, J.M., Newell, P.D., Douglas, A.E., 2014. Metagenome-wide association of microbial determinants of host phenotype in drosophila melanogaster. mBio 5, 1–12. https://doi.org/10.1128/mBio.01631-14
Consuegra, J., Grenier, T., Akherraz, H., Rahioui, I., Gervais, H., da Silva, P., Leulier, F., 2020a. Metabolic cooperation among commensal bacteria supports Drosophila juvenile growth under nutritional stress. iScience 101232. https://doi.org/10.1016/j.isci.2020.101232
Consuegra, J., Grenier, T., Baa-Puyoulet, P., Rahioui, I., Akherraz, H., Gervais, H., Parisot, N., Silva, P. da, Charles, H., Calevro, F., Leulier, F., 2020b. Drosophila-associated bacteria differentially shape the nutritional requirements of their host during juvenile growth. PLOS Biology 18, e3000681. https://doi.org/10.1371/journal.pbio.3000681
Coplan, J.D., Fathy, H.M., Abdallah, C.G., Ragab, S.A., Kral, J.G., Mao, X., Shungu, D.C., Mathew, S.J., 2014. Reduced hippocampal N-acetyl-aspartate (NAA) as a biomarker for overweight. NeuroImage: Clinical 4, 326–335. https://doi.org/10.1016/j.nicl.2013.12.014
Cox, J.E., Thummel, C.S., Tennessen, J.M., 2017. Metabolomic Studies in *Drosophila*. Genetics 206, 1169–1185. https://doi.org/10.1534/genetics.117.200014
Coyte, K.Z., Rakoff-Nahoum, S., 2019. Understanding Competition and Cooperation within the Mammalian Gut Microbiome. Current Biology 29, R538–R544. https://doi.org/10.1016/j.cub.2019.04.017
Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., Fahy, E., Steinbeck, C., Subramanian, S., Bolton, E., Greiner, R., Wishart, D.S., 2016. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. J Cheminform 8. https://doi.org/10.1186/s13321-016-0174-y
Dobson, A.J., Chaston, J.M., Douglas, A.E., 2016. The Drosophila transcriptional network is structured by microbiota. BMC Genomics 17, 975. https://doi.org/10.1186/s12864-016-3307-9
Douglas, A.E., 2019. Simple animal models for microbiome research. Nature Reviews Microbiology 17, 764–775. https://doi.org/10.1038/s41579-019-0242-1
Douglas, A.E., 2014. Symbiosis as a General Principle in Eukaryotic Evolution. Cold Spring Harb Perspect Biol 6, a016113. https://doi.org/10.1101/cshperspect.a016113
Dow, J. a. T., Davies, S.A., 2003. Integrative Physiology and Functional Genomics of Epithelial Function in a Genetic Model Organism. Physiological Reviews 83, 687–729. https://doi.org/10.1152/physrev.00035.2002
Efron, B., 2007. Size, power and false discovery rates. Ann. Statist. 35, 1351–1377. https://doi.org/10.1214/009053606000001460

Engel, P., Moran, N.A., 2013. The gut microbiota of insects - diversity in structure and function. FEMS Microbiology Reviews 37, 699–735. https://doi.org/10.1111/1574-6976.12025

Erkosar, B., Storelli, G., Defaye, A., Leulier, F., 2013. Host-intestinal microbiota mutualism: "learning on the fly." Cell Host and Microbe 13, 8–14. https://doi.org/10.1016/j.chom.2012.12.004

Figueiredo, A.R.T., Kramer, J., 2020. Cooperation and Conflict Within the Microbiota and Their Effects On Animal Hosts. Front. Ecol. Evol. 8, 132. https://doi.org/10.3389/fevo.2020.00132

Fischer, C., Trautman, E.P., Crawford, J.M., Stabb, E.V., Handelsman, J., Broderick, N.A., 2017. Metabolite exchange between microbiome members produces compounds that influence drosophila behavior. eLife 6. https://doi.org/10.7554/eLife.18855

Fox, J., Weisberg, S., 2019. An {R} Companion to Applied Regression, Third edit. ed. Thousand Oaks, CA.

Fromont, C., Adair, K.L., Douglas, A.E., 2019. Correlation and causation between the microbiome, Wolbachia and host functional traits in natural populations of drosophilid flies. Mol. Ecol. 28, 1826–1841. https://doi.org/10.1111/mec.15041

Gansemer, E.R., McCommis, K.S., Martino, M., King-McAlpin, A.Q., Potthoff, M.J., Finck, B.N., Taylor, E.B., Rutkowski, D.T., 2020. NADPH and Glutathione Redox Link TCA Cycle Activity to Endoplasmic Reticulum Homeostasis. iScience 23, 101116. https://doi.org/10.1016/j.isci.2020.101116

Gérard, C., Vidal, H., 2019. Impact of Gut Microbiota on Host Glycemic Control. Front. Endocrinol. 10. https://doi.org/10.3389/fendo.2019.00029

Goncalves, R.L.S., Rothschild, D.E., Quinlan, C.L., Scott, G.K., Benz, C.C., Brand, M.D., 2014. Sources of superoxide/H2O2 during mitochondrial proline oxidation. Redox Biology 2, 901–909. https://doi.org/10.1016/j.redox.2014.07.003

Gould, A.L., Zhang, V., Lamberti, L., Jones, E.W., Obadia, B., Korasidis, N., Gavryushkin, A., Carlson, J.M., Beerenwinkel, N., Ludington, W.B., 2018. Microbiome interactions shape host fitness. Proceedings of the National Academy of Sciences of the United States of America 115, E11951–E11960. https://doi.org/10.1073/pnas.1809349115

Granato, E.T., Meiller-Legrand, T.A., Foster, K.R., 2019. The Evolution and Ecology of Bacterial Warfare. Current Biology 29, R521–R537. https://doi.org/10.1016/j.cub.2019.04.024

Harvanek, Z.M., Lyu, Y., Gendron, C.M., Johnson, J.C., Kondo, S., Promislow, D.E.L., Pletcher, S.D., 2017. Perceptive costs of reproduction drive ageing and physiology in male Drosophila. Nature Ecology & Evolution 1, 1–8. https://doi.org/10.1038/s41559-017-0152

Hooper, L.V., Midtvedt, T., Gordon, J.I., 2002. How Host-Microbial Interactions Shape the Nutrient Environment of the Mammalian Intestine. Annual Review of Nutrition 22, 283–307. https://doi.org/10.1146/annurev.nutr.22.011602.092259

Huang, J., Douglas, A.E., 2015. Consumption of dietary sugar by gut bacteria determines Drosophila lipid content. Biology Letters 11, 20150469–20150469.
Jehrke, L., Stewart, F.A., Droste, A., Beller, M., 2018. The impact of genome variation and diet on the metabolic phenotype and microbiome composition of Drosophila melanogaster. Scientific Reports 8, 1–15. https://doi.org/10.1038/s41598-018-24542-5
Jensen, K., McClure, C., Priest, N.K., Hunt, J., 2015. Sex-specific effects of protein and carbohydrate intake on reproduction but not lifespan in Drosophila melanogaster. Aging Cell 14, 605–615. https://doi.org/10.1111/acel.12333
Jin, K., Wilson, K.A., Beck, J.N., Nelson, C.S., Iii, G.W.B., Harrison, B.R., Djukovic, D., Raftery, D., Brem, R.B., Yu, S., Drton, M., Shojaie, A., Kapahi, P., Promislow, D., 2020. Genetic and metabolomic architecture of variation in diet restriction-mediated lifespan extension in Drosophila. PLOS Genetics 16, e1008835. https://doi.org/10.1371/journal.pgen.1008835
Judd, A.M., Matthews, M.K., Hughes, R., Veloz, M., Sexton, C.E., Chaston, J.M., 2018. Bacterial Methionine Metabolism Genes Influence *Drosophila melanogaster* Starvation Resistance. Appl Environ Microbiol 84, e00662-18, /aem/84/17/e00662-18.atom. https://doi.org/10.1128/AEM.00662-18
Kanehisa, M., Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28, 27–30. https://doi.org/10.1093/nar/28.1.27
Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L., Gordon, J.I., 2011. Human nutrition, the gut microbiome and the immune system. Nature 474, 327–336. https://doi.org/10.1038/nature10213
Kaur, S., Birdsill, A.C., Steward, K., Pasha, E., Kruzliak, P., Tanaka, H., Haley, A.P., 2017. Higher visceral fat is associated with lower cerebral N-acetyl-aspartate ratios in middle-aged adults. Metab Brain Dis 32, 727–733. https://doi.org/10.1007/s11011-017-9961-z
Keebaugh, E.S., Yamada, R., Obadia, B., Ludington, W.B., Ja, W.W., 2018. Microbial Quantity Impacts Drosophila Nutrition, Development, and Lifespan. iScience 4, 247–259. https://doi.org/10.1016/j.isci.2018.06.004
Kešnerová, L., Mars, R.A.T., Ellegaard, K.M., Troilo, M., Sauer, U., Engel, P., 2017. Disentangling metabolic functions of bacteria in the honey bee gut. PLoS Biology 15, 1–28. https://doi.org/10.1371/journal.pbio.2003467
Kolde, R., 2019. pheatmap: Pretty Heatmaps. R package version 1.0.12.
Koyle, M.L., Veloz, M., Judd, A.M., Wong, A.C.-N., Newell, P.D., Douglas, A.E., Chaston, J.M., 2016. Rearing the Fruit Fly Drosophila melanogaster Under Axenic and Gnotobiotic Conditions. Journal of Visualized Experiments. https://doi.org/10.3791/54219
Lenth, R., 2019. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.
Li, H., Tennessen, J.M., 2017. Methods for studying the metabolic basis of Drosophila development. Wiley Interdiscip Rev Dev Biol 6. https://doi.org/10.1002/wdev.280
Mattila, J., Hietakangas, V., 2017. Regulation of Carbohydrate Energy Metabolism in Drosophila melanogaster. Genetics 207, 1231–1253.

https://doi.org/10.1534/genetics.117.199885
McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H.V., Domazet-Lošo, T., Douglas, A.E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S.F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A.H., Kremer, N., Mazmanian, S.K., Metcalf, J.L., Nealson, K., Pierce, N.E., Rawls, J.F., Reid, A., Ruby, E.G., Rumpho, M., Sanders, J.G., Tautz, D., 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, 3229–36. https://doi.org/10.1073/pnas.1218525110
McKean, K.A., Nunney, L., 2001. Increased sexual activity reduces male immune function in Drosophila melanogaster. PNAS 98, 7904–7909. https://doi.org/10.1073/pnas.131216398
McMullen, J.G., Peters-Schulze, G., Cai, J., Patterson, A.D., Douglas, A.E., 2020. How gut microbiome interactions affect nutritional traits of *Drosophila melanogaster*. J Exp Biol 223, jeb227843. https://doi.org/10.1242/jeb.227843
Morais, L.H., Schreiber, H.L., Mazmanian, S.K., 2020. The gut microbiota–brain axis in behaviour and brain disorders. Nature Reviews Microbiology 1–15. https://doi.org/10.1038/s41579-020-00460-0
Musselman, L.P., Kühnlein, R.P., 2018. Drosophila as a model to study obesity and metabolic disease. Journal of Experimental Biology 121. https://doi.org/10.1242/jeb.163881
Newell, P.D., Douglas, A.E., 2014. Interspecies Interactions Determine the Impact of the Gut Microbiota on Nutrient Allocation in Drosophila melanogaster. Applied and Environmental Microbiology 80, 788–796. https://doi.org/10.1128/AEM.02742-13
Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., Pettersson, S., 2012. Host-Gut Microbiota Metabolic Interactions. Science 336, 1262–1267. https://doi.org/10.1126/science.1223813
Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., Wagner, H., 2019. vegan: Community Ecology Package. R package version 2.5-5.
Quan, A.S., Eisen, M.B., 2018. The ecology of the drosophila-yeast mutualism in wineries. PLoS ONE 13, 1–25. https://doi.org/10.1371/journal.pone.0196440
R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Read, M.N., Holmes, A.J., 2017. Towards an integrative understanding of diet-host-gut microbiome interactions. Frontiers in Immunology 8, 1–9. https://doi.org/10.3389/fimmu.2017.00538
Ridley, E.V., Wong, A.C.N., Douglas, A.E., 2013. Microbe-dependent and nonspecific effects of procedures to eliminate the resident microbiota from drosophila melanogaster. Applied and Environmental Microbiology 79, 3209–3214. https://doi.org/10.1128/AEM.00206-13
Rolhion, N., Chassaing, B., 2016. When pathogenic bacteria meet the intestinal microbiota. Philosophical Transactions of the Royal Society B: Biological

Sciences 371. https://doi.org/10.1098/rstb.2015.0504
Roy, S., Saha, T.T., Zou, Z., Raikhel, A.S., 2018. Regulatory Pathways Controlling Female Insect Reproduction. Annual Review of Entomology 63, 489–511. https://doi.org/10.1146/annurev-ento-020117-043258
Ruggirello, M., Nucera, D., Cannoni, M., Peraino, A., Rosso, F., Fontana, M., Cocolin, L., Dolci, P., 2019. Antifungal activity of yeasts and lactic acid bacteria isolated from cocoa bean fermentations. Food Research International 115, 519–525. https://doi.org/10.1016/j.foodres.2018.10.002
Shin, S.C., Kim, S.-H., You, H., Kim, B., Kim, A.C., Lee, K.-A., Yoon, J.-H., Ryu, J.-H., Lee, W.-J., 2011. Drosophila microbiome modulates host developmental and metabolic homeostasis via insulin signaling. Science 334, 670–4. https://doi.org/10.1126/science.1212782
Sommer, A.J., Newell, P.D., 2019. Metabolic Basis for Mutualism between Gut Bacteria and Its Impact on the Drosophila melanogaster Host. Applied and environmental microbiology 85, 1–15. https://doi.org/10.1128/AEM.01882-18
Sommer, F., Bäckhed, F., 2013. The gut microbiota-masters of host development and physiology. Nature Reviews Microbiology 11, 227–238. https://doi.org/10.1038/nrmicro2974
St. Clair, S.L., Li, H., Ashraf, U., Karty, J.A., Tennessen, J.M., 2017. Metabolomic Analysis Reveals That the Drosophila melanogaster Gene lysine Influences Diverse Aspects of Metabolism. Genetics 207, 1255–1261. https://doi.org/10.1534/genetics.117.300201
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 Metabolism 14, 403–414. https://doi.org/10.1016/j.cmet.2011.07.012
Storelli, G., Strigini, M., Grenier, T., Bozonnet, L., Schwarzer, M., Daniel, C., Matos, R., Leulier, F., 2018. Drosophila Perpetuates Nutritional Mutualism by Promoting the Fitness of Its Intestinal Symbiont Lactobacillus plantarum. Cell Metab 27, 362-377.e8. https://doi.org/10.1016/j.cmet.2017.11.011
Svanström, Å., Boveri, S., Boström, E., Melin, P., 2013. The lactic acid bacteria metabolite phenyllactic acid inhibits both radial growth and sporulation of filamentous fungi. BMC Res Notes 6, 464. https://doi.org/10.1186/1756-0500-6-464
Terashima, J., Bownes, M., 2004. Translating available food into the number of eggs laid by Drosophila melanogaster. Genetics 167, 1711–1719. https://doi.org/10.1534/genetics.103.024323
Thaiss, C.A., Zmora, N., Levy, M., Elinav, E., 2016. The microbiome and innate immunity. Nature 535, 65–74. https://doi.org/10.1038/nature18847
Thévenot, E.A., Roux, A., Xu, Y., Ezan, E., Junot, C., 2015. Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. J. Proteome Res. 14, 3322–3335. https://doi.org/10.1021/acs.jproteome.5b00354
Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M.,

VanderGheynst, J., Fiehn, O., Arita, M., 2015. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nature Methods 12, 523–526. https://doi.org/10.1038/nmeth.3393

van den Berg, R.A., Hoefsloot, H.C., Westerhuis, J.A., Smilde, A.K., van der Werf, M.J., 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics 7, 142. https://doi.org/10.1186/1471-2164-7-142

Vernocchi, P., Del Chierico, F., Putignani, L., 2016. Gut Microbiota Profiling: Metabolomics Based Approach to Unravel Compounds Affecting Human Health. Front Microbiol 7. https://doi.org/10.3389/fmicb.2016.01144

Wilinski, D., Winzeler, J., Duren, W., Persons, J.L., Holme, K.J., Mosquera, J., Khabiri, M., Kinchen, J.M., Freddolino, P.L., Karnovsky, A., Dus, M., 2019. Rapid metabolic shifts occur during the transition between hunger and satiety in Drosophila melanogaster. Nature Communications 10, 4052. https://doi.org/10.1038/s41467-019-11933-z

Wohlgemuth, G., Haldiya, P.K., Willighagen, E., Kind, T., Fiehn, O., 2010. The Chemical Translation Service—a web-based tool to improve standardization of metabolomic reports. Bioinformatics 26, 2647–2648. https://doi.org/10.1093/bioinformatics/btq476

Wong, A.C.-N., Chaston, J.M., Douglas, A.E., 2013. The inconstant gut microbiota of Drosophila species revealed by 16S rRNA gene analysis. The ISME journal 7, 1922–32. https://doi.org/10.1038/ismej.2013.86

Wong, A.C.-N., Dobson, A.J., Douglas, A.E., 2014. Gut microbiota dictates the metabolic response of Drosophila to diet. The Journal of experimental biology 217, 1894–1901. https://doi.org/10.1242/jeb.101725

Wong, A.C.N., Vanhove, A.S., Watnick, P.I., 2016. The interplay between intestinal bacteria and host metabolism in health and disease: lessons from Drosophila melanogaster. Disease Models & Mechanisms 9. https://doi.org/10.1242/dmm.023408

Wu, Q., Zhang, H., Dong, X., Chen, X.-F., Zhu, Z.-Y., Hong, Z.-Y., Chai, Y.-F., 2014. UPLC-Q-TOF/MS based metabolomic profiling of serum and urine of hyperlipidemic rats induced by high fat diet. Journal of Pharmaceutical Analysis 4, 360–367. https://doi.org/10.1016/j.jpha.2014.04.002

Zhou, S., Morgante, F., Geisz, M., Ma, J., Anholt, R., Mackay, T., 2019. Systems Genetics of the Drosophila Metabolome. Genome Res. gr.243030.118. https://doi.org/10.1101/gr.243030.118

CHAPTER 5

DISCUSSION

Research conclusions

The goal of this dissertation was to investigate the effect of among-microbe interactions on *Drosophila* nutrient allocation and metabolic function. Results indicated that among the bacterial microbiome members, taxa of different orders differ metabolically which may facilitate niche overlap between taxa. In addition, co-associations between representative members of three key taxa, *Acetobacter fabarum* (Acetobacteraceae), *Lactobacillus brevis* (Lactobacillales) and yeast (*Hanseniaspora uvarum*), demonstrated that mono-associations cannot predict how the microbiome influences key aspects of the nutritional physiology and metabolome of *Drosophila*. This research contributes to our understanding of the relationship between animal hosts and the gut microbiome, including how metabolic functions of individual microbial taxa and metabolic interactions among microorganisms influence host traits. At the start of this research, there were few examples on the importance of among-microbe interactions on *Drosophila* traits, e.g. Newell and Douglas (2014). Most research on host-microbiome interactions in *Drosophila* had emphasized the functions of individual bacteria in host-microbiome associations, e.g. (Chaston et al., 2014; Erkosar et al., 2015; Shin et al., 2011; Storelli et al., 2011) with the assumption that individual traits would be additive in a community context, or focused on the combined effect multi-species communities with conventional (i.e. unmanipulated microbiome) fly, while some studies used a standard synthetic microbiome without

assessing underlying community processes (Chaston et al., 2016; Combe et al., 2014; Ridley et al., 2012; Wong et al., 2014). However, there have been several notable recent findings that highlight that pairwise and higher-order interactions between gut microorganisms can expand the metabolic capabilities of individual microbes and their influence on *Drosophila* health and fitness (Consuegra et al., 2020a; Fischer et al., 2017; Gould et al., 2018; Sommer and Newell, 2019), and these studies complement this dissertation work.

In the main section of this Discussion (below), I address the patterns identified in my research in the context of the literature on the *Drosophila* system, and the implications of this research for the ecology and evolution of *Drosophila* and its associated gut microbiome. I conclude with some perspectives for future research.

***Patterns of interactions between microorganisms and* Drosophila**

It has been known from the first studies on the *Drosophila* microbiome using culture-independent molecular methods that the taxonomic composition of the *Drosophila* gut microbiome is variable and dominated by bacterial taxa of three orders (Enterobacterales, Lactobacillales, and Rhodospirillales) (Chandler et al., 2011; Corby-Harris et al., 2007). The genome sequence analyses conducted in **Chapter 2** revealed that bacteria isolated from wild *Drosophila* are functionally distinct at various taxonomic levels from order to strain-level variation. At higher order taxonomy, the metabolic traits varied widely between the three dominant bacterial orders with a high number of traits related to fermentation, amino acid, and nucleotide metabolism in bacteria belonging to the Rhodospirillales and Lactobacillales. These

metabolic functions are candidate processes contributing to host traits that vary with gut microbiome composition. For example, there is evidence that conserved metabolic capabilities of Rhodospirillales and Lactobacillales can suppress or reverse locally adapted life history traits of *Drosophila* with faster or slower development time, respectively, and that the natural composition of the microbiome in wild *Drosophila* populations along a well-established latitudinal gradient along the U.S. East coast matches the latitudinal gradient of host life history traits (Walters et al., 2020).

Strain-level variation among prevalent bacteria associated with *Drosophila* indicated that some of the bacteria, particularly *Lactobacillus plantarum* and *Gluconobacter cerinus*, were enriched in diverse functions for utilization of carbon. These results provide a strong caution against the continued use of 16S rRNA gene-based data as a proxy for the study of functional variation in the gut microbiome of *Drosophila*. In addition, these findings may also explain the different and occasionally incompatible results obtained from different laboratories using different bacterial strains, e.g. Blum et al. (2013); Douglas (2018); Inamine et al. (2018); Matos and Leulier (2014). Looking ahead, they also provide a rationale for future research to extend the laboratory study of *L. plantarum* interactions with *Drosophila*, which currently is conducted on single strains (e.g. Newell and Douglas, 2014; Storelli et al., 2018, 2011; Téfit and Leulier, 2017), to multiple strains, to ensure that the range of results are representative of the species. Similarly, there is the opportunity to expand investigations of Acetobacteraceae to genera other than *Acetobacter* as we have a limited understanding of taxa outside of this genus, but see Ryu et al. (2008); Solomon et al. (2019).

Research to date on the *Drosophila* microbiome has focused primarily on *Acetobacter* and *Lactobacillus* spp., (e.g. Aranda-Díaz et al., 2020; Consuegra et al., 2020b; Dobson et al., 2015; Gould et al., 2018; Newell and Douglas, 2014; Shin et al., 2011; Storelli et al., 2011; Walters et al., 2020). In part, this is because these taxa are commonly found associated with laboratory *Drosophila* colonies. The genome data for representatives of the Enterobacterales can also provide the basis for further investigation of these taxa. Notably, the Enterobacterales analyzed in **Chapter 2** included a higher diversity of functional traits than for the Acetobacteraceae or Lactobacillales strains investigated. The distinctive features of the Enterobacterales studied were associated primarily with urea cycle and polyamine metabolism. Important issues for future research are whether and how this metabolism may be linked to functional aspects of the interaction between Enterobacterales and *Drosophila*. Some members of the Enterobacterales, particularly *Providencia* spp., can be pathogenic to *Drosophila* (Galac and Lazzaro, 2011) and there may be host filtering processes that limit their abundance in the gut (Wang et al., 2020). A further prevalent and abundant genus of the Enterobacterales found in multiple studies in the wild *Drosophila* gut microbiome is *Tatumella* (Adair et al., 2018; Kang and Douglas, 2020; Wong et al., 2013). *Tatumella* is a member of the Erwiniaceae, which also includes important pathogens of plants and *Drosophila*, as well as insect endosymbionts (Adeolu et al., 2016; Manzano-Marín et al., 2020; Troha et al., 2018). In summary, the data are currently fragmentary but offer a strong basis to investigate *Drosophila* interactions with Enterobacterales. These interactions may differ from *Drosophila* interactions with Acetobacteraceae and Lactobacillales, may vary among

taxa within the Enterobacterales, and may be context-dependent, varying with environmental circumstance and co-occurring microorganisms.

The study summarized in **Chapter 3** was founded on the known effects of *Acetobacter* and *Lactobacillus* spp. on *Drosophila* performance and nutritional physiology. For example, *Acetobacter* spp. tend to decrease lipid and glucose content and development time, while *Lactobacillus* spp. lower glucose content and increased starvation resistance of female flies on nutrient poor conditions (Newell and Douglas, 2014; Téfit and Leulier, 2017). In co-association, *Acetobacter* and *Lactobacillus* display cooperative metabolism, with exchange of *Lactobacillus*-derived lactic acid for *Acetobacter*-produced essential amino acids and B vitamins, which in turn results in synergistic reductions in both development time and lipid storage of *Drosophila*. These effects are likely mediated through some combination of insulin-like signaling by acetic acid from *Acetobacter* produced using *Lactobacillus*-derived ethanol and consumption of dietary glucose (Consuegra et al., 2020a; Shin et al., 2011; Sommer and Newell, 2019). However, apart from Fischer et al (2017), interactions between bacteria and yeasts have rarely been investigated.

A key feature of the combinatorial experimental design of the research conducted in **Chapter 3** is the inclusion of yeast and yeast-bacteria interactions in the analysis of microbiome effects on fly nutritional status. In particular, the interaction between *Acetobacter fabarum* and *Hanseniaspora uvarum* resulted in increased titers of acetic acid and decreased lipid storage, with evidence suggesting that an unknown yeast-derived factor suppressed the acetic acid effect and increased lipid storage (Fig. 3.6). The metabolic basis for this interaction is a high priority for future research. The data

in **Chapter 3** offers a clear indication that yeasts may be an active constituent of the microbiome, which has been an open question in the field despite the early recognition of their importance in the *Drosophila* life cycle (Broderick and Lemaitre, 2012).

Most of the early research conducted on *Drosophila*-yeast interactions has focused on yeast as a food source, providing protein, B vitamins and sterols (Broderick and Lemaitre, 2012; Sang, 1956) and as an attractant (via volatile fermentation products) for *Drosophila* to rotting fruits (Becher et al., 2012). Nevertheless, there are several indications in the recent literature that yeasts can influence host functions in ways that may be contingent on their metabolic activity. For example, yeast increased flux of amino acids to the fly to extend fly longevity, although this function can be replaced with a re-occurring high dose of heat killed yeast cells (Yamada et al., 2015). *Drosophila* performance was dramatically influence, in particular with decreased development time (Murgier et al., 2019; Solomon et al., 2019; Fig. 3.2). In addition, yeast differ in their effect from individual bacteria on *Drosophila* molecular functions and resemble transcriptional profiles of conventional flies (Elya et al., 2016). To date, evidence for metabolic interactions between yeast and bacteria in the context of *Drosophila* come from a single study (Fischer et al., 2017), which indicated that *Drosophila* prefer co-culture of *Acetobacter-Saccharomyces* for feeding and oviposition compared to individual cultures. This is primarily due to the utilization of yeast-derived ethanol by *Acetobacter* to produce acetic acid, acetate esters, and acetaldehyde derivatives. However, synthesis of these different studies is limited by their use of different yeast species.

Much of our knowledge on *Drosophila*-yeast interactions comes from investigations

with *Saccharomyces* c*erevisiae*, e.g. Becher et al. (2012; Christiaens et al. (2014; Ha et al. (2009, 2005; Min et al. (2008), but may not be representative of yeast associated with wild *Drosophila* populations. Generally S. *cerevisiae* does not colonize the fly gut in the wild (Chandler et al., 2012; Quan and Eisen, 2018), is not found to survive the gut environment well (except as a spore) (Coluccio et al., 2008; Hoang et al., 2015), and its effect on *Drosophila* traits can differ from other yeasts (Anagnostou et al., 2010; Hoang et al., 2015; Murgier et al., 2019; Palanca et al., 2013). In nature, yeast diversity is lower compared to bacterial gut microbiome with *Hanseniaspora* and *Pichia* (also known as *Issatchenkia* (Kurtzman et al., 2008)) spp. as the dominant taxa (Chandler et al., 2012). At least for the larval gut, some *H. uvarum* vegetative cells appear to survive passage, but cell viability has not been quantified (Solomon et al., 2019). Based on research with *S. cerevisiae*, spores are likely to be the dominant cell type to survive the harsh conditions of the gut, which appears to be important for dispersal, outbreeding, and overwintering protection (Coluccio et al., 2008; Stefanini et al., 2016); while vegetative cells are likely transmitted via the bodily surface (Christiaens et al., 2014). Further investigation is required to understand the role of yeasts in the *Drosophila* gut microbiome as the current information is fragmentary.

In summary, our current understanding of the metabolic basis of microbiome effects on the nutritional and fitness traits of *Drosophila* comprises two broad processes. *Acetobacter* strains can lower *Drosophila* lipid storage through the competition for dietary glucose (Chaston et al., 2014; Huang and Douglas, 2015), and microbial production of acetic acid stimulates insulin-like signaling to reduce development time and glucose and lipid content (Shin et al., 2011). The metabolomics presented in

Chapter 4 provides evidence that co-associations can influence the metabolic profile of *Drosophila* and particularly highlights that co-associations between *A. fabarum* and *L. brevis* promotes titers of TCA cycle intermediates, particularly in male flies. It can be reasoned from the work presented here that the high titers of acetic acid observed in di-association between *A. fabarum* and *H. uvarum* may be causal for elevated titers of TCA cycle intermediates. This hypothesis is testable, for example by analysis of metabolite profiles in flies administered acetic acid (using the experimental design adopted in Chapter 3).

The identification of phenyllactic acid in the *Drosophila* metabolome, likely produced by *L. brevis* and *H. uvarum* (Ruggirello et al., 2019; Svanström et al., 2013), suggests that this metabolite it is a biomarker for metabolic health with negative association with glucose content in both sexes and a positive correlation with male fly lipid content. Direct evidence of microbial production of this metabolite by members of the microbiome and its effect on *Drosophila* remain to be investigated. The low titer of phenyllactic acid in the tri-association, if shown by further work to be general, could be indicative of *Acetobacter* effects on the production of this compound. These further studies are potentially very important because phenyllactic acid has a known antimicrobial effect on filamentous fungi (Svanström et al., 2013), and could represent a novel role of the microbiome in protection against filamentous fungal pathogens and competitors.

Perspectives

The key general point to emerge from these studies is that the interactions between

Drosophila and its microbiome members are complex and context-dependent. Multiple processes have been identified, such as nutritional interactions, defensive functions, and modulation of host signaling, and microbial partners have been shown to have diverse effects that are strongly dependent on the presence and composition of other microorganisms. Many other factors, including *Drosophila* genotype, diet, age, temperature, endosymbiont status (*Wolbachia, Spiroplasma*), and among-yeast interactions, e.g. (Jehrke et al., 2018; Newton and Rice, 2020; Rohlfs and Kürschner, 2010; Yamauchi et al., 2020), were not considered in these studies, but undoubtedly compound the context-dependency of microbial effects on *Drosophila* traits. *Drosophila* is a superb system to address the complexity of animal-microbiome interactions, which would be difficult or ethically impossible with other animals, especially humans. For example, the genetic and genomic resources available for *Drosophila*, such as DGRP lines (*Drosophila* Genetic Reference Panel), RNA interference, and CRISPR (Heigwer et al., 2018; Mackay et al., 2012), allows for follow-up investigations to determine the host genetic basis for metabolic patterns identified. In addition, genome-scale metabolic modeling with host signaling networks (Consuegra et al., 2020b; Imam et al., 2015; Schönborn et al., 2019) can provide a useful basis for combining *in silico* predictions and empirical studies to identify novel and unexpected metabolic targets for future investigations. Furthermore, the flexibility to study *Drosophila* in the laboratory and field (Adair et al., 2018; Kang and Douglas, 2020; Rudman et al., 2019) allows for studies concerning questions on the ecology and evolution of animal-microbiomes. In conclusion, the *Drosophila* gut microbiome has a continuing important role to play in study of animal-microbiome interactions.

References

Adair, K.L., Bost, A., Bueno, E., Kaunisto, S., Kortet, R., Peters-Schulze, G., Martinson, V.G., Douglas, A.E., 2020. Host determinants of among-species variation in microbiome composition in drosophilid flies. The ISME Journal 14, 217–229. https://doi.org/10.1038/s41396-019-0532-7

Adair, K.L., Wilson, M., Bost, A., Douglas, A.E., 2018. Microbial community assembly in wild populations of the fruit fly Drosophila melanogaster. ISME Journal 12, 959–972. https://doi.org/10.1038/s41396-017-0020-x

Adeolu, M., Alnajar, S., Naushad, S., S. Gupta, R., 2016. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. International Journal of Systematic and Evolutionary Microbiology 66, 5575–5599. https://doi.org/10.1099/ijsem.0.001485

Anagnostou, C., Dorsch, M., Rohlfs, M., 2010. Influence of dietary yeasts on Drosophila melanogaster life-history traits. Entomologia Experimentalis et Applicata 136, 1–11. https://doi.org/10.1111/j.1570-7458.2010.00997.x

Aranda-Díaz, A., Obadia, B., Dodge, R., Thomsen, T., Hallberg, Z.F., Güvener, Z.T., Ludington, W.B., Huang, K.C., 2020. Bacterial interspecies interactions modulate pH-mediated antibiotic tolerance. eLife 9, e51493. https://doi.org/10.7554/eLife.51493

Basset, A., Khush, R.S., Braun, A., Gardan, L., Boccard, F., Hoffmann, J.A., Lemaitre, B., 2000. The phytopathogenic bacteria Erwinia carotovora infects Drosophila and activates an immune response. PNAS 97, 3376–3381. https://doi.org/10.1073/pnas.97.7.3376

Becher, P.G., Flick, G., Rozpedowska, E., Schmidt, A., Hagman, A., Lebreton, S., Larsson, M.C., Hansson, B.S., Piškur, J., Witzgall, P., Bengtsson, M., 2012. Yeast, not fruit volatiles mediate Drosophila melanogaster attraction, oviposition and development. Functional Ecology 26, 822–828. https://doi.org/10.1111/j.1365-2435.2012.02006.x

Becher, P.G., Hagman, A., Verschut, V., Chakraborty, A., Rozpędowska, E., Lebreton, S., Bengtsson, M., Flick, G., Witzgall, P., Piškur, J., 2018. Chemical signaling and insect attraction is a conserved trait in yeasts. Ecol Evol 8, 2962–2974. https://doi.org/10.1002/ece3.3905

Blum, J.E., Fischer, C.N., Miles, J., Handelsman, J., 2013. Frequent replenishment sustains the beneficial microbiome of Drosophila melanogaster. mBio 4. https://doi.org/10.1128/mBio.00860-13

Bost, A., Franzenburg, S., Adair, K.L., Martinson, V.G., Loeb, G., Douglas, A.E., 2018. How gut transcriptional function of Drosophila melanogaster varies with the presence and composition of the gut microbiota. Molecular Ecology 27, 1848–1859. https://doi.org/10.1111/mec.14413

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

Chandler, J.A., Eisen, J.A., Kopp, A., 2012. Yeast Communities of Diverse Drosophila Species : Comparison of Two Symbiont Groups in the Same Hosts 78, 7327–7336. https://doi.org/10.1128/AEM.01741-12

Chandler, J.A., Lang, J.M., Bhatnagar, S., Eisen, J.A., 2011. Bacterial Communities of Diverse Drosophila Species : Ecological Context of a Host – Microbe Model System 7. https://doi.org/10.1371/journal.pgen.1002272

Chaston, J.M., Dobson, A.J., Newell, P.D., Douglas, A.E., 2016. Host genetic control of the microbiota mediates the Drosophila Nutritional Phenotype. Applied and Environmental Microbiology 82, 671–679. https://doi.org/10.1128/AEM.03301-15.Editor

Chaston, J.M., Newell, P.D., Douglas, A.E., 2014. Metagenome-wide association of microbial determinants of host phenotype in drosophila melanogaster. mBio 5, 1–12. https://doi.org/10.1128/mBio.01631-14

Christiaens, J.F., Franco, L.M., Cools, T.L., De Meester, L., Michiels, J., Wenseleers, T., Hassan, B. a, Yaksi, E., Verstrepen, K.J., 2014. The fungal aroma gene ATF1 promotes dispersal of yeast cells through insect vectors. Cell reports 9, 425–32. https://doi.org/10.1016/j.celrep.2014.09.009

Coluccio, A.E., Rodriguez, R.K., Kernan, M.J., Neiman, A.M., 2008. The Yeast Spore Wall Enables Spores to Survive Passage through the Digestive Tract of Drosophila. PLoS One 3. https://doi.org/10.1371/journal.pone.0002873

Combe, B.E., Defaye, A., Bozonnet, N., Puthier, D., Royet, J., Leulier, F., 2014. Drosophila microbiota modulates host metabolic gene expression via IMD/NF-kB signaling. PloS one 9, e94729–e94729. https://doi.org/10.1371/journal.pone.0094729

Connor, R.C., 1995. The Benefits of Mutualism: A Conceptual Framework. Biological Reviews 70, 427–457. https://doi.org/10.1111/j.1469-185X.1995.tb01196.x

Consuegra, J., Grenier, T., Akherraz, H., Rahioui, I., Gervais, H., da Silva, P., Leulier, F., 2020a. Metabolic Cooperation among Commensal Bacteria Supports Drosophila Juvenile Growth under Nutritional Stress. iScience 23, 101232. https://doi.org/10.1016/j.isci.2020.101232

Consuegra, J., Grenier, T., Baa-Puyoulet, P., Rahioui, I., Akherraz, H., Gervais, H., Parisot, N., Silva, P. da, Charles, H., Calevro, F., Leulier, F., 2020b. Drosophila-associated bacteria differentially shape the nutritional requirements of their host during juvenile growth. PLOS Biology 18, e3000681. https://doi.org/10.1371/journal.pbio.3000681

Dobson, A.J., Chaston, J.M., Newell, P.D., Donahue, L., Hermann, S.L., Sannino, D.R., Westmiller, S., Wong, A.C.-N., Clark, A.G., Lazzaro, B.P., Douglas, A.E., 2015. Host genetic determinants of microbiota-dependent nutrition revealed by genome-wide analysis of Drosophila melanogaster. Nature Communications 6, 6312–6312. https://doi.org/10.1038/ncomms7312

Douglas, A.E., 2018. Contradictory Results in Microbiome Science Exemplified by Recent Drosophila Research. mBio 9. https://doi.org/10.1128/mBio.01758-18

Duar, R.M., Lin, X.B., Zheng, J., Martino, M.E., Grenier, T., Pérez-Muñoz, M.E., Leulier, F., Gänzle, M., Walter, J., 2017. Lifestyles in transition: evolution and natural history of the genus Lactobacillus. FEMS microbiology reviews 41,

S27–S48. https://doi.org/10.1093/femsre/fux030
Elya, C., Zhang, V., Ludington, W.B., Eisen, M.B., 2016. Stable Host Gene Expression in the Gut of Adult Drosophila melanogaster with Different Bacterial Mono-Associations. PLOS ONE 11, e0167357. https://doi.org/10.1371/journal.pone.0167357
Erkosar, B., Storelli, G., Mitchell, M., Bozonnet, L., Bozonnet, N., Leulier, F., 2015. Pathogen Virulence Impedes Mutualist-Mediated Enhancement of Host Juvenile Growth via Inhibition of Protein Digestion. Cell Host and Microbe 18. https://doi.org/10.1016/j.chom.2015.09.001
Fischer, C., Trautman, E.P., Crawford, J.M., Stabb, E.V., Handelsman, J., Broderick, N.A., 2017. Metabolite exchange between microbiome members produces compounds that influence drosophila behavior. eLife 6. https://doi.org/10.7554/eLife.18855
Galac, M.R., Lazzaro, B.P., 2011. Comparative pathology of bacteria in the genus Providencia to a natural host, Drosophila melanogaster. Microbes and Infection 13, 673–683. https://doi.org/10.1016/j.micinf.2011.02.005
Gilbert, D.G., 1980. Dispersal of yeasts and bacteria by Drosophila in a temperate forest 46, 135–137.
Gould, A.L., Zhang, V., Lamberti, L., Jones, E.W., Obadia, B., Korasidis, N., Gavryushkin, A., Carlson, J.M., Beerenwinkel, N., Ludington, W.B., 2018. Microbiome interactions shape host fitness. Proceedings of the National Academy of Sciences of the United States of America 115, E11951–E11960. https://doi.org/10.1073/pnas.1809349115
Ha, E.-M., Lee, K.-A., Park, S.H., Kim, S.-H., Nam, H.-J., Lee, H.-Y., Kang, D., Lee, W.-J., 2009. Regulation of DUOX by the Gαq-Phospholipase Cβ-Ca2+ Pathway in Drosophila Gut Immunity. Developmental Cell 16, 386–397. https://doi.org/10.1016/j.devcel.2008.12.015
Ha, E.-M., Oh, C.-T., Ryu, J.-H., Bae, Y.-S., Kang, S.-W., Jang, I., Brey, P.T., Lee, W.-J., 2005. An Antioxidant System Required for Host Protection against Gut Infection in Drosophila. Developmental Cell 8, 125–132. https://doi.org/10.1016/j.devcel.2004.11.007
Heigwer, F., Port, F., Boutros, M., 2018. RNA Interference (RNAi) Screening in Drosophila. Genetics 208, 853–874. https://doi.org/10.1534/genetics.117.300077
Hoang, D., Kopp, A., Chandler, J.A., 2015. Interactions between Drosophila and its natural yeast symbionts—Is Saccharomyces cerevisiae a good model for studying the fly-yeast relationship? PeerJ 3, e1116–e1116. https://doi.org/10.7717/peerj.1116
Huang, J., Douglas, A.E., 2015. Consumption of dietary sugar by gut bacteria determines Drosophila lipid content. Biology Letters 11, 20150469–20150469.
Imam, S., Schäuble, S., Brooks, A.N., Baliga, N.S., Price, N.D., 2015. Data-driven integration of genome-scale regulatory and metabolic network models. Front. Microbiol. 6. https://doi.org/10.3389/fmicb.2015.00409
Inamine, H., Ellner, S.P., Newell, P.D., Luo, Y., Buchon, N., Douglas, A.E., 2018. Spatiotemporally Heterogeneous Population Dynamics of Gut Bacteria

Inferred from Fecal Time Series Data. mBio 9. https://doi.org/10.1128/mBio.01453-17
Jehrke, L., Stewart, F.A., Droste, A., Beller, M., 2018. The impact of genome variation and diet on the metabolic phenotype and microbiome composition of Drosophila melanogaster. Scientific Reports 8, 1–15. https://doi.org/10.1038/s41598-018-24542-5
Jolly, N.P., Varela, C., Pretorius, I.S., 2014. Not your ordinary yeast: Non-Saccharomyces yeasts in wine production uncovered. FEMS Yeast Research 14, 215–237. https://doi.org/10.1111/1567-1364.12111
Joseph, R.M., Devineni, A.V., King, I.F.G., Heberlein, U., 2009. Oviposition preference for and positional avoidance of acetic acid provide a model for competing behavioral drives in Drosophila. Proceedings of the National Academy of Sciences of the United States of America 106, 11352–11357. https://doi.org/10.1073/pnas.0901419106
Kamareddine, L., Robins, W.P., Berkey, C.D., Mekalanos, J.J., Watnick, P.I., 2018. The Drosophila Immune Deficiency Pathway Modulates Enteroendocrine Function and Host Metabolism. Cell Metabolism 28, 449-462.e5. https://doi.org/10.1016/j.cmet.2018.05.026
Kang, D., Douglas, A.E., 2020. Functional traits of the gut microbiome correlated with host lipid content in a natural population of *Drosophila melanogaster*. Biol. Lett. 16, 20190803. https://doi.org/10.1098/rsbl.2019.0803
Keesey, I.W., Koerte, S., Khallaf, M.A., Retzke, T., Guillou, A., Grosse-Wilde, E., Buchon, N., Knaden, M., Hansson, B.S., 2017. Pathogenic bacteria enhance dispersal through alteration of Drosophila social communication. Nature Communications 8, 265. https://doi.org/10.1038/s41467-017-00334-9
Kim, G., Huang, J.H., McMullen, J.G., Newell, P.D., Douglas, A.E., 2018. Physiological responses of insects to microbial fermentation products: Insights from the interactions between Drosophila and acetic acid. Journal of Insect Physiology 106. https://doi.org/10.1016/j.jinsphys.2017.05.005
Kurtzman, C.P., Robnett, C.J., Basehoar-Powers, E., 2008. Phylogenetic relationships among species of Pichia, Issatchenkia and Williopsis determined from multigene sequence analysis, and the proposal of Barnettozyma gen. nov., Lindnera gen. nov. and Wickerhamomyces gen. nov. FEMS Yeast Research 8, 939–954. https://doi.org/10.1111/j.1567-1364.2008.00419.x
Mackay, T.F.C., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., Casillas, S., Han, Y., Magwire, M.M., Cridland, J.M., Richardson, M.F., Anholt, R.R.H., Barrón, M., Bess, C., Blankenburg, K.P., Carbone, M.A., Castellano, D., Chaboub, L., Duncan, L., Harris, Z., Javaid, M., Jayaseelan, J.C., Jhangiani, S.N., Jordan, K.W., Lara, F., Lawrence, F., Lee, S.L., Librado, P., Linheiro, R.S., Lyman, R.F., Mackey, A.J., Munidasa, M., Muzny, D.M., Nazareth, L., Newsham, I., Perales, L., Pu, L.-L., Qu, C., Ràmia, M., Reid, J.G., Rollmann, S.M., Rozas, J., Saada, N., Turlapati, L., Worley, K.C., Wu, Y.-Q., Yamamoto, A., Zhu, Y., Bergman, C.M., Thornton, K.R., Mittelman, D., Gibbs, R.A., 2012. The Drosophila melanogaster Genetic Reference Panel. Nature 482, 173–178. https://doi.org/10.1038/nature10811

Manzano-Marín, A., Coeur D'acier, A., Clamens, A.-L., Orvain, C., Cruaud, C., Barbe, V., Jousselin, E., 2020. Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of new nutritional symbionts in aphids' di-symbiotic systems. The ISME Journal 14, 259–273. https://doi.org/10.1038/s41396-019-0533-6

Matos, R.C., Leulier, F., 2014. Lactobacilli-Host mutualism: "learning on the fly." Microbial Cell Factories 13, S6–S6. https://doi.org/10.1186/1475-2859-13-S1-S6

Min, K.-J., Yamamoto, R., Buch, S., Pankratz, M., Tatar, M., 2008. Drosophila lifespan control by dietary restriction independent of insulin-like signaling. Aging Cell 7, 199–206. https://doi.org/10.1111/j.1474-9726.2008.00373.x

Murgier, J., Everaerts, C., Farine, J.P., Ferveur, J.F., 2019. Live yeast in juvenile diet induces species-specific effects on Drosophila adult behaviour and fitness. Scientific Reports 9, 1–12. https://doi.org/10.1038/s41598-019-45140-z

Newell, P.D., Douglas, A.E., 2014. Interspecies Interactions Determine the Impact of the Gut Microbiota on Nutrient Allocation in Drosophila melanogaster. Applied and Environmental Microbiology 80, 788–796. https://doi.org/10.1128/AEM.02742-13

Newton, I.L.G., Rice, D.W., 2020. The Jekyll and Hyde Symbiont: Could Wolbachia Be a Nutritional Mutualist? Journal of Bacteriology 202. https://doi.org/10.1128/JB.00589-19

Obadia, B., Güvener, Z.T., Zhang, V., Ceja-Navarro, J.A., Brodie, E.L., Ja, W.W., Ludington, W.B., 2017. Probabilistic Invasion Underlies Natural Gut Microbiome Stability. Current Biology 27, 1999-2006.e8. https://doi.org/10.1016/j.cub.2017.05.034

Pais, I.S., Valente, R.S., Sporniak, M., Teixeira, L., 2018. Drosophila melanogaster establishes a species-specific mutualistic interaction with stable gut-colonizing bacteria. PLOS Biology 16, e2005710. https://doi.org/10.1371/journal.pbio.2005710

Palanca, L., Gaskett, A.C., Günther, C.S., Newcomb, R.D., Goddard, M.R., 2013. Quantifying Variation in the Ability of Yeasts to Attract Drosophila melanogaster. PLoS ONE 8, e75332–e75332. https://doi.org/10.1371/journal.pone.0075332

Quan, A.S., Eisen, M.B., 2018. The ecology of the drosophila-yeast mutualism in wineries. PLoS ONE 13, 1–25. https://doi.org/10.1371/journal.pone.0196440

Reuter, M., Bell, G., Greig, D., 2007. Increased outbreeding in yeast in response to dispersal by an insect vector. Current Biology 17, R81–R83. https://doi.org/10.1016/j.cub.2006.11.059

Ridley, E.V., Wong, A.C.N., Westmiller, S., Douglas, A.E., 2012. Impact of the resident microbiota on the nutritional phenotype of drosophila melanogaster. PLoS ONE 7. https://doi.org/10.1371/journal.pone.0036765

Rohlfs, M., Kürschner, L., 2010. Saprophagous insect larvae, Drosophila melanogaster, profit from increased species richness in beneficial microbes. Journal of Applied Entomology 134, 667–671. https://doi.org/10.1111/j.1439-0418.2009.01458.x

Rudman, S.M., Greenblum, S., Hughes, R.C., Rajpurohit, S., Kiratli, O., Lowder, D.B., Lemmon, S.G., Petrov, D.A., Chaston, J.M., Schmidt, P., 2019. Microbiome composition shapes rapid genomic adaptation of Drosophila melanogaster. PNAS 116, 20025–20032. https://doi.org/10.1073/pnas.1907787116

Ruggirello, M., Nucera, D., Cannoni, M., Peraino, A., Rosso, F., Fontana, M., Cocolin, L., Dolci, P., 2019. Antifungal activity of yeasts and lactic acid bacteria isolated from cocoa bean fermentations. Food Research International 115, 519–525. https://doi.org/10.1016/j.foodres.2018.10.002

Ryu, J.-H., Kim, S.-H., Lee, H.-Y., Bai, J.Y., Nam, Y.-D., Bae, J.-W., Lee, D.G., Shin, S.C., Ha, E.-M., Lee, W.-J., 2008. Innate Immune Homeostasis by the Homeobox Gene Caudal and Commensal-Gut Mutualism in Drosophila. Science 319, 777–782. https://doi.org/10.1126/science.1149357

Sang, J.H., 1956. The Quantitative Nutritional Requirements of Drosophila Melanogaster. Journal of Experimental Biology 33, 45–72.

Schönborn, J.W., Jehrke, L., Mettler-Altmann, T., Beller, M., 2019. FlySilico: Flux balance modeling of Drosophila larval growth and resource allocation. Scientific Reports 9, 17156. https://doi.org/10.1038/s41598-019-53532-4

Shin, S.C., Kim, S.-H., You, H., Kim, B., Kim, a. C., Lee, K. -a., Yoon, J.-H., Ryu, J.-H., 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

Solomon, G.M., Dodangoda, H., McCarthy-Walker, T.T., Ntim-Gyakari, R.R., Newell, P.D., 2019. The microbiota of Drosophila suzukii influences the larval development of Drosophila melanogaster. PeerJ 2019. https://doi.org/10.7717/peerj.8097

Sommer, A.J., Newell, P.D., 2019. Metabolic Basis for Mutualism between Gut Bacteria and Its Impact on the Drosophila melanogaster Host. Applied and environmental microbiology 85, 1–15. https://doi.org/10.1128/AEM.01882-18

Stefanini, I., Dapporto, L., Berná, L., Polsinelli, M., Turillazzi, S., Cavalieri, D., 2016. Social wasps are a *Saccharomyces* mating nest. Proc Natl Acad Sci USA 113, 2247–2251. https://doi.org/10.1073/pnas.1516453113

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 Metabolism 14, 403–414. https://doi.org/10.1016/j.cmet.2011.07.012

Storelli, G., Strigini, M., Grenier, T., Bozonnet, L., Schwarzer, M., Daniel, C., Matos, R., Leulier, F., 2018. Drosophila Perpetuates Nutritional Mutualism by Promoting the Fitness of Its Intestinal Symbiont Lactobacillus plantarum. Cell Metab 27, 362-377.e8. https://doi.org/10.1016/j.cmet.2017.11.011

Stoy, K.S., Gibson, A.K., Gerardo, N.M., Morran, L.T., 2020. A need to consider the evolutionary genetics of host–symbiont mutualisms. Journal of Evolutionary Biology n/a. https://doi.org/10.1111/jeb.13715

Svanström, Å., Boveri, S., Boström, E., Melin, P., 2013. The lactic acid bacteria metabolite phenyllactic acid inhibits both radial growth and sporulation of

filamentous fungi. BMC Res Notes 6, 464. https://doi.org/10.1186/1756-0500-6-464
Téfit, M.A., Leulier, F., 2017. Lactobacillus plantarum favors the early emergence of fit and fertile adult Drosophila upon chronic undernutrition. Journal of Experimental Biology 220, 900–907. https://doi.org/10.1242/jeb.151522
Troha, K., Im, J.H., Revah, J., Lazzaro, B.P., Buchon, N., 2018. Comparative transcriptomics reveals CrebA as a novel regulator of infection tolerance in D. melanogaster. PLoS Pathog 14. https://doi.org/10.1371/journal.ppat.1006847
Walters, A.W., Hughes, R.C., Call, T.B., Walker, C.J., Wilcox, H., Petersen, S.C., Rudman, S.M., Newell, P.D., Douglas, A.E., Schmidt, P.S., Chaston, J.M., 2020. The microbiota influences the Drosophila melanogaster life history strategy. Molecular Ecology 29, 639–653. https://doi.org/10.1111/mec.15344
Wang, Y., Kapun, M., Waidele, L., Kuenzel, S., Bergland, A.O., Staubach, F., 2020. Common structuring principles of the Drosophila melanogaster microbiome on a continental scale and between host and substrate. Environmental Microbiology Reports 12, 220–228. https://doi.org/10.1111/1758-2229.12826
Winans, N.J., Walter, A., Chouaia, B., Chaston, J.M., Douglas, A.E., Newell, P.D., 2017. A genomic investigation of ecological differentiation between free-living and Drosophila-associated bacteria. Mol Ecol 26, 4536–4550. https://doi.org/10.1111/mec.14232
Wolfe, A.J., 2005. The Acetate Switch The Acetate Switch 69, 12–50. https://doi.org/10.1128/MMBR.69.1.12
Wong, A.C.-N., Chaston, J.M., Douglas, A.E., 2013. The inconstant gut microbiota of Drosophila species revealed by 16S rRNA gene analysis. The ISME journal 7, 1922–32. https://doi.org/10.1038/ismej.2013.86
Wong, A.C.-N., Dobson, A.J., Douglas, A.E., 2014. Gut microbiota dictates the metabolic response of Drosophila to diet. The Journal of experimental biology 217, 1894–1901. https://doi.org/10.1242/jeb.101725
Yamada, R., Deshpande, S.A., Bruce, K.D., Mak, E.M., Ja, W.W., 2015. Microbes Promote Amino Acid Harvest to Rescue Undernutrition in Drosophila. Cell reports 10, 1–8. https://doi.org/10.1016/j.celrep.2015.01.018
Yamauchi, T., Oi, A., Kosakamoto, H., Akuzawa-Tokita, Y., Murakami, T., Mori, H., Miura, M., Obata, F., 2020. Gut Bacterial Species Distinctively Impact Host Purine Metabolites during Aging in Drosophila. iScience 101477. https://doi.org/10.1016/j.isci.2020.101477

APPENDIX A

SUPPLEMENTAL MATERIAL FOR CHAPTER 2

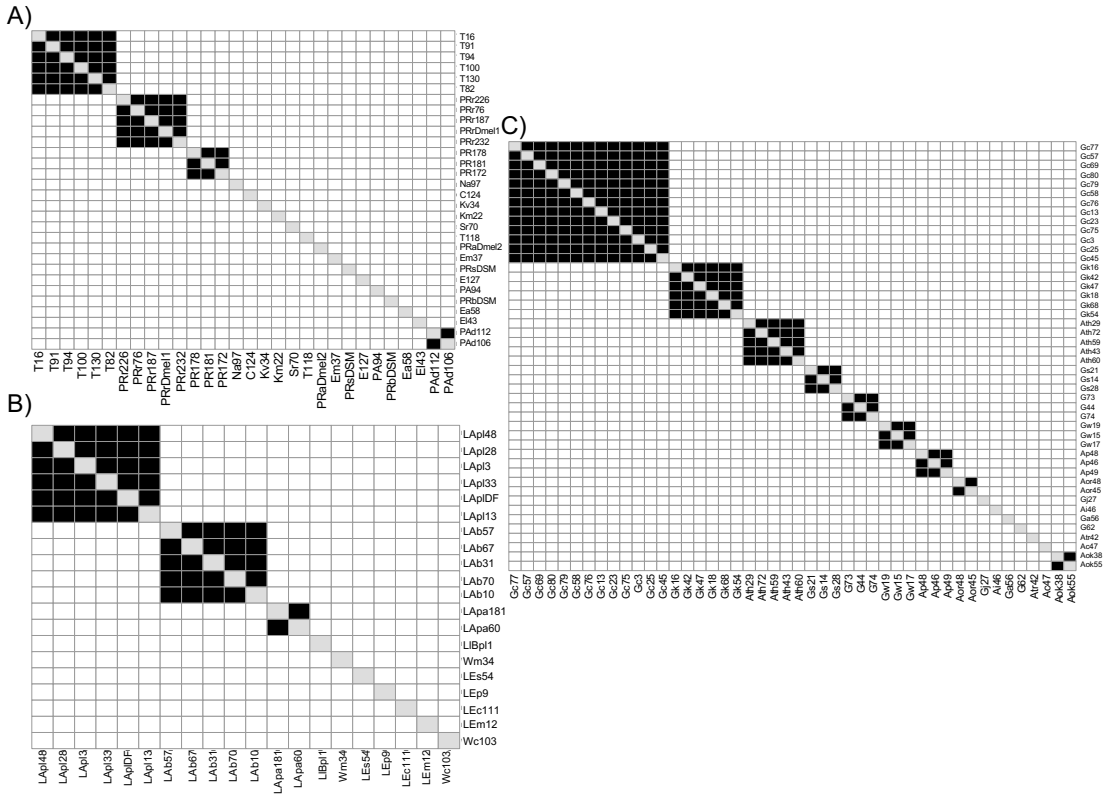


Figure S1.1. Species boundary delineation among *Drosophila*-associated strains. (A) Enterobacterales. (B) Lactobacillales. (C) Rhodospirillales. Black cells of heatmap correspond to pairwise average nucleotide identity (ANI) score of 95% or greater, while white cells indicate scores < 95%.

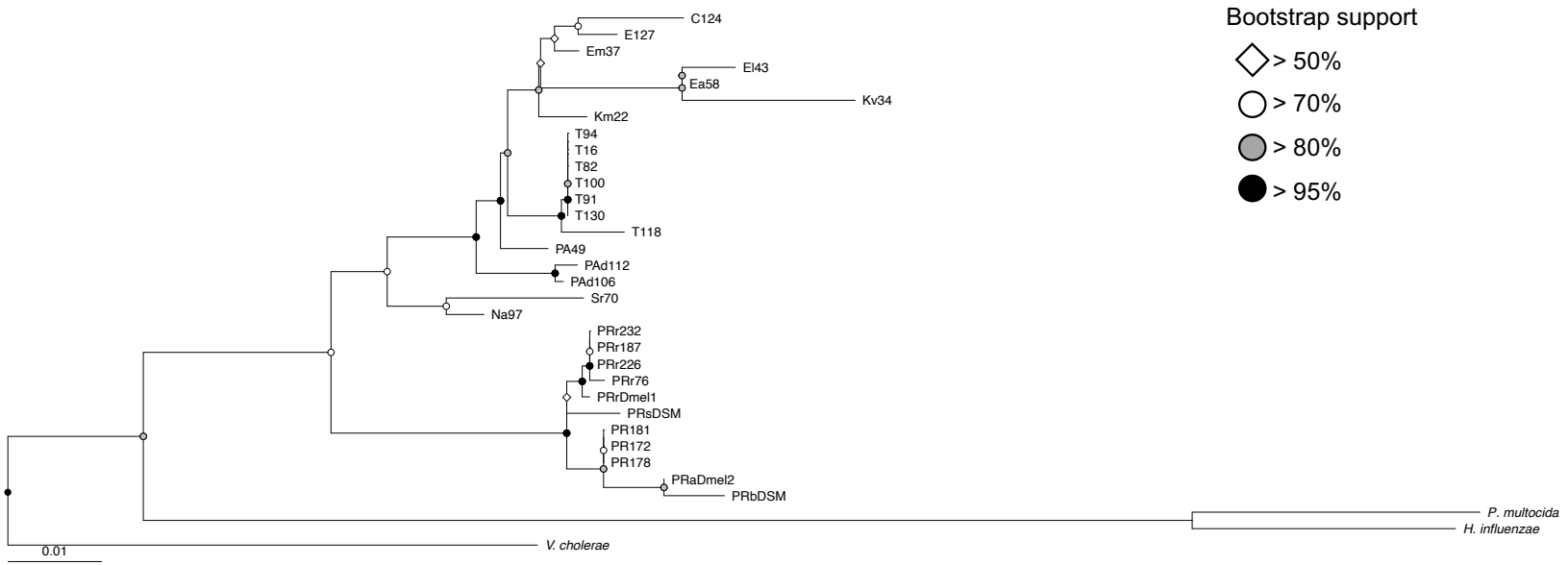


Figure S1.2. 16S rRNA gene phylogeny. (A) Enterobacterales clade extraction. *V. cholerae*, *H. influenzae*, and *P. multocida* are used as reference strains for studied taxa. Maximum likelihood tree was generated with a TIM3+F+I+G4 model of evolution (length = 1441 bp) with 10,000 ultrafast bootstrap replicates. Tree is scaled by number of nucleotide substitutions per site.

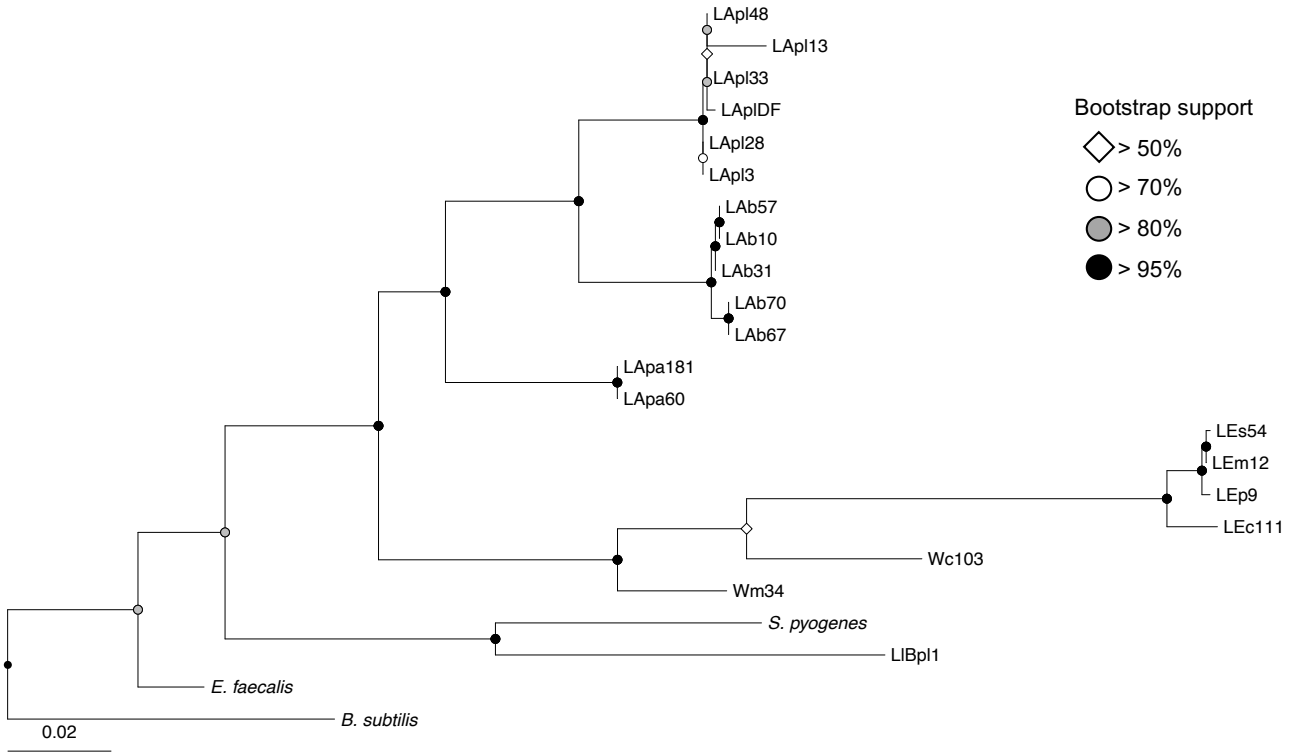


Figure S1.2. 16S rRNA gene phylogeny. (B) Lactobacillales clade extraction. *B. subtilis*, *E. faecalis*, and *S. pyogenes* are used as reference strains for studied taxa. Maximum likelihood tree was generated with a TIM3+F+I+G4 model of evolution (length = 1441 bp) with 10,000 ultrafast bootstrap replicates. Tree is scaled by number of nucleotide substitutions per site.

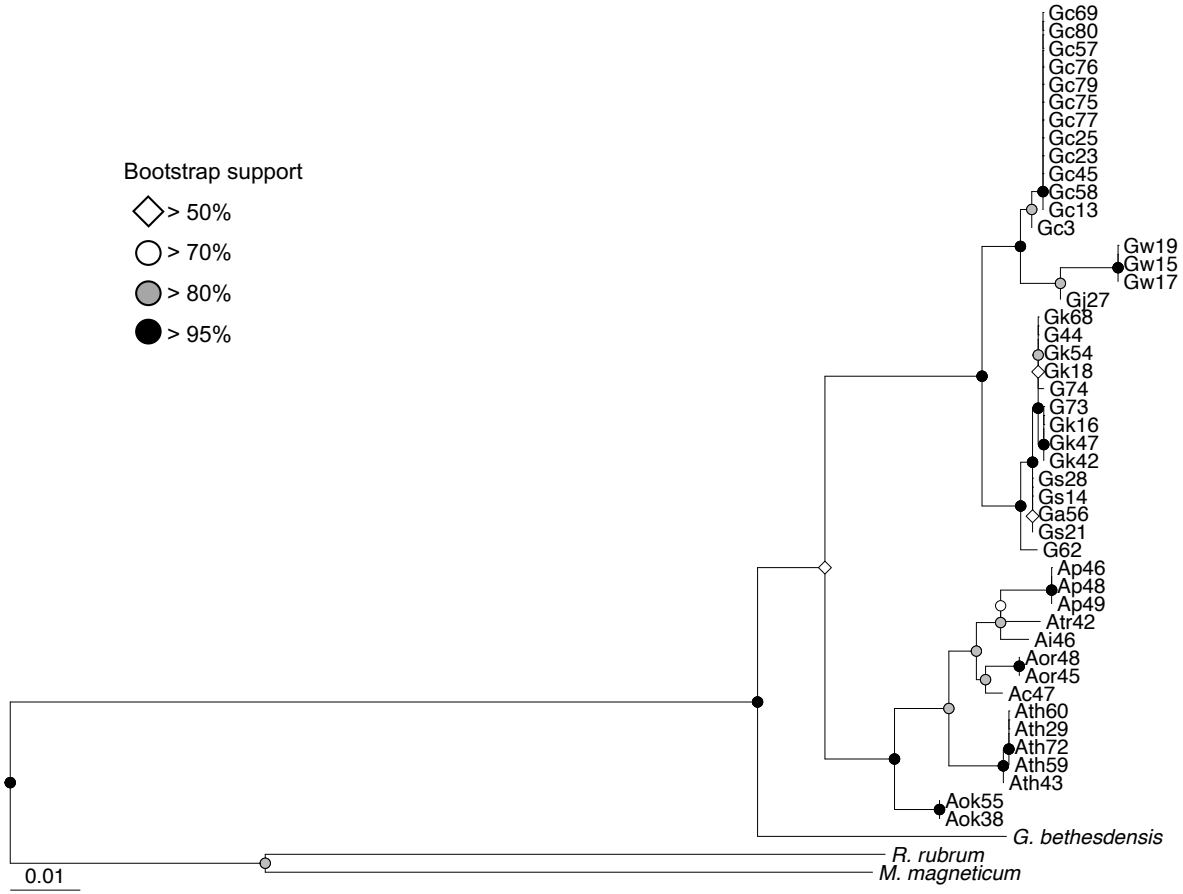


Figure S1.2. 16S rRNA gene phylogeny. (C) Rhodospirillales clade extraction. *M. magneticum*, *R. rubrum*, and *G. bethesdensis* are used as reference strains for studied taxa. Maximum likelihood tree was generated with a TIM3+F+I+G4 model of evolution (length = 1441 bp) with 10,000 ultrafast bootstrap replicates. Tree is scaled by number of nucleotide substitutions per site.

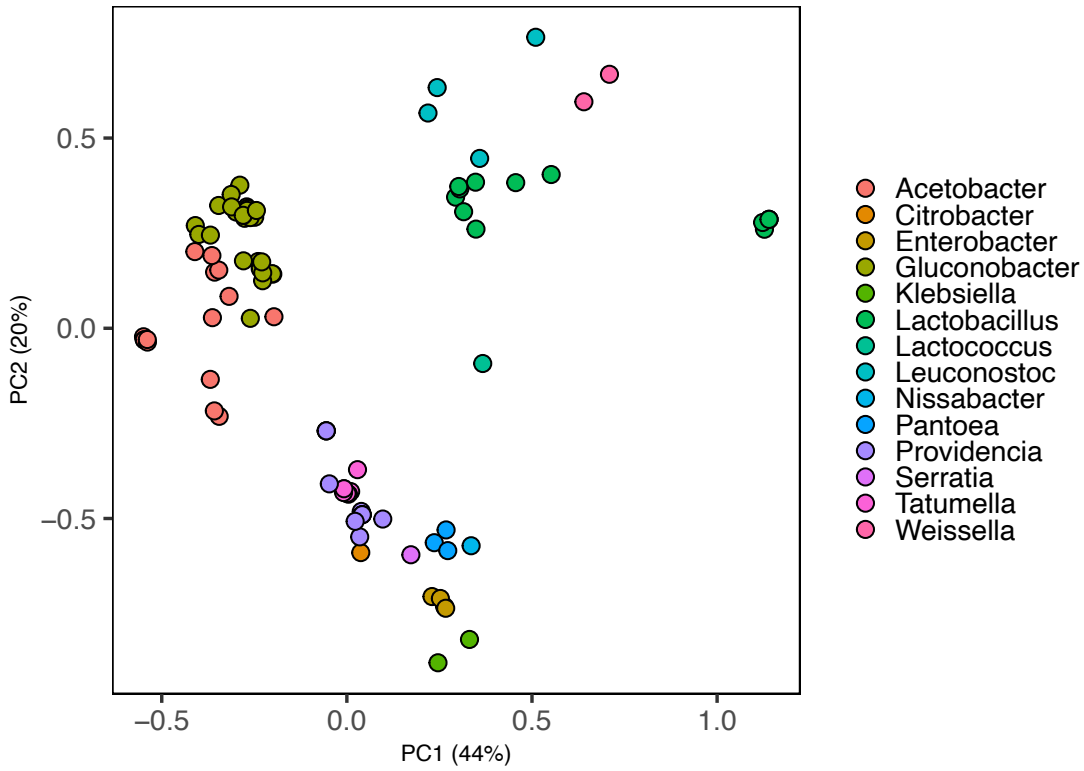


Figure S1.3. Principal coordinates analysis (PCoA) of metabolic functions by genus. All points represent composite of metabolic functional traits for each genome analyzed with colors corresponding to genus-level taxonomy. The percentages on each axis correspond to the amount of variation explained.

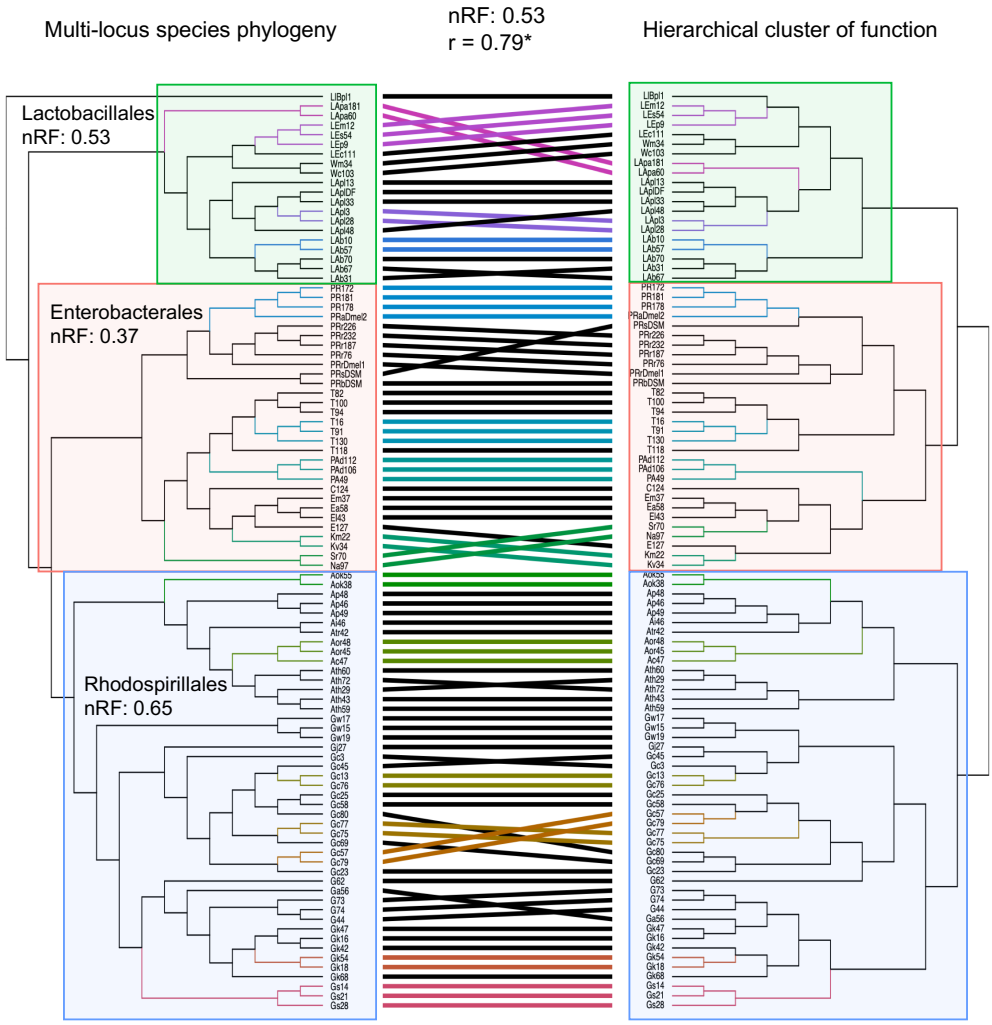


Figure S1.4. Mapping function onto phylogeny. (A) Correlation between multi-locus species phylogeny and Bray-Curtis dissimilarity based hierarchical cluster of relative metabolic function counts. The normalized Robinson-Foulds index (nRF) and Mantel's test (based on Spearman's rank correlation) for the entire tanglegram are reported at the top. Each bacterial order is indicated by shaded box with the specific nRF obtained for each order. The tips of each dendrogram are connected by either black or colored line; the latter represents subtrees within each dendrogram that have the same topology. * represents p = 0.001.

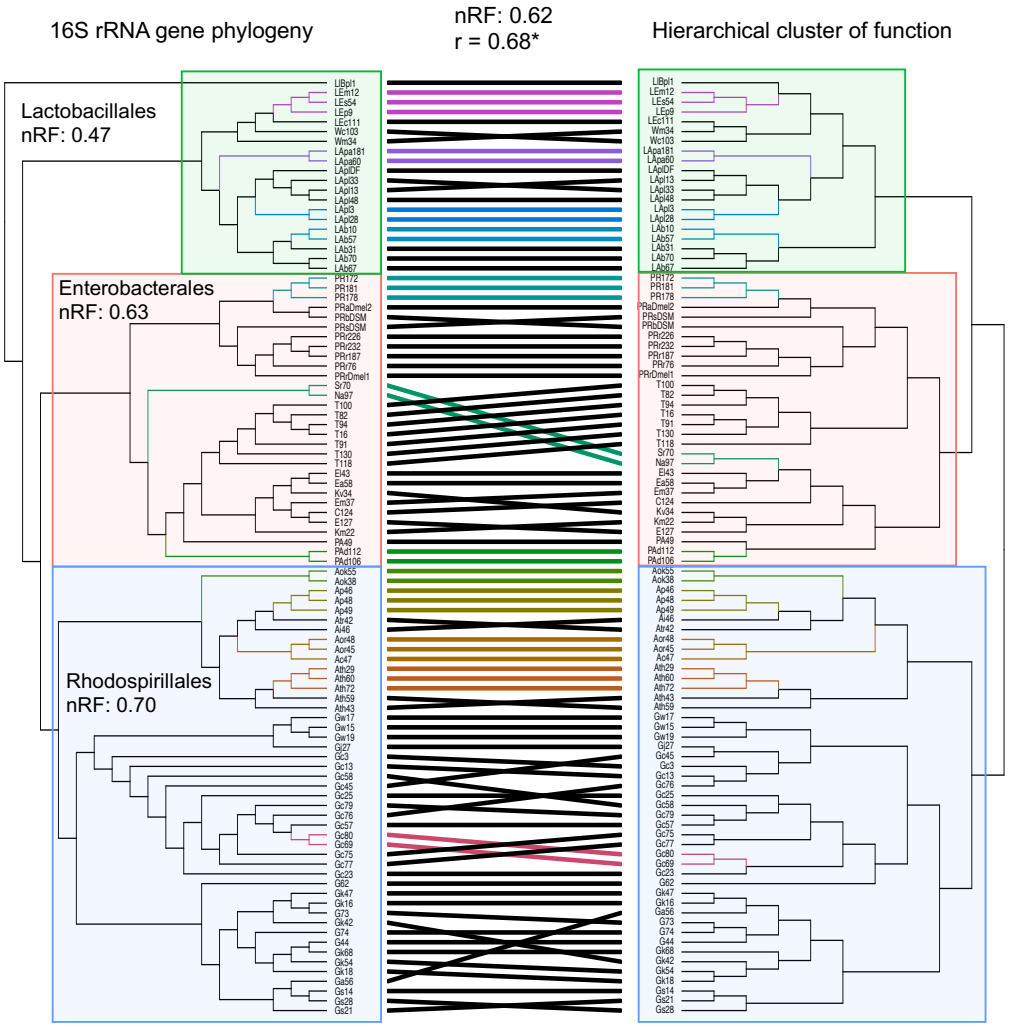


Figure S1.4. Mapping function onto phylogeny. (B) Correlation between 16S rRNA gene phylogeny and Bray-Curtis dissimilarity based hierarchical cluster of relative metabolic function counts. The normalized Robinson-Foulds index (nRF) and Mantel's test (based on Spearman's rank correlation) for the entire tanglegram are reported at the top. Each bacterial order is indicated by shaded box with the specific nRF obtained for each order. The tips of each dendrogram are connected by either black or colored line; the latter represents subtrees within each dendrogram that have the same topology. * represents p = 0.001.

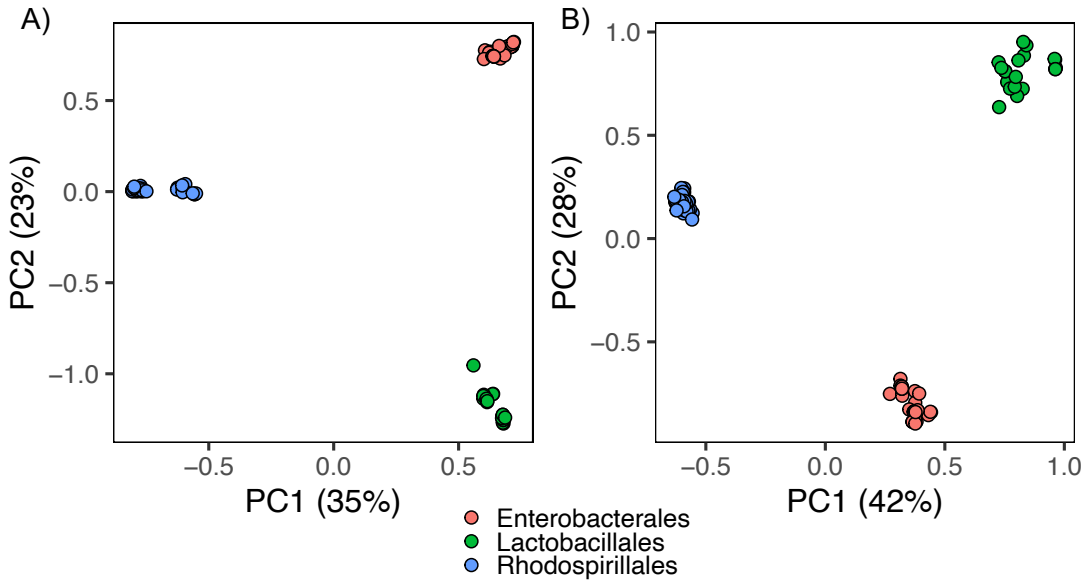


Figure S1.5. Principal coordinates anlaysis (PCoA) of orthogroup incidence. (A) Representation of all 13,170 orthogroups with at least three genomes present in each. (B) Visualization of the 1,055 metabolism-related orthogroups extracted from the full dataset. All points represent composite of orthogroup incidence for each genome analyzed with colors corresponding to order-level taxonomy. The percentages on each axis correspond to the amount of variation explained.

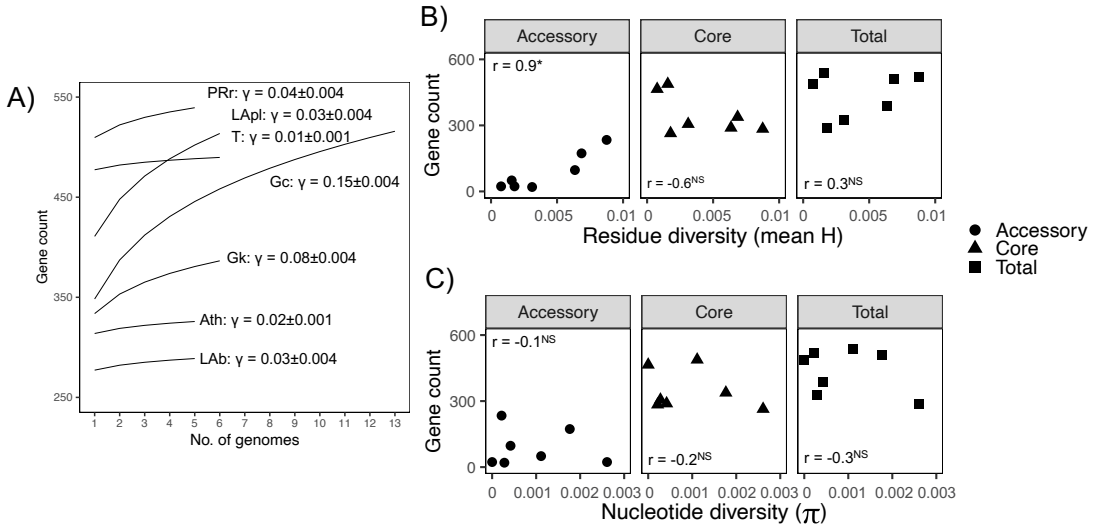


Figure S1.6. Strain diversity among metabolic pangenomes and gene distribution. (A) Rarefaction curves generated from Roary analysis using log-log linear model. (B) Correlation between pangenome distribution for each species with residue diversity (Shannon's entropy calculated from phylogenomic amino acid sequence alignment) and (C) nucleotide diversity among 16S rRNA gene alignments. γ in panel A represents the slope and standard error from the log-log linear model, which indicates an "open" pangenome for values < 1. For panels B and C, Pearson's correlation statistic is show. NS = not significant; * p = 0.003. Species identifiers: Ath = *A. thailandicus*, Gc = *G. cerinus*, Gk = *G. kondonii*, LAb = *L. brevis*, LApl = *L. plantarum*, PRr = *P. rettgeri*, and T = *Tatumella* sp.

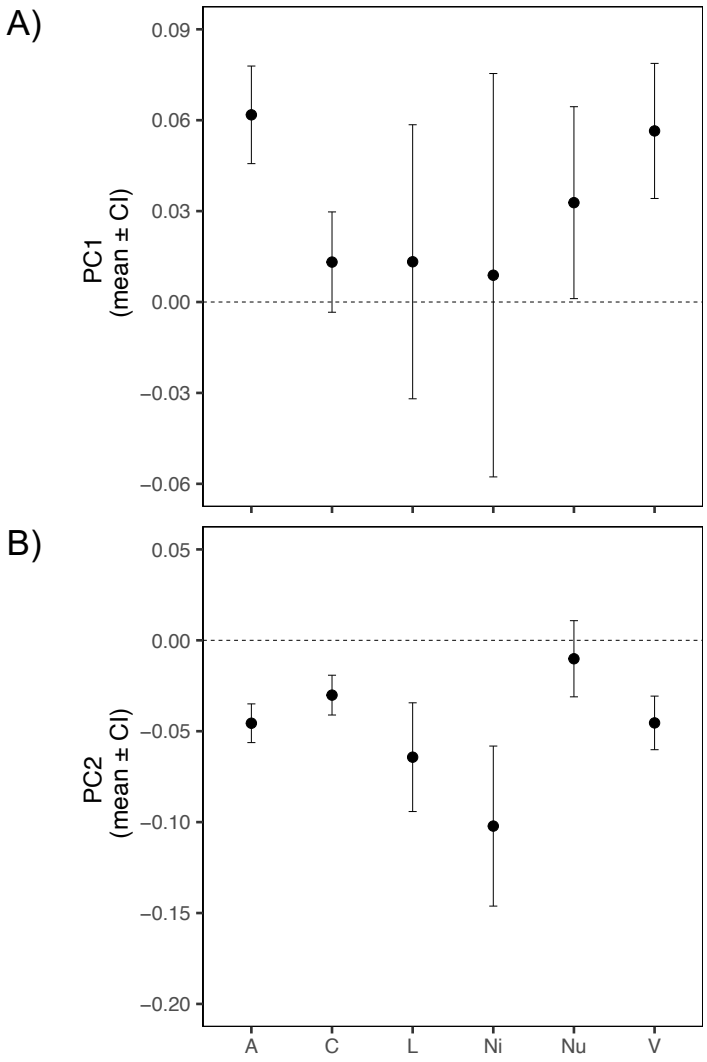


Figure S1.7. Association of metabolic function with composite representation of orthogroup incidence among prevalent species. (A) PC1 and (B) PC2 from PCoA in Figure 4. The means and 95% confidence interval are plotted. Non-overlapping confidence intervals indicate significant association with function and taxa deviations from PCoA center (0,0). RAST categories: A = amino acids and derivatives, C = Carbohydrates, V = cofactors, vitamins, prosthetic groups, and pigments, L = fatty acids, lipids, and isoprenoids, Ni = nitrogen metabolism, and Nu = nucleosides and nucleotides.

Table S1.1. List of bacteria used in comparative genomics analysis (A) Summary of genome sequencing and strain isolation for new *Drosophila*-associated bacteria (see Table S1.1C for media recipes)

Year isolated	**Fly replicate (sex)**	**Collection site**	**Isolation medium**	**Order**	**Family**	**Genus**	**Species**	**strain**	**ID**	**Median insert size (SD) [bp]**	**Total number of reads**
2015	1 (F)	Kitchen A	YPD	Enterobacterales	Enterobacteriaceae	Citrobacter	sp.	JGM124	C124	249 (109.79)	3,100,133
2015	2 (F)	Kitchen B	YPD	Enterobacterales	Enterobacteriaceae	Enterobacter	asburiae	JGM58	Ea58	386 (254.94)	3,426,835
2015	3 (F)	Kitchen B	YPD	Enterobacterales	Enterobacteriaceae	Enterobacter	ludwigii	JGM43	El43	264 (384.27)	2,663,001
2015	2 (F)	Kitchen B	YPD	Enterobacterales	Enterobacteriaceae	Enterobacter	mori	JGM37	Em37	379 (91.19)	2,931,967
2015	1 (F)	Kitchen A	YPD	Enterobacterales	Enterobacteriaceae	Enterobacter	sp.	JGM127	E127	364 (1,307.28)	2,228,975
2015	3 (F)	Kitchen B	YPD	Enterobacterales	Enterobacteriaceae	Klebsiella	michiganensis	JGM22	Km22	277 (130.64)	1,785,485

2015	4 M	Kitchen C	YPD	Enterobacterales	Enterobacteriaceae	Klebsiella	variicola	JGM34	Kv34	299 (147.33)	1,745,945
2015	4 M	Kitchen C	YPD	Enterobacterales	Erwiniaceae	Pantoea	dispersa	JGM106	PAd106	398 (273.46)	3,162,348
2015	4 M	Kitchen C	YPD	Enterobacterales	Erwiniaceae	Pantoea	dispersa	JGM112	PAd112	234 (110.56)	3,736,174
2015	5 (F)	Kitchen B	YPD	Enterobacterales	Erwiniaceae	Pantoea	sp.	JGM49	PA49	208 (63.18)	27,430,863
2015	6 (F)	Kitchen B	YPD	Enterobacterales	Erwiniaceae	Tatumella	sp.	JGM94	T94	379 (84.41)	5,141,122
2015	7 (F)	Kitchen B	YPD	Enterobacterales	Erwiniaceae	Tatumella	sp.	JGM100	T100	336 (106.24)	2,900,815
2015	8 (F)	Kitchen B	YPD	Enterobacterales	Erwiniaceae	Tatumella	sp.	JGM91	T91	370 (86.75)	3,728,278
2015	9 (F)	Kitchen B	YPD	Enterobacterales	Erwiniaceae	Tatumella	sp.	JGM82	T82	271 (121.19)	1,728,130
2015	10 M	Kitchen A	YPD	Enterobacterales	Erwiniaceae	Tatumella	sp.	JGM118	T118	236 (259.18)	4,595,051
2015	1 (F)	Kitchen A	YPD	Enterobacterales	Erwiniaceae	Tatumella	sp.	JGM130	T130	213 (86.26)	3,750,073
2015	11 (F)	Kitchen B	YPD	Enterobacterales	Erwiniaceae	Tatumella	sp.	JGM16	T16	272 (126.57)	1,916,022
2015	12 (F)	Kitchen D	YPD + tet	Enterobacterales	Morganellaceae	Providencia	rettgeri	JGM232	PRr232	375 (83.61)	5,187,746

2015	14 (F)	Kitchen B	YPD	Enterobacterales	Morganellaceae	Providencia	rettgeri	JGM76	PRr76	364 (692.08)	4,465,980
2015	17 (F)	Kitchen D	YPD + tet	Enterobacterales	Morganellaceae	Providencia	rettgeri	JGM226	PRr226	357 (679.79)	5,152,318
2015	13 (F)	Kitchen D	YPD + tet	Enterobacterales	Morganellaceae	Providencia	rettgeri	JGM187	PRr187	355 (870.92)	3,025,022
2015	13 (F)	Kitchen D	YPD + tet	Enterobacterales	Morganellaceae	Providencia	sp.	JGM181	PR181	178 (40.44)	20,779,498
2015	15 (F)	Kitchen D	YPD + tet	Enterobacterales	Morganellaceae	Providencia	sp.	JGM172	PR172	337 (2,376.93)	2,354,416
2015	16 (F)	Kitchen D	YPD + tet	Enterobacterales	Morganellaceae	Providencia	sp.	JGM178	PR178	371 (388.03)	4,675,512
2015	3 (F)	Kitchen B	YPD	Enterobacterales	Yersiniaceae	Nissabacter	archeti	JGM97	Na97	399 (117.85)	3,650,488
2015	18 (F)	Kitchen A	YPD	Enterobacterales	Yersiniaceae	Serratia	rubidaea	JGM70	Sr70	262 (130.50)	3,580,730
2019	19 M	Apple dumpster	mMRS + azide	Lactobacilales	Lactobacillaceae	Lactobacillus	brevis	Dm-2019-31	LAb31	225 (70.91)	8,912,486
2019	20 M	Apple dumpster	mMRS + azide	Lactobacilales	Lactobacillaceae	Lactobacillus	brevis	Dm-2019-57	LAb57	225 (70.99)	16,482,166
2019	21 M	Kitchen E	mMRS + azide	Lactobacilales	Lactobacillaceae	Lactobacillus	brevis	Dm-2019-70	LAb70	230 (71.49)	11,434,133

2019	22 M	Apple dumpster	mMRS + azide	Lactobacilales	Lactobacillaceae	Lactobacillus	brevis	Dm-2019-10	LAb10	240 (77.71)	9,450,050
2019	23 M	Kitchen E	mMRS + azide	Lactobacilales	Lactobacillaceae	Lactobacillus	brevis	Dm-2019-67	LAb67	236 (80.56)	9,853,100
2019	24 M	Apple dumpster	mMRS + azide	Lactobacilales	Lactobacillaceae	Lactobacillus	paracasei	Dm-2019-60	LApa60	245 (78.76)	7,155,294
2019	25 M	Apple dumpster	mMRS + azide	Lactobacilales	Lactobacillaceae	Lactobacillus	plantarum	Dm-2019-13	LApl13	237 (76.10)	6,852,346
2019	19 M	Apple dumpster	mMRS + azide	Lactobacilales	Lactobacillaceae	Lactobacillus	plantarum	Dm-2019-28	LApl28	193 (52.38)	6,947,542
2019	26 M	Kitchen E	mMRS + azide	Lactobacilales	Lactobacillaceae	Lactobacillus	plantarum	Dm-2019-48	LApl48	239 (76.91)	7,792,794
2019	22 M	Apple dumpster	mMRS + azide	Lactobacilales	Lactobacillaceae	Lactobacillus	plantarum	Dm-2019-33	LApl33	239 (188.00)	9,388,764

2019	27 M	Apple dumpster	mMRS + azide	Lactobacilales	Lactobacillaceae	Lactobacillus	plantarum	Dm-2019-3	LApl3	233 (74.32)	11,535,963
2019	29 M	Kitchen E	mMRS + azide	Lactobacilales	Leuconostocaceae	Leuconostoc	mesenteroides	Dm-2019-12	LEm12	250 (78.67)	5,190,537
2017	30 M	Apple dumpster	mMRS	Lactobacilales	Leuconostocaceae	Leuconostoc	pseudomesenteroides	Dm-9	LEp9	253 (87.89)	3,650,652
2019	28 M	Kitchen E	mMRS + azide	Lactobacilales	Leuconostocaceae	Leuconostoc	suionicum	Dm-2019-54	LEs54	239 (91.89)	14,269,857
2019	21 M	Kitchen E	mMRS + azide	Lactobacilales	Leuconostocaceae	Weissella	minor	Dm-2019-34	Wm34	182 (42.88)	28,432,172
2017	31 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Acetobacter	okinawensis	Dm-38	Aok38	359 (88.86)	2,339,511
2017	31 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Acetobacter	okinawensis	Dm-55	Aok55	337 (75.90)	1,760,234
2017	32 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Acetobacter	persici	Dm-48	Ap48	396 (1063.92)	3,229,753

2017	33 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Acetobacter	persici	Dm-49	Ap49	394 (651.97)	1,276,255
2017	32 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Acetobacter	persici	Dm-46	Ap46	377 (183.63)	2,368,062
2017	34 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Acetobacter	thailandicus	Dm-59	Ath59	373 (164.34)	4,765,208
2017	32 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Acetobacter	thailandicus	Dm-60	Ath60	388 (93.78)	4,040,653
2017	30 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Acetobacter	thailandicus	Dm-29	Ath29	370 (580.29)	2,055,491
2017	35 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Acetobacter	thailandicus	Dm-72	Ath72	376 (85.42)	2,705,121
2017	34 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	albidus	Dm-56	Ga56	397 (215.13)	3,874,546
2017	30 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	cerinus	Dm-3	Gc3	216 (133.06)	5,680,602
2017	36 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	cerinus	Dm-75	Gc75	378 (842.67)	2,439,056

2017	35 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	cerinus	Dm-76	Gc76	395 (376.08)	3,215,107
2017	35 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	cerinus	Dm-77	Gc77	378 (88.94)	3,494,791
2017	35 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	cerinus	Dm-69	Gc69	398 (82.52)	2,889,676
2017	35 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	cerinus	Dm-79	Gc79	397 (1,291.83)	2,247,038
2017	30 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	cerinus	Dm-13	Gc13	397 (573.75)	1,033,730
2017	32 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	cerinus	Dm-45	Gc45	362 (743.30)	1,533,908
2017	31 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	cerinus	Dm-57	Gc57	373 (410.21)	2,859,856
2017	31 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	cerinus	Dm-25	Gc25	380 (91.07)	2,215,489
2017	31 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	cerinus	Dm-58	Gc58	376 (86.46)	2,177,527

2017	31 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	cerinus	Dm-23	Gc23	362 (89.75)	1,908,386
2017	36 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	cerinus	Dm-80	Gc80	377 (748.31)	2,242,173
2017	34 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	japonicus	Dm-27	Gj27	372 (3141.69)	2,122,970
2017	33 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	kondonii	Dm-16	Gk16	276 (287.36)	6,114,446
2017	33 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	kondonii	Dm-18	Gk18	378 (92.06)	2,569,747
2017	32 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	kondonii	Dm-47	Gk47	392 (84.09)	4,477,981
2017	31 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	kondonii	Dm-54	Gk54	395 (82.80)	3,284,054
2017	36 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	kondonii	Dm-68	Gk68	380 (217.01)	3,357,111

2017	34 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	kondonii	Dm-42	Gk42	373 (91.80)	3,766,484
2017	36 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	sp.	Dm-73	G73	387 (87.33)	3,839,698
2017	36 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	sp.	Dm-74	G74	390 (98.86)	4,092,688
2017	32 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	sp.	Dm-62	G62	357 (210.00)	3,059,058
2017	34 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	sp.	Dm-44	G44	291 (162.42)	8,349,074
2017	34 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	sphaericus	Dm-28	Gs28	337 (731.22)	2,827,494
2017	34 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	sphaericus	Dm-21	Gs21	378 (206.20)	2,334,687
2017	33 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	sphaericus	Dm-14	Gs14	363 (92.66)	1,470,047
2017	33 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	wancherniae	Dm-17	Gw17	335 (1206.82)	1,303,785
2017	33 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	wancherniae	Dm-19	Gw19	268 (422.47)	7,352,214

2017	33 M	Apple dumpster	mMRS	Rhodospirillales	Acetobacteraceae	Gluconobacter	wancherniae	Dm-15	Gw15	379 (1686.07)	2,539,861

[M] indicates BLAST top matches with more than one species match.

Table S1.1. List of bacteria used in comparative genomics analysis (B) Publicly available genomes used in analyses

Order	**Family**	**Genus**	**Species**	**Strain**	**Accession ID**	**Referece genome**	***Drosophila*-associated**
Bacillales	Bacillaceae	Bacillus	subtilis	168	NC_000964	Yes	No
Enterobacterales	Morganellaceae	Providencia	burhodogranariea	DSM 19968	NZ_AKKL01000000	No	Yes
Enterobacterales	Morganellaceae	Providencia	sneebia	DSM 19967	NZ_AKKN01000000	No	Yes
Enterobacterales	Morganellaceae	Providencia	alcalifaciens	Dmel2	NZ_AKKM01000000	No	Yes
Enterobacterales	Morganellaceae	Providencia	rettgeri	Dmel1	NZ_AJSB01000000	No	Yes
Lactobacillales	Enterococcaceae	Enterococcus	faecalis	68EA1	NZ_KZ846072	Yes	No
Lactobacillales	Lactobacillaceae	Lactobacillus	paracasei	DmW181	NDXH00000000	No	Yes
Lactobacillales	Lactobacillaceae	Lactobacillus	plantarum	DF	NZ_CP013753	No	Yes
Lactobacillales	Lactobacillaceae	Lactococcus	lactis	BPL1	JRFX00000000	No	Yes
Lactobacillales	Lactobacillaceae	Leuconostoc	citreum	DmW_111	NDXG00000000	No	Yes
Lactobacillales	Lactobacillaceae	Weissella	cibaria	DmW_103	NDXJ00000000	No	Yes
Lactobacillales	Streptococcaceae	Streptococcus	pyogenes	NGAS638	NZ_CP010450	Yes	No
Pasteurellales	Pasteurellaceae	Haemophilus	influenzae	Hi375	NZ_CP009610	Yes	No
Pasteurellales	Pasteurellaceae	Pasteurella	multocida	20N	NZ_CP028926	Yes	No
Rhodospirillales	Acetobacteraceae	Acetobacter	cibinongensis	DmW_047	JOMQ00000000	No	Yes
Rhodospirillales	Acetobacteraceae	Acetobacter	indonesiensis	DmW_046	JOMP00000000	No	Yes
Rhodospirillales	Acetobacteraceae	Acetobacter	orientalis	DmW_045	JOMO00000000	No	Yes
Rhodospirillales	Acetobacteraceae	Acetobacter	orientalis	DmW_048	JOOY00000000	No	Yes
Rhodospirillales	Acetobacteraceae	Acetobacter	thailandicus (fromerly sp.)	DmW_043	JOMN00000000	No	Yes
Rhodospirillales	Acetobacteraceae	Acetobacter	tropicalis	DmW_042	JOMM00000000	No	Yes

Rhodospirillales	Acetobacteraceae	Granulibacter	bethesdensis	CGDNIH4	NZ_CP003182	Yes	No
Rhodospirillales	Rhodospirillaceae	Magnetospirillum	magneticum	AMB-1	AP007255	Yes	No
Rhodospirillales	Rhodospirillaceae	Rhodospirillum	rubrum	F11	CP003046	Yes	No
Vibrionales	Vibrionaceae	Vibrio	cholerae	MS6	AP014524 & AP014525	Yes	No

Table S1.1. List of bacteria used in comparative genomics analysis (C) Recipes for media used to isolate new bacteria strains

Medium (abbreviated name)	**Ingredients**
Yeast-peptone-dextrose (YPD)	1% yeast extract, 2% bacto-peptone, 2% glucose, and 1.5% agar (for plates only). All ingredients are from Sigma, except bacto-peptone from Becton Dickinson.
Yeast-peptone-dextrose + tetracycline (YPD + tet)	YPD medium supplemented with 25 μl ml^{-1} tetracycline (Sigma)
modified De Man, Rogosa, Sharpe (mMRS)	1.25% bacto-proteose peptone, 0.75% yeast extract, 2% glucose, 0.5% sodium acetate, 0.2% dipotassium hydrogen phosphate, 0.2% tirammonium citrate, 0.02% magnesium sulphate heptahydrate, 0.005% manganese sulfate tetrahydrate, and 1.2% agar (for plates only). All ingredients are from Sigma, except bacto-proteose peptone from Becton Dickinson.
modified De Man, Rogosa, Sharpe + azide (mMRS + azide)	mMRS medium supplemented with 0.005% azide (Acros), 0.005% bromocresol purple (Sigma), and 0.1% tween-80 (Sigma) and grown under low oxygen condition in a candle jar

Table S1.2. Gene annotation and model selection for 52 amino acid sequences used in phylogenomic reconstruction of *Drosophila*-associated bacteria

Gene family	Name	Length (aa)	Model of evolution	Eggnog mapper annotation
1	*gpsA*	309	LG+I+G4	Glycerol-3-phosphate dehydrogenase
2	*rpsB*	224	LG+G4	Belongs to the universal ribosomal protein uS2 family
3	*tsf*	262	LG+I+G4	Associates with the EF-Tu.GDP complex and induces the exchange of GDP to GTP. It remains bound to the aminoacyl-tRNA.EF-Tu.GTP complex up to the GTP hydrolysis stage on the ribosome
4	*rsmD*	113	LG+I+G4	Specifically methylates the guanine in position 966 of 16S rRNA in the assembled 30S particle
5	*rimP*	55	LG+G4	Required for maturation of 30S ribosomal subunits
6	*nusA*	385	LG+G4	Participates in both transcription termination and antitermination
7	*rbfA*	93	LG+G4	One of several proteins that assist in the late maturation steps of the functional core of the 30S ribosomal subunit. Associates with free 30S ribosomal subunits (but not with 30S subunits that are part of 70S ribosomes or polysomes). Required for efficient processing of 16S rRNA. May interact with the 5'-terminal helix region of 16S rRNA
8	*atpB*	131	mtZOA+G4	it plays a direct role in the translocation of protons across the membrane
9	*coaD*	137	LG+I+G4	Reversibly transfers an adenylyl group from ATP to 4'- phosphopantetheine, yielding dephospho-CoA (dPCoA) and pyrophosphate
10	*recJ*	413	LG+I+G4	single-stranded-DNA-specific exonuclease recJ
11	*ribF*	258	WAG+I+G4	Belongs to the ribF family
12	*plsY*	101	LG+I+G4	Catalyzes the transfer of an acyl group from acyl- phosphate (acyl-PO(4)) to glycerol-3-phosphate (G3P) to form lysophosphatidic acid (LPA). This enzyme utilizes acyl-phosphate as fatty acyl donor, but not acyl-CoA or acyl-ACP
13	*pyrG*	509	LG+I+G4	Catalyzes the ATP-dependent amination of UTP to CTP with either L-glutamine or ammonia as the source of nitrogen. Regulates intracellular CTP levels through interactions with the four ribonucleotide triphosphates
14	*prmA*	161	LG+G4	ribosomal protein L11
15	*ybeY*	77	LG+G4	Single strand-specific metallo-endoribonuclease

				involved in late-stage 70S ribosome quality control and in maturation of the 3' terminus of the 16S rRNA
16	*prfA*	341	LG+F+I+G4	Peptide chain release factor 1 directs the termination of translation in response to the peptide chain termination codons UAG and UAA
17	*smpB*	135	LG+I+G4	Required for rescue of stalled ribosomes mediated by trans-translation. Binds to transfer-messenger RNA (tmRNA), required for stable association of tmRNA with ribosomes. tmRNA and SmpB together mimic tRNA shape, replacing the anticodon stem-loop with SmpB. tmRNA is encoded by the ssrA gene
18	*glyQ*	274	LG+I+G4	glycyl-tRNA synthetase, alpha subunit
19	*mutY*	223	LG+I+G4	TIGRFAM A G-specific adenine glycosylase
20	*atpC*	70	LG+G4	Produces ATP from ADP in the presence of a proton gradient across the membrane
21	*atpD*	434	LG+I+G4	Produces ATP from ADP in the presence of a proton gradient across the membrane. The catalytic sites are hosted primarily by the beta subunits
22	*atpG*	230	LG+I+G4	Produces ATP from ADP in the presence of a proton gradient across the membrane. The gamma chain is believed to be important in regulating ATPase activity and the flow of protons through the CF(0) complex
23	*atpA*	484	LG+I+G4	Produces ATP from ADP in the presence of a proton gradient across the membrane. The alpha chain is a regulatory subunit
24	*atpH*	92	LG+G4	F(1)F(0) ATP synthase produces ATP from ADP in the presence of a proton or sodium gradient. F-type ATPases consist of two structural domains, F(1) containing the extramembraneous catalytic core and F(0) containing the membrane proton channel, linked together by a central stalk and a peripheral stalk. During catalysis, ATP synthesis in the catalytic domain of F(1) is coupled via a rotary mechanism of the central stalk subunits to proton translocation
25	*nusG*	172	LG+G4	Participates in transcription elongation, termination and antitermination
26	*tsaD*	273	LG+I+G4	Required for the formation of a threonylcarbamoyl group on adenosine at position 37 (t(6)A37) in tRNAs that read

				codons beginning with adenine. Is involved in the transfer of the threonylcarbamoyl moiety of threonylcarbamoyl-AMP (TC-AMP) to the N6 group of A37, together with TsaE and TsaB. TsaD likely plays a direct catalytic role in this reaction
27	*alaS*	745	LG+I+G4	Catalyzes the attachment of alanine to tRNA(Ala) in a two-step reaction alanine is first activated by ATP to form Ala- AMP and then transferred to the acceptor end of tRNA(Ala). Also edits incorrectly charged Ser-tRNA(Ala) and Gly-tRNA(Ala) via its editing domain
28	*miaA*	256	LG+F+I+G4	Catalyzes the transfer of a dimethylallyl group onto the adenine at position 37 in tRNAs that read codons beginning with uridine, leading to the formation of N6-(dimethylallyl)adenosine (i(6)A)
29	*obg*	311	LG+I+G4	An essential GTPase which binds GTP, GDP and possibly (p)ppGpp with moderate affinity, with high nucleotide exchange rates and a fairly low GTP hydrolysis rate. Plays a role in control of the cell cycle, stress response, ribosome biogenesis and in those bacteria that undergo differentiation, in morphogenesis control
30	*der*	394	LG+G4	GTPase that plays an essential role in the late steps of ribosome biogenesis
31	*rsmE*	150	LG+I+G4	Specifically methylates the N3 position of the uracil ring of uridine 1498 (m3U1498) in 16S rRNA. Acts on the fully assembled 30S ribosomal subunit
32	*ffh*	406	LG+F+I+G4	Involved in targeting and insertion of nascent membrane proteins into the cytoplasmic membrane. Binds to the hydrophobic signal sequence of the ribosome-nascent chain (RNC) as it emerges from the ribosomes. The SRP-RNC complex is then targeted to the cytoplasmic membrane where it interacts with the SRP receptor FtsY. Interaction with FtsY leads to the transfer of the RNC complex to the Sec translocase for insertion into the membrane, the hydrolysis of GTP by both Ffh and FtsY, and the dissociation of the SRP-FtsY complex into the individual components
33	*rpsP*	70	LG+G4	Belongs to the bacterial ribosomal protein bS16 family
34	*rimM*	66	LG+G4	An accessory protein needed during the final

				step in the assembly of 30S ribosomal subunit, possibly for assembly of the head region. Probably interacts with S19. Essential for efficient processing of 16S rRNA. May be needed both before and after RbfA during the maturation of 16S rRNA. It has affinity for free ribosomal 30S subunits but not for 70S ribosomes
35	*trmD*	212	LG+G4	tRNA (Guanine-1)-methyltransferase
36	*rplS*	106	LG+G4	This protein is located at the 30S-50S ribosomal subunit interface and may play a role in the structure and function of the aminoacyl-tRNA binding site
37	*clpX*	372	LG+I+G4	ATP-dependent specificity component of the Clp protease. It directs the protease to specific substrates. Can perform chaperone functions in the absence of ClpP
38	*proS*	384	LG+G4	Catalyzes the attachment of proline to tRNA(Pro) in a two-step reaction proline is first activated by ATP to form Pro- AMP and then transferred to the acceptor end of tRNA(Pro). As ProRS can inadvertently accommodate and process non-cognate amino acids such as alanine and cysteine, to avoid such errors it has two additional distinct editing activities against alanine. One activity is designated as 'pretransfer' editing and involves the tRNA(Pro)-independent hydrolysis of activated Ala-AMP. The other activity is designated 'posttransfer' editing and involves deacylation of mischarged Ala-tRNA(Pro). The misacylated Cys- tRNA(Pro) is not edited by ProRS
39	*typA*	582	LG+I+G4	GTP-binding protein TypA
40	*engB*	155	LG+I+G4	Necessary for normal cell division and for the maintenance of normal septation
41	*rseP*	232	LG+I+G4	zinc metalloprotease
42	*uppS*	224	LG+I+G4	Catalyzes the condensation of isopentenyl diphosphate (IPP) with allylic pyrophosphates generating different type of terpenoids
43	*frr*	170	LG+I+G4	Responsible for the release of ribosomes from messenger RNA at the termination of protein biosynthesis. May increase the efficiency of translation by recycling ribosomes from one round of translation to another
44	*pyrH*	230	LG+G4	Catalyzes the reversible phosphorylation of UMP to UDP

45	*era*	262	LG+G4	An essential GTPase that binds both GDP and GTP, with rapid nucleotide exchange. Plays a role in 16S rRNA processing and 30S ribosomal subunit biogenesis and possibly also in cell cycle regulation and energy metabolism
46	*aspS*	542	LG+I+G4	Aspartyl-tRNA synthetase with relaxed tRNA specificity since it is able to aspartylate not only its cognate tRNA(Asp) but also tRNA(Asn). Reaction proceeds in two steps L-aspartate is first activated by ATP to form Asp-AMP and then transferred to the acceptor end of tRNA(Asp Asn)
47	*apt*	156	LG+G4	Catalyzes a salvage reaction resulting in the formation of AMP, that is energically less costly than de novo synthesis
48	*rpsA*	485	LG+F+G4	thus facilitating recognition of the initiation point. It is needed to translate mRNA with a short Shine-Dalgarno (SD) purine-rich sequence
49	*recR*	169	LG+G4	May play a role in DNA repair. It seems to be involved in an RecBC-independent recombinational process of DNA repair. It may act with RecF and RecO
50	*pgk*	369	WAG+I+G4	Belongs to the phosphoglycerate kinase family
51	*yqgF*	108	LG+G4	Could be a nuclease involved in processing of the 5'-end of pre-16S rRNA
52	*grpE*	126	LG+I+G4	Participates actively in the response to hyperosmotic and heat shock by preventing the aggregation of stress-denatured proteins, in association with DnaK and GrpE. It is the nucleotide exchange factor for DnaK and may function as a thermosensor. Unfolded proteins bind initially to DnaJ

Table S1.3. Percent identity among 16S rRNA gene nucleotide sequences. (A) Enterobacterales.

	PRsDSM	PRbDSM	PRaDmel2	PR172	PR178	PR181	PRRrDmel1	PRr76	PRr187	PRr226	PRr232	Na97	Sr70	PAd106	PAd112	T118	Kv3
PRsDSM		98.8	98.6	99.2	99.2	99.2	99.2	99.2	99.2	99.2	99.2	94.4	96.7	93.5	96.1	93.2	96.
PRbDSM	98.8		99.5	98.8	98.8	98.8	99.1	98.7	98.9	98.9	98.9	94	96.2	93.5	95.9	93	95.
PRaDmel2	98.6	99.5		99.3	99.3	99.3	99.1	98.7	98.9	98.9	98.9	94	96.3	93.5	96	93	95.
PR172	99.2	98.8	99.3		100	100	99.4	99.2	99.4	99.4	99.4	94.7	96.8	93.7	96.5	93.5	96.
PR178	99.2	98.8	99.3	100		100	99.4	99.2	99.4	99.4	99.4	94.7	96.8	93.7	96.5	93.5	96.
PR181	99.2	98.8	99.3	100	100		99.4	99.2	99.4	99.4	99.4	94.7	96.8	93.7	96.5	93.5	96.
PRRrDmel1	99.2	99.1	99.1	99.4	99.4	99.4		99.7	99.8	99.8	99.8	94.6	96.5	93.8	96.2	93.6	95.
PRr76	99.2	98.7	98.7	99.2	99.2	99.2	99.7		99.8	99.8	99.8	94.5	96.5	93.8	96.4	93.4	95.
PRr187	99.2	98.9	98.9	99.4	99.4	99.4	99.8	99.8		100	100	94.7	96.7	93.8	96.4	93.6	96.
PRr226	99.2	98.9	98.9	99.4	99.4	99.4	99.8	99.8	100		100	94.7	96.7	93.8	96.4	93.6	96.
PRr232	99.2	98.9	98.9	99.4	99.4	99.4	99.8	99.8	100	100		94.7	96.7	93.8	96.4	93.6	96.
Na97	94.4	94	94	94.7	94.7	94.7	94.6	94.5	94.7	94.7	94.7		98.4	96.4	97.6	95.9	97.
Sr70	96.7	96.2	96.3	96.8	96.8	96.8	96.5	96.5	96.7	96.7	96.7	98.4		97.4	97.6	96.1	97.
PAd106	93.5	93.5	93.5	93.7	93.7	93.7	93.8	93.8	93.8	93.8	93.8	96.4	97.4		99.9	97	97.
PAd112	96.1	95.9	96	96.5	96.5	96.5	96.2	96.4	96.4	96.4	96.4	97.6	97.6	99.9		97	97.
T118	93.2	93	93	93.5	93.5	93.5	93.6	93.4	93.6	93.6	93.6	95.9	96.1	97	97		97.
Kv34	96.2	95.8	95.9	96.3	96.3	96.3	95.9	95.9	96.1	96.1	96.1	97.6	97.1	97.6	97.7	97.5	
E127	91.9	91.9	91.8	91.9	91.9	91.9	92	92	92	92	92	93.5	93.5	93.9	94	94.3	93.
C124	95.3	94.8	94.8	95.4	95.4	95.4	95.1	94.9	95.1	95.1	95.1	96.7	97.3	96.6	97.5	96.4	9
T16	93.7	93.5	93.5	94.1	94.1	94.1	94.1	94.1	94.1	94.1	94.1	96.4	96.8	97.3	97.8	98.3	97.
T91	93.7	93.5	93.5	94.1	94.1	94.1	94.1	94.1	94.1	94.1	94.1	96.4	96.8	97.3	97.8	98.3	97.
T82	93.8	93.6	93.6	94.1	94.1	94.1	94.1	94.2	94.2	94.2	94.2	96.4	96.8	97.3	97.8	98.3	97.
T94	93.8	93.6	93.6	94.1	94.1	94.1	94.1	94.2	94.2	94.2	94.2	96.4	96.8	97.3	97.8	98.3	97.
T100	93.8	93.6	93.6	94.1	94.1	94.1	94.1	94.2	94.2	94.2	94.2	96.4	96.8	97.3	97.8	98.3	97.
T130	95.6	95.2	95.3	96	96	96	95.7	95.8	95.8	95.8	95.8	97.3	96.8	97.4	97.8	98.3	97.

Km22	92	92.4	92.2	92	92	92	92.3	92	92.1	92.1	92.1	94.7	96.2	96.3	97.3	96.6	98.
Ea58	94.8	94.3	94.3	94.7	94.7	94.7	94.7	94.7	94.7	94.7	94.7	97.1	97.5	97.9	98.1	97.6	98.
El43	95.8	95.4	95.3	95.8	95.8	95.8	95.5	95.7	95.7	95.7	95.7	97.2	97.6	97.5	97.8	97.5	98.
Em37	93.2	92.8	92.8	93.2	93.2	93.2	93.1	93.1	93.1	93.1	93.1	96.5	96.9	96.6	97.2	96.1	97.
PA49	95.1	95.5	95.3	95.3	95.3	95.3	94.9	94.7	94.9	94.9	94.9	97.4	97.5	97.7	97.9	97.7	98.

Table S1.3. Percent identity among 16S rRNA gene nucleotide sequences. (B) Lactobacillales.

	LlBpl1	LEc111	LEp9	LEm12	LEs54	Wc103	Wm34	LApa60	LApa181	LAb10	LAb57	LAb31	LAb67	LAb70	LAplDF	LApl3	LApl28	LApl13	LApl33	LApl48
LlBpl1		82.8	83.2	83	83.1	76.7	85.8	85.7	87.4	85	85	85	86.1	86.1	85.5	85.5	85.5	86.9	85.5	87.9
LEc111	82.8		97.8	97.9	97.9	81.4	90.4	86.6	88.2	84.9	84.9	84.9	85.6	85.6	85	85	85	87.1	85.1	87
LEp9	83.2	97.8		99.6	99.6	81.6	91.4	86.7	88.5	85.1	85.1	85.1	86.1	86.1	85.8	85.7	85.7	88.3	85.8	88.1
LEm12	83	97.9	99.6		99.9	81.7	91.3	86.7	88.5	85.2	85.2	85.2	86.3	86.3	85.9	85.8	85.8	88.4	86	88.2
LEs54	83.1	97.9	99.6	99.9		81.7	91.2	86.7	88.6	85.1	85.1	85.1	86.2	86.2	85.9	85.8	85.8	88.3	85.9	88.1
Wc103	76.7	81.4	81.6	81.7	81.7		96.9	86.8	89.2	79.8	79.8	79.9	87.5	87.5	79.2	79.6	79.6	89.3	79.5	89.3
Wm34	85.8	90.4	91.4	91.3	91.2	96.9		92.7	92.7	91.9	91.9	92	92	92	92	92.1	92.1	92.3	92.3	92.3
LApa60	85.7	86.6	86.7	86.7	86.7	86.8	92.7		100	92	92	92.1	92.1	92.1	92.1	92.1	92.1	93.3	92.2	93.4
LApa181	87.4	88.2	88.5	88.5	88.6	89.2	92.7	100		93.5	93.5	93.5	93.5	93.5	93.3	93.4	93.4	93.3	93.5	93.5
LAb10	85	84.9	85.1	85.2	85.1	79.8	91.9	92	93.5		100	99.9	99.9	99.9	93.4	93.4	93.4	95.6	93.4	95.8
LAb57	85	84.9	85.1	85.2	85.1	79.8	91.9	92	93.5	100		99.9	99.9	99.9	93.4	93.4	93.4	95.6	93.4	95.8
LAb31	85	84.9	85.1	85.2	85.1	79.9	92	92.1	93.5	99.9	99.9		100	100	93.4	93.4	93.4	95.7	93.4	95.9
LAb67	86.1	85.6	86.1	86.3	86.2	87.5	92	92.1	93.5	99.9	99.9	100		100	94.6	94.8	94.8	95.7	94.7	95.9
LAb70	86.1	85.6	86.1	86.3	86.2	87.5	92	92.1	93.5	99.9	99.9	100	100		94.6	94.8	94.8	95.7	94.7	95.9
LAplDF	85.5	85	85.8	85.9	85.9	79.2	92	92.1	93.3	93.4	93.4	93.4	94.6	94.6		99.7	99.7	99.8	99.9	99.8
LApl3	85.5	85	85.7	85.8	85.8	79.6	92.1	92.1	93.4	93.4	93.4	93.4	94.8	94.8	99.7		100	99.9	99.9	99.9

LApl28	85.5	85	85.7	85.8	85.8	79.6	92.1	92.1	93.4	93.4	93.4	93.4	94.8	94.8	99.7	100		99.9	99.9	99.9
LApl13	86.9	87.1	88.3	88.4	88.3	89.3	92.3	93.3	93.3	95.6	95.6	95.7	95.7	95.7	99.8	99.9	99.9		100	100
LApl33	85.5	85.1	85.8	86	85.9	79.5	92.3	92.2	93.5	93.4	93.4	93.4	94.7	94.7	99.9	99.9	99.9	100		100
LApl48	87.9	87	88.1	88.2	88.1	89.3	92.3	93.4	93.5	95.8	95.8	95.9	95.9	95.9	99.8	99.9	99.9	100	100	

Table S1.3. Percent identity among 16S rRNA gene nucleotide sequences. (C) Rhodospirillales.

	G62	Gs14	Gs21	Gs28	Ga56	Gk16	Gk42	Gk47	G73	G74	G44	Gk18	Gk54	Gk68	Gw15	Gw17	Gw19	Gj27	Gc80	Gc13	Gc25
G62		99.7	99.7	99.7	99.7	99.5	99.5	99.5	99.5	99.5	99.6	99.6	99.6	99.6	98.4	98.4	98.4	98	98.1	98.1	98.1
Gs14	99.7		100	100	100	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	98.3	98.3	98.3	98.1	98.2	98.2	98.2
Gs21	99.7	100		100	100	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	98.3	98.3	98.3	98.1	98.2	98.2	98.2
Gs28	99.7	100	100		100	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	98.3	98.3	98.3	98.1	98.2	98.2	98.2
Ga56	99.7	100	100	100		99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	98.3	98.3	98.3	98.1	98.2	98.2	98.2
Gk16	99.5	99.9	99.9	99.9	99.9		100	100	100	99.9	99.9	99.9	99.9	99.9	98.2	98.2	98.2	97.9	98.2	98.2	98.2
Gk42	99.5	99.9	99.9	99.9	99.9	100		100	100	99.9	99.9	99.9	99.9	99.9	98.2	98.2	98.2	97.9	98.2	98.2	98.2
Gk47	99.5	99.9	99.9	99.9	99.9	100	100		100	99.9	99.9	99.9	99.9	99.9	98.2	98.2	98.2	97.9	98.2	98.2	98.2
G73	99.5	99.9	99.9	99.9	99.9	100	100	100		99.9	99.9	99.9	99.9	99.9	98.2	98.2	98.2	97.9	98.2	98.2	98.2
G74	99.5	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9		99.9	99.9	99.9	99.9	98.2	98.2	98.2	97.9	98.2	98.2	98.2
G44	99.6	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9		100	100	100	98.3	98.3	98.3	98	98.3	98.3	98.3
Gk18	99.6	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	100		100	100	98.3	98.3	98.3	98	98.3	98.3	98.3
Gk54	99.6	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	100	100		100	98.3	98.3	98.3	98	98.3	98.3	98.3

Gk68	99.6	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	100	100	100		98.3	98.3	98.3	98	98.3	98.3	98.3
Gw15	98.4	98.3	98.3	98.3	98.3	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3		100	100	99.2	98.5	98.5	98.5
Gw17	98.4	98.3	98.3	98.3	98.3	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3	100		100	99.2	98.5	98.5	98.5
Gw19	98.4	98.3	98.3	98.3	98.3	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3	100	100		99.2	98.5	98.5	98.5
Gj27	98	98.1	98.1	98.1	98.1	97.9	97.9	97.9	97.9	97.9	98	98	98	98	99.2	99.2	99.2		99.3	99.3	99.3
Gc80	98.1	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3	98.5	98.5	98.5	99.3		100	100
Gc13	98.1	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3	98.5	98.5	98.5	99.3	100		100
Gc25	98.1	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3	98.5	98.5	98.5	99.3	100	100	
Gc45	98.1	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3	98.5	98.5	98.5	99.3	100	100	100
Gc57	98.1	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3	98.5	98.5	98.5	99.3	100	100	100
Gc58	98.1	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3	98.5	98.5	98.5	99.3	100	100	100
Gc69	98.1	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3	98.5	98.5	98.5	99.3	100	100	100
Gc75	98.1	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3	98.5	98.5	98.5	99.3	100	100	100
Gc76	98.1	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3	98.5	98.5	98.5	99.3	100	100	100
Gc77	98.1	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3	98.5	98.5	98.5	99.3	100	100	100

Gc79	98.1	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3	98.5	98.5	98.5	99.3	100	100	100
Gc23	98.1	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.3	98.3	98.3	98.3	98.5	98.5	98.5	99.3	100	100	100
Gc3	98.3	98.4	98.4	98.4	98.4	98.4	98.4	98.4	98.4	98.4	98.5	98.5	98.5	98.5	98.6	98.6	98.6	99.4	99.8	99.8	99.8
Aok38	95.3	95.5	95.5	95.5	95.5	95.4	95.4	95.4	95.4	95.5	95.5	95.5	95.5	95.5	94.9	94.9	94.9	95.3	95.3	95.3	95.3
Aok55	95.3	95.5	95.5	95.5	95.5	95.4	95.4	95.4	95.4	95.5	95.5	95.5	95.5	95.5	94.9	94.9	94.9	95.3	95.3	95.3	95.3
Ath43	95.7	96	96	96	96	95.9	95.9	95.9	95.9	96	96	96	96	96	94.7	94.7	94.7	95.1	95.2	95.2	95.2
Ath29	95.5	95.9	95.9	95.9	95.9	95.7	95.7	95.7	95.7	95.9	95.8	95.8	95.8	95.8	94.5	94.5	94.5	94.9	95.1	95.1	95.1
Ath59	95.5	95.9	95.9	95.9	95.9	95.7	95.7	95.7	95.7	95.9	95.8	95.8	95.8	95.8	94.5	94.5	94.5	94.9	95.1	95.1	95.1
Ath60	95.5	95.9	95.9	95.9	95.9	95.7	95.7	95.7	95.7	95.9	95.8	95.8	95.8	95.8	94.5	94.5	94.5	94.9	95.1	95.1	95.1
Ath72	95.5	95.9	95.9	95.9	95.9	95.7	95.7	95.7	95.7	95.9	95.8	95.8	95.8	95.8	94.5	94.5	94.5	94.9	95.1	95.1	95.1
Atr42	95.5	95.7	95.7	95.7	95.7	95.5	95.5	95.5	95.5	95.7	95.6	95.6	95.6	95.6	94.6	94.6	94.6	94.9	95	95	95
Ap48	95.1	95.5	95.5	95.5	95.5	95.3	95.3	95.3	95.3	95.5	95.4	95.4	95.4	95.4	94.3	94.3	94.3	94.7	94.9	94.9	94.9
Ap49	95.2	95.5	95.5	95.5	95.5	95.4	95.4	95.4	95.4	95.5	95.5	95.5	95.5	95.5	94.4	94.4	94.4	94.8	94.9	94.9	94.9
Ap46	95.1	95.5	95.5	95.5	95.5	95.3	95.3	95.3	95.3	95.5	95.4	95.4	95.4	95.4	94.3	94.3	94.3	94.7	94.9	94.9	94.9
Ai46	95.4	95.7	95.7	95.7	95.7	95.6	95.6	95.6	95.6	95.7	95.7	95.7	95.7	95.7	94.6	94.6	94.6	95.2	95.3	95.3	95.3

Ac47	95.1	95.5	95.5	95.5	95.5	95.3	95.3	95.3	95.3	95.5	95.4	95.4	95.4	95.4	94.4	94.4	94.4	94.8	94.9	94.9	94.9
Aor45	95.7	96	96	96	96	95.9	95.9	95.9	95.9	96	95.9	95.9	95.9	95.9	94.9	94.9	94.9	95.2	95.3	95.3	95.3
Aor48	95.7	96	96	96	96	95.9	95.9	95.9	95.9	96	95.9	95.9	95.9	95.9	94.9	94.9	94.9	95.2	95.3	95.3	95.3

Table S1.4. *Post hoc* pairwise comparisons for PERMANOVAs

(A) Comparisons among orders for RAST normalized function counts (Figure 3)			
Factor	F statistic	Effect size (R^2)	p-value (*post hoc* FDR adjusted)
Order	61.72	0.57	0.001
Enterobacterales vs. Lactobacillales	36.10	0.43	0.001 (0.001)
Enterobacterales vs. Rhodospirillales	79.96	0.52	0.001 (0.001)
Lactobacillales vs. Rhodospirillales	73.89	0.54	0.001 (0.001)
(B) Comparisons among orders for orthogroup incidence for all taxa (Figure S5A)			
Factor	F statistic	Effect size (R^2)	p-value (*post hoc* FDR adjusted)
Order	63.24	0.58	0.001
Enterobacterales vs. Lactobacillales	44.32	0.48	0.001 (0.001)
Enterobacterales vs. Rhodospirillales	75.67	0.51	0.001 (0.001)
Lactobacillales vs. Rhodospirillales	66.66	0.51	0.001 (0.001)
(C) Comparisons among orders for metabolic orthogroup incidence for all taxa (Figure S5B)			
Factor	F statistic	Effect size (R^2)	p-value (*post hoc* FDR adjusted)
Order	106.52	0.70	0.001
Enterobacterales vs. Lactobacillales	69.81	0.59	0.001 (0.001)
Enterobacterales vs. Rhodospirillales	133.93	0.64	0.001 (0.001)
Lactobacillales vs. Rhodospirillales	116.00	0.64	0.001 (0.001)
(D) Comparisons among orders for metabolic orthogroup incidence for prevalent species (Figure 4)			
Factor	F statistic	Effect size (R^2)	p-value (*post hoc* FDR adjusted)
Species	775.23	0.99	0.001
Ath vs. Gc	326.80	0.95	0.002 (0.005)
Ath vs. Gk	369.79	0.98	0.001 (0.005)
Ath vs. LAb	1293.69	0.99	0.006 (0.006)
Ath vs. LApl	1008.25	0.99	0.004 (0.005)
Ath vs. PRr	4709.46	1.00	0.013 (0.013)
Ath vs. T	8199.83	1.00	0.004 (0.005)
Gc vs. Gk	117.45	0.87	0.001 (0.005)
Gc vs. LAb	988.79	0.98	0.002 (0.005)
Gc vs. LApl	786.48	0.98	0.001 (0.005)
Gc vs. PRr	710.41	0.98	0.002 (0.005)
Gc vs. T	814.54	0.98	0.001 (0.005)
Gk vs. LAb	912.16	0.99	0.003 (0.005)

Gk vs. LApl	711.58	0.99	0.003 (0.005)
Gk vs. PRr	1003.73	0.99	0.004 (0.005)
Gk vs. T	1190.39	0.99	0.005 (0.006)
LAb vs. LApl	382.28	0.98	0.002 (0.005)
LAb vs. PRr	976.67	0.99	0.006 (0.006)
LAb vs. T	1287.88	0.99	0.004 (0.005)
LApl vs. PRr	801.15	0.99	0.003 (0.005)
LApl vs. T	991.91	0.99	0.004 (0.005)
PRr vs. T	2565.03	1.00	0.005 (0.006)

Table S1.5. Top metabolic orthogroups of prevalent species.

Orthogroup	RAST category	Function	Ath	Gc	Gk	LAb	LApl	PRr	T	p-value	FDR adjusted p-value
OG0001426	A	5-methylthioribose kinase (EC 2.7.1.100)		■	■				■	4.32E-13	8.89E-12
OG0000885	A	L-asparaginase (EC 3.5.1.1)	■			■	■	■	■	7.22E-13	8.89E-12
OG0002254	A	Cysteine desulfurase (EC 2.8.1.7)		■	■					7.22E-13	8.89E-12
OG0002337	A	4-hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27)		■					■	7.22E-13	8.89E-12
OG0002466	A	Periplasmic aromatic amino acid aminotransferase beta precursor (EC 2.6.1.57)		■	■					7.22E-13	8.89E-12
OG0002521	A	Ornithine cyclodeaminase (EC 4.3.1.12)		■	■					7.22E-13	8.89E-12
OG0000138	C	2-Keto-D-gluconate dehydrogenase (EC 1.1.99.4), membrane-bound, flavoprotein		■	■				■	4.32E-13	8.89E-12
OG0000348	C	2-Keto-D-gluconate dehydrogenase (EC 1.1.99.4), membrane-bound, gamma subunit		■	■				■	4.32E-13	8.89E-12
OG0000746	C	Gluconate 2-dehydrogenase (EC 1.1.99.3), membrane-bound, gamma subunit		■	■				■	4.32E-13	8.89E-12
OG0000769	C	Gluconate 2-dehydrogenase (EC 1.1.99.3), membrane-bound, flavoprotein		■	■				■	4.32E-13	8.89E-12
OG0001120	C	N-acetylglucosamine kinase of eukaryotic type (EC 2.7.1.59)		■	■		■			4.32E-13	8.89E-12
OG0001605	C	Glucokinase (EC 2.7.1.2)/Galactokinase (EC 2.7.1.6)		■	■				■	4.32E-13	8.89E-12
OG0001656	C	Inner membrane protein YghQ, probably involved in polysaccharide biosynthesis		■	■				■	4.32E-13	8.89E-12
OG0000161	C	Alcohol dehydrogenase (EC 1.1.1.1)/Acetaldehyde dehydrogenase (EC 1.2.1.10)	■			■	■	■	■	7.22E-13	8.89E-12
OG0001300	C	N-acetyl glucosamine transporter, NagP		■	■					7.22E-13	8.89E-12
OG0001429	C	Hydroxyacylglutathione hydrolase (EC 3.1.2.6)		■	■					7.22E-13	8.89E-12

OG0002046	C	1,4-alpha-glucan (glycogen) branching enzyme, GH-13-type (EC 2.4.1.18)		■			■			7.22E-13	8.89E-12
OG0002119	C	Xylonolactonase (EC 3.1.1.68)		■	■					7.22E-13	8.89E-12
OG0002176	C	Phosphoenolpyruvate carboxylase (EC 4.1.1.31)		■	■					7.22E-13	8.89E-12
OG0002207	C	Broad-specificity glycerol dehydrogenase (EC 1.1.99.22), subunit SldB/Glucose dehydrogenase, PQQ-dependent (EC 1.1.5.2)		■	■					7.22E-13	8.89E-12
OG0000275	L	Acetyl-CoA acetyltransferase (EC 2.3.1.9)/3-ketoacyl-CoA thiolase (EC 2.3.1.16)	■			■		■	■	4.32E-13	8.89E-12
OG0001484	L	CDP-diacylglycerol pyrophosphatase (EC 3.6.1.26)		■	■				■	4.32E-13	8.89E-12
OG0001643	Nu	Cytosine/purine/uracil/thiamine/allantoin permease family protein		■			■		■	4.32E-13	8.89E-12
OG0001428	Nu	Xanthine dehydrogenase, molybdenum binding subunit (EC 1.17.1.4)		■	■					7.22E-13	8.89E-12
OG0002047	Nu	Xanthine and CO dehydrogenases maturation factor, XdhC/CoxF family		■	■					7.22E-13	8.89E-12
OG0002105	Nu	Xanthine dehydrogenase iron-sulfur subunit (EC 1.17.1.4)/Xanthine dehydrogenase, FAD binding subunit (EC 1.17.1.4)		■	■					7.22E-13	8.89E-12
OG0002258	Nu	Adenosine deaminase (EC 3.5.4.4)		■	■					7.22E-13	8.89E-12
OG0002563	Nu	Thymidine kinase (EC 2.7.1.21)		■	■					7.22E-13	8.89E-12
OG0002573	Nu	CTP:molybdopterin cytidylyltransferase		■	■					7.22E-13	8.89E-12
OG0000808	V	2-dehydropantoate 2-reductase (EC 1.1.1.169)	■			■	■	■	■	7.22E-13	8.89E-12
OG0001649	V	Hydroxymethylpyrimidine ABC transporter, transmembrane component		■					■	7.22E-13	8.89E-12

Black indicates an orthogroup is present in all strains and white indicates absent in all strains.

Table S1.6. Subsystem enrichment analysis of prevalent species. (A) *Acetobacter thailandicus*.

	Present		Absent				
Subsystems	**Accessory**	**Core**	**Accessory**	**Core**	**Odds ratio**	**Fisher's p-value**	**FDR corrected p-value**
L-2-amino-thiazoline-4-carboxylic acid-Lcysteine conversion	3	0	23	473	20.17078	0.002233	0.167458
Nonmevalonate Branch of Isoprenoid Biosynthesis	2	4	24	470	6.473515	0.059617	1
5-FCL-like protein	3	13	23	460	3.733465	0.068749	1
Thiamin biosynthesis	2	5	24	469	5.543674	0.074238	1
Pyridoxin (Vitamin B6) Biosynthesis	2	8	24	466	3.864988	0.123729	1
Histidine Biosynthesis	2	9	24	465	3.507673	0.141681	1
Pyruvate Alanine Serine Interconversions	1	3	25	472	4.690937	0.234332	1
Citrate Metabolism, Transport, and Regulation	1	3	25	472	4.690937	0.234332	1
Pyruvate metabolism I: anaplerotic r	1	3	25	472	4.690937	0.234332	1

eactions, PEP							
Biotin biosynthesis	1	4	25	471	3.749867	0.274387	1
Glutamine, Glutamate, Aspartate and Asparagine Biosynthesis	2	16	24	458	2.116021	0.277532	1
Biotin biosynthesis Experimental	1	5	25	470	3.121559	0.312423	1
Folate biosynthesis cluster	1	6	25	469	2.672094	0.348538	1
Coenzyme A Biosynthesis	1	8	25	467	2.071412	0.415374	1
Fatty Acid Biosynthesis FASII	1	9	25	466	1.861057	0.446269	1
Threonine and Homoserine Biosynthesis	1	10	25	465	1.688686	0.475591	1
Coenzyme B 12 biosynthesis	1	15	25	460	1.14966	0.601177	1
Cobalamin synthesis	0	11	26	465	0	1	1
Pentose phosphate pathway	0	9	26	467	0	1	1
Ammonia assimilation	0	10	26	466	0	1	1
Lysine Biosynthesis DAP	0	8	26	468	0	1	1

Pathway, GJO scratch							
Lysine Biosynthesis DAP Pathway	0	8	26	468	0	1	1
Polyamine Metabolism	0	3	26	473	0	1	1
Glycine cleavage system	0	3	26	473	0	1	1
Glycine and Serine Utilization	0	13	26	463	0	1	1
Threonine degradation	0	2	26	474	0	1	1
Methylglyoxal Metabolism	0	7	26	469	0	1	1
Pyruvate metabolism II: acetyl-CoA, acetogenesis from pyruvate	0	7	26	469	0	1	1
Creatine and Creatinine Degradation	0	4	26	472	0	1	1
D-gluconate and ketogluconates metabolism	0	4	26	472	0	1	1
Mannose Metabolism	0	3	26	473	0	1	1
Branched-Chain Amino Acid Biosynthesis	0	9	26	467	0	1	1

Acetolactate synthase subunits	0	2	26	474	0	1	1
Methionine Biosynthesis	0	18	26	458	0	1	1
Cysteine Biosynthesis	0	13	26	463	0	1	1
Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilate and more.	0	11	26	465	0	1	1
Tryptophan synthesis	0	9	26	467	0	1	1
Heme and Siroheme Biosynthesis	0	12	26	464	0	1	1
Methionine Degradation	0	7	26	469	0	1	1
NAD and NADP cofactor biosynthesis global	0	8	26	468	0	1	1
Alanine biosynthesis	0	6	26	470	0	1	1
Dehydrogenase complexes	0	7	26	469	0	1	1
Methionine Salvage	0	5	26	471	0	1	1
Proline, 4-hydroxyproline uptake a	0	2	26	474	0	1	1

nd utilizatio n							
De Novo Pur ine Biosynth esis	0	12	26	464	0	1	1
Purine conv ersions	0	11	26	465	0	1	1
Glycolate, gl yoxylate int erconversio ns	0	2	26	474	0	1	1
Common Pa thway For Sy nthesis of A romatic Co mpounds (D AHP synthas e to chorism ate)	0	6	26	470	0	1	1
Proline Synt hesis	0	4	26	472	0	1	1
pyrimidine c onversions	0	9	26	467	0	1	1
Fermentatio ns: Lactate	0	1	26	475	0	1	1
Riboflavin, F MN and FAD metabolism	0	8	26	468	0	1	1
Histidine De gradation	0	7	26	469	0	1	1
Leucine Bios ynthesis	0	5	26	471	0	1	1
Arginine an d Ornithine Degradation	0	6	26	470	0	1	1
Trehalose Bi osynthesis	0	2	26	474	0	1	1
Purine Utiliz ation	0	4	26	472	0	1	1

Serine Biosynthesis	0	7	26	469	0	1	1
Phenylalanine and Tyrosine Branches from Chorismate	0	3	26	473	0	1	1
riboflavin to FAD	0	4	26	472	0	1	1
Triacylglycerol metabolism	0	1	26	475	0	1	1
One-carbon metabolism by tetrahydropterines	0	6	26	470	0	1	1
Glycine Biosynthesis	0	1	26	475	0	1	1
Urea decomposition	0	3	26	473	0	1	1
Urea carboxylase and Allophanate hydrolase cluster	0	2	26	474	0	1	1
Arginine Deiminase Pathway	0	2	26	474	0	1	1
Lactate utilization	0	4	26	472	0	1	1
Isoprenoinds for Quinones	0	3	26	473	0	1	1
Polyprenyl Diphosphate Biosynthesis	0	2	26	474	0	1	1
Glycerol and Glycerol-3-	0	5	26	471	0	1	1

phosphate Uptake and Utilization							
Xanthine Metabolism in Bacteria	0	1	26	475	0	1	1
Coenzyme A Biosynthesis cluster	0	3	26	473	0	1	1
Glutamine synthetases	0	1	26	475	0	1	1
Isoprenoid Biosynthesis: Interconversions	0	1	26	475	0	1	1
GMP synthase	0	2	26	474	0	1	1

Table S1.6. Subsystem enrichment analysis of prevalent species. (B) *Gluconobacter cerinus*.

	Present		Absent				
Subsystems	**Accessory**	**Core**	**Accessory**	**Core**	**Odds ratio**	**Fisher's p-value**	**FDR corrected p-value**
Histidine Degradation	15	0	310	712	2.294779	0.020655	1
Methylglyoxal Metabolism	15	3	310	709	1.904718	0.052763	1
Urease subunits	10	0	315	717	2.274239	0.056043	1
Urea decomposition	10	0	315	717	2.274239	0.056043	1
Methionine Degradation	9	0	316	718	2.27021	0.068874	1
D-gluconate and ketogluconates metabolism	20	11	305	696	1.471674	0.123357	1
Pyruvate Alanine Serine Interconversions	11	4	314	712	1.661978	0.144819	1
Arginine and Ornithine Degradation	8	2	317	717	1.808258	0.158823	1
Glycine and Serine Utilization	16	10	309	701	1.395624	0.19334	1
Pyruvate metabolism II: acetyl-CoA, acetogenesis from pyruvate	10	5	315	712	1.506257	0.215239	1

Mannose Metabolism	5	1	320	721	1.876467	0.229384	1
Fatty Acid Biosynthesis FASII	12	8	313	707	1.354893	0.260902	1
Chitin and N-acetylglucosamine utilization	12	8	313	707	1.354893	0.260902	1
Cobalamin synthesis	10	7	315	710	1.32553	0.305899	1
Isoprenoid Biosynthesis: Interconversions	2	0	323	725	2.242699	0.36403	1
Sucrose utilization	2	0	323	725	2.242699	0.36403	1
Formaldehyde assimilation: Ribulose monophosphate pathway	2	0	323	725	2.242699	0.36403	1
Molybdenum cofactor biosynthesis	2	0	323	725	2.242699	0.36403	1
Threonine degradation	3	1	322	723	1.683105	0.373122	1
5-FCL-like protein	15	14	310	698	1.164439	0.374644	1
Biotin biosynthesis	4	3	321	720	1.281441	0.456772	1
Creatine and Creatinine Degradation	4	3	321	720	1.281441	0.456772	1
Proline, 4-hydroxyprol	4	3	321	720	1.281441	0.456772	1

ine uptake and utilization							
Polyamine Metabolism	3	2	322	722	1.344976	0.471071	1
Pyruvate metabolism I: anaplerotic reactions, PEP	3	2	322	722	1.344976	0.471071	1
Flavodoxin	2	1	323	724	1.493721	0.490316	1
Polyprenyl Diphosphate Biosynthesis	2	1	323	724	1.493721	0.490316	1
Serine Biosynthesis	6	6	319	715	1.120558	0.499437	1
Alanine biosynthesis	6	6	319	715	1.120558	0.499437	1
Glutamate dehydrogenases	1	0	324	726	2.238866	0.522633	1
Arginine Deiminase Pathway	1	0	324	726	2.238866	0.522633	1
VC0266	1	0	324	726	2.238866	0.522633	1
Biotin biosynthesis Experimental	4	4	321	719	1.119807	0.535408	1
Xanthine dehydrogenase subunits	4	4	321	719	1.119807	0.535408	1
Glutamine, Glutamate, Aspartate and Asparagine Biosynthesis	10	12	315	705	1.0173	0.55028	1

Pyridoxin (Vitamin B6) Biosynthesis	7	9	318	711	0.97819	0.598234	1
Isoprenoinds for Quinones	2	2	323	723	1.119066	0.599474	1
Purine Utilization	6	8	319	713	0.957951	0.618886	1
Dihydroxyacetone kinases	1	1	324	725	1.118698	0.670389	1
Threonine and Homoserine Biosynthesis	6	9	319	712	0.892888	0.672144	1
Proline Synthesis	2	3	323	722	0.894214	0.690014	1
pyrimidine conversions	7	11	318	709	0.867172	0.696673	1
NAD and NADP cofactor biosynthesis global	5	8	320	714	0.858301	0.69859	1
Histidine Biosynthesis	5	8	320	714	0.858301	0.69859	1
Tryptophan synthesis	3	7	322	717	0.668246	0.818012	1
Folate biosynthesis cluster	2	5	323	720	0.637135	0.820597	1
Coenzyme A Biosynthesis	2	5	323	720	0.637135	0.820597	1
Methionine Biosynthesis	8	16	317	703	0.739401	0.822146	1
Branched-Chain Amino Acid Biosynthesis	4	10	321	713	0.634873	0.856795	1

Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilate and more.	3	8	322	716	0.606696	0.857558	1
Trehalose Biosynthesis	1	4	324	722	0.445963	0.891774	1
Thiamin biosynthesis	2	7	323	718	0.494263	0.899628	1
Nonmevalonate Branch of Isoprenoid Biosynthesis	1	5	324	721	0.371161	0.925401	1
One-carbon metabolism by tetrahydropterines	2	8	323	717	0.444248	0.925649	1
Dehydrogenase complexes	1	6	324	720	0.317735	0.948601	1
Common Pathway For Synthesis of Aromatic Compounds (DAHP synthase to chorismate)	1	6	324	720	0.317735	0.948601	1
Pentose phosphate pathway	1	7	324	719	0.277665	0.964602	1

Lysine Biosynthesis DAP Pathway, GJO scratch	1	8	324	718	0.246434	0.975632	1
Lysine Biosynthesis DAP Pathway	1	8	324	718	0.246434	0.975632	1
De Novo Purine Biosynthesis	2	12	323	713	0.315617	0.978663	1
Purine conversions	5	23	320	699	0.390351	0.989606	1
Heme and Siroheme Biosynthesis	0	11	325	716	0	1	1
riboflavin to FAD	0	4	325	723	0	1	1
Riboflavin, FMN and FAD metabolism	0	8	325	719	0	1	1
Glycine Biosynthesis	0	1	325	726	0	1	1
GMP synthase	0	2	325	725	0	1	1
Ammonia assimilation	0	5	325	722	0	1	1
Glutamine synthetases	0	1	325	726	0	1	1
Cysteine Biosynthesis	0	10	325	717	0	1	1
Methionine Salvage	0	6	325	721	0	1	1
Leucine Biosynthesis	0	5	325	722	0	1	1
Acetolactate synthase subunits	0	2	325	725	0	1	1

Acetoin, butanediol metabolism	0	3	325	724	0	1	1
Phenylalanine and Tyrosine Branches from Chorismate	0	4	325	723	0	1	1
Glycine cleavage system	0	3	325	724	0	1	1
Glycogen metabolism	0	3	325	724	0	1	1
Fermentations: Lactate	0	1	325	726	0	1	1
Glycerol and Glycerol-3-phosphate Uptake and Utilization	0	5	325	722	0	1	1
D-ribose utilization	0	1	325	726	0	1	1
Xylose utilization	0	2	325	725	0	1	1

Table S1.6. Subsystem enrichment analysis of prevalent species. (C) *Gluconobacter kondonii*.

Subsystems	Present		Absent		Odds ratio	Fisher's p-value	FDR corrected p-value
	Accessory	Core	Accessory	Core			
D-gluconate and ketogluconates metabolism	21	14	95	501	3.156881	0.000191	0.014542
Histidine Degradation	8	0	108	528	4.871778	0.002911	0.110621
Chitin and N-acetylglucosamine utilization	9	3	107	524	3.662794	0.005801	0.146955
Proline, 4-hydroxyproline uptake and utilization	6	1	110	529	4.109461	0.016329	0.310246
Purine Utilization	5	5	111	526	2.365404	0.109733	1
Pyruvate metabolism II: acetyl-CoA, acetogenesis from pyruvate	4	5	112	527	2.088397	0.186203	1
Sucrose utilization	2	1	114	533	3.109661	0.217778	1
Pyruvate metabolism I: anaplerotic r	2	2	114	532	2.329494	0.290296	1

eactions, PEP							
Citrate Metabolism, Transport, and Regulation	2	2	114	532	2.329494	0.290296	1
Dihydroxyacetone kinases	1	0	115	535	4.636434	0.324399	1
Formaldehyde assimilation: Ribulose monophosphate pathway	1	0	115	535	4.636434	0.324399	1
Heme and Siroheme Biosynthesis	4	9	112	523	1.435923	0.357608	1
Mannose Metabolism	2	3	114	531	1.860986	0.361961	1
Cysteine Biosynthesis	4	10	112	522	1.330995	0.402069	1
Methylglyoxal Metabolism	2	4	114	530	1.548519	0.430793	1
pyrimidine conversions	4	11	112	521	1.240039	0.445997	1
Thiamin biosynthesis	2	5	114	529	1.325189	0.495532	1
Xanthine dehydrogenase subunits	2	5	114	529	1.325189	0.495532	1
Coenzyme A Biosynthesis cluster	1	2	115	533	1.54375	0.544173	1
Polyamine Metabolism	1	2	115	533	1.54375	0.544173	1
5-FCL-like protein	3	10	113	523	1.067976	0.563363	1

Biotin biosynthesis	1	3	115	532	1.156213	0.625771	1
Glycine cleavage system	1	3	115	532	1.156213	0.625771	1
Methionine Biosynthesis	4	16	112	516	0.921561	0.644227	1
Serine Biosynthesis	2	8	114	526	0.922945	0.659726	1
Pyruvate Alanine Serine Interconversions	2	8	114	526	0.922945	0.659726	1
One-carbon metabolism by tetrahydropterines	1	4	115	531	0.923618	0.692866	1
Ammonia assimilation	1	4	115	531	0.923618	0.692866	1
Pyridoxin (Vitamin B6) Biosynthesis	2	10	114	524	0.766372	0.743641	1
Coenzyme A Biosynthesis	1	5	115	530	0.768403	0.748017	1
Glycine and Serine Utilization	3	17	113	516	0.685314	0.808107	1
Glutamine, Glutamate, Aspartate and Asparagine Biosynthesis	2	12	114	522	0.654512	0.809341	1
NAD and NADP cofactor biosynthesis global	1	7	115	528	0.574326	0.830561	1

Alanine biosynthesis	1	8	115	527	0.509607	0.861128	1
Tryptophan synthesis	1	8	115	527	0.509607	0.861128	1
Histidine Biosynthesis	1	9	115	526	0.457823	0.88622	1
Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilate and more.	1	9	115	526	0.457823	0.88622	1
Purine conversions	2	16	114	518	0.505308	0.897741	1
Threonine and Homoserine Biosynthesis	1	10	115	525	0.415435	0.906811	1
Branched-Chain Amino Acid Biosynthesis	1	10	115	525	0.415435	0.906811	1
Fatty Acid Biosynthesis FASII	1	13	115	522	0.324609	0.948906	1
Biotin biosynthesis Experimental	0	5	116	531	0	1	1
Nonmevalonate Branch of Isoprenoid Biosynthesis	0	5	116	531	0	1	1
riboflavin to FAD	0	4	116	532	0	1	1

Riboflavin, FMN and FAD metabolism	0	8	116	528	0	1	1
Flavodoxin	0	1	116	535	0	1	1
Folate biosynthesis cluster	0	7	116	529	0	1	1
Glycine Biosynthesis	0	1	116	535	0	1	1
De Novo Purine Biosynthesis	0	12	116	524	0	1	1
Dehydrogenase complexes	0	6	116	530	0	1	1
Pentose phosphate pathway	0	8	116	528	0	1	1
GMP synthase	0	2	116	534	0	1	1
Isoprenoinds for Quinones	0	3	116	533	0	1	1
Polyprenyl Diphosphate Biosynthesis	0	2	116	534	0	1	1
Isoprenoid Biosynthesis: Interconversions	0	1	116	535	0	1	1
Glutamine synthetases	0	1	116	535	0	1	1
Arginine and Ornithine Degradation	0	4	116	532	0	1	1
Methionine Degradation	0	7	116	529	0	1	1

Threonine degradation	0	3	116	533	0	1	1
Lysine Biosynthesis DAP Pathway, GJO scratch	0	8	116	528	0	1	1
Lysine Biosynthesis DAP Pathway	0	8	116	528	0	1	1
Methionine Salvage	0	6	116	530	0	1	1
Creatine and Creatinine Degradation	0	3	116	533	0	1	1
Leucine Biosynthesis	0	5	116	531	0	1	1
Acetolactate synthase subunits	0	2	116	534	0	1	1
Acetoin, butanediol metabolism	0	3	116	533	0	1	1
Common Pathway For Synthesis of Aromatic Compounds (DAHP synthase to chorismate)	0	6	116	530	0	1	1
Phenylalanine and Tyrosine Branches from Chorismate	0	4	116	532	0	1	1
Proline Synthesis	0	4	116	532	0	1	1

Trehalose Biosynthesis	0	2	116	534	0	1	1
Lactate utilization	0	5	116	531	0	1	1
Fermentations: Lactate	0	1	116	535	0	1	1
Glycerol and Glycerol-3-phosphate Uptake and Utilization	0	5	116	531	0	1	1
VC0266	0	1	116	535	0	1	1
D-ribose utilization	0	1	116	535	0	1	1
Xylose utilization	0	2	116	534	0	1	1

Table S1.6. Subsystem enrichment analysis of prevalent species. (D) *Lactobacillus brevis*.

	Present		Absent				
Subsystems	**Accessory**	**Core**	**Accessory**	**Core**	**Odds ratio**	**Fisher's p-value**	**FDR corrected p-value**
Citrate Metabolism, Transport, and Regulation	5	0	22	399	17.78866	0.000135	0.009031
Glycolate, glyoxylate interconversions	3	3	24	398	8.199947	0.014339	0.415199
D-Galacturonate and D-Glucuronate Utilization	5	18	22	381	3.746413	0.023667	0.415199
Xylose utilization	4	11	23	389	4.482147	0.024788	0.415199
Pyruvate metabolism I: anaplerotic reactions, PEP	2	7	25	395	3.494042	0.146586	1
Trehalose Biosynthesis	1	1	26	402	7.646196	0.176792	1
Chitin and N-acetylglucosamine utilization	1	3	26	400	3.824602	0.277493	1
D-ribose utilization	1	3	26	400	3.824602	0.277493	1

L-Arabinose utilization	1	4	26	399	3.056405	0.323286	1
Glycerate metabolism	1	5	26	398	2.543036	0.366277	1
Fatty Acid Biosynthesis FASII	1	12	26	391	1.156389	0.601532	1
pyrimidine conversions	1	16	26	387	0.875812	0.69544	1
Purine conversions	1	26	26	377	0.537649	0.846297	1
Biotin biosynthesis Experimental	0	1	27	403	0	1	1
riboflavin to FAD	0	4	27	400	0	1	1
Riboflavin, FMN and FAD metabolism	0	9	27	395	0	1	1
Flavodoxin	0	4	27	400	0	1	1
Pyridoxin(Vitamin B6) Degradation Pathway	0	1	27	403	0	1	1
NAD and NADP cofactor biosynthesis global	0	7	27	397	0	1	1
Folate Biosynthesis	0	5	27	399	0	1	1
5-FCL-like protein	0	15	27	389	0	1	1
One-carbon metabolism by tetrahydropterines	0	4	27	400	0	1	1

Pyruvate metabolism II: acetyl-CoA, acetogenesis from pyruvate	0	11	27	393	0	1	1
Fermentations: Lactate	0	10	27	394	0	1	1
Glycine Biosynthesis	0	1	27	403	0	1	1
Serine Biosynthesis	0	1	27	403	0	1	1
Dehydrogenase complexes	0	3	27	401	0	1	1
Lipoic acid metabolism	0	3	27	401	0	1	1
Coenzyme A Biosynthesis	0	6	27	398	0	1	1
Creatine and Creatinine Degradation	0	1	27	403	0	1	1
Deoxyribose and Deoxynucleoside Catabolism	0	9	27	395	0	1	1
De Novo Pyrimidine Synthesis	0	7	27	397	0	1	1
Xanthine Metabolism in Bacteria	0	3	27	401	0	1	1
Purine Utilization	0	6	27	398	0	1	1
GMP synthase	0	2	27	402	0	1	1

Triacylglycerol metabolism	0	1	27	403	0	1	1
Fatty acid metabolism cluster	0	2	27	402	0	1	1
Mevalonate Branch of Isoprenoid Biosynthesis	0	6	27	398	0	1	1
Isoprenoid Biosynthesis	0	9	27	395	0	1	1
Acetyl-CoA fermentation to Butyrate	0	5	27	399	0	1	1
Isoprenoinds for Quinones	0	3	27	401	0	1	1
Polyprenyl Diphosphate Biosynthesis	0	3	27	401	0	1	1
Isoprenoid Biosynthesis: Interconversions	0	2	27	402	0	1	1
Glutamine, Glutamate, Aspartate and Asparagine Biosynthesis	0	8	27	396	0	1	1
Glutamine synthetases	0	1	27	403	0	1	1
Glutamate and Aspartate uptake in Bacteria	0	2	27	402	0	1	1

Arginine Deiminase Pathway	0	11	27	393	0	1	1
Polyamine Metabolism	0	11	27	393	0	1	1
Arginine and Ornithine Degradation	0	13	27	391	0	1	1
Methionine Degradation	0	6	27	398	0	1	1
Methionine Biosynthesis	0	7	27	397	0	1	1
Common Pathway For Synthesis of Aromatic Compounds (DAHP synthase to chorismate)	0	1	27	403	0	1	1
Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilate and more.	0	1	27	403	0	1	1
Proline Synthesis	0	1	27	403	0	1	1
Alanine biosynthesis	0	2	27	402	0	1	1
Pyruvate Alanine Serine Interconversions	0	4	27	400	0	1	1

Pentose phosphate pathway	0	12	27	392	0	1	1
D-gluconate and ketogluconates metabolism	0	6	27	398	0	1	1
Trehalose Uptake and Utilization	0	4	27	400	0	1	1
Lactose and Galactose Uptake and Utilization	0	8	27	396	0	1	1
D-galactarate, D-glucarate and D-glycerate catabolism - gjo	0	4	27	400	0	1	1
D-galactarate, D-glucarate and D-glycerate catabolism	0	4	27	400	0	1	1
Alpha-acetolactate operon	0	2	27	402	0	1	1
Acetoin, butanediol metabolism	0	5	27	399	0	1	1
Glycerol and Glycerol-3-phosphate U	0	9	27	395	0	1	1

ptake and U tilization							
VC0266	0	1	27	403	0	1	1
Mannose Me tabolism	0	1	27	403	0	1	1

Table S1.6. Subsystem enrichment analysis of prevalent species. (E) *Lactobacillus plantarum*.

	Present		Absent				
Subsystems	**Accessory**	**Core**	**Accessory**	**Core**	**Odds ratio**	**Fisher's p-value**	**FDR corrected p-value**
Trehalose Uptake and Utilization	21	0	204	711	3.47937	0.000117	0.009502
Fructose utilization	8	0	217	724	3.331103	0.018044	0.616989
Inositol catabolism	6	0	219	726	3.309842	0.040298	0.616989
Xylose utilization	6	0	219	726	3.309842	0.040298	0.616989
Glycogen metabolism	6	1	219	725	2.833771	0.060501	0.616989
Nitrate and nitrite ammonification	5	0	220	727	3.299355	0.060937	0.616989
Denitrifying reductase gene clusters	5	0	220	727	3.299355	0.060937	0.616989
L-Arabinose utilization	5	0	220	727	3.299355	0.060937	0.616989
Molybdenum cofactor biosynthesis	6	2	219	724	2.476619	0.085714	0.771389
Beta-Glucoside Metabolism	12	11	213	709	1.735548	0.095233	0.771389
Common Pathway For Synthesis of Aromatic Compounds (D	7	5	218	720	1.925192	0.134516	0.990529

AHP synthase to chorismate)							
D-ribose utilization	5	4	220	723	1.8244	0.214785	1
Pentose phosphate pathway	9	11	216	712	1.482663	0.222409	1
Tryptophan synthesis	5	5	220	722	1.639956	0.264264	1
Methionine Degradation	7	10	218	715	1.350057	0.325985	1
Flavodoxin	2	1	223	729	2.17731	0.335737	1
Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilate and more.	5	7	220	720	1.36316	0.36881	1
Trehalose Biosynthesis	1	0	224	731	3.258341	0.415131	1
Deoxyribose and Deoxynucleoside Catabolism	1	0	224	731	3.258341	0.415131	1
Sucrose utilization	3	4	222	725	1.39908	0.4295	1
Mannose Metabolism	2	2	223	728	1.631348	0.430629	1
Glycerol and Glycerol-3-phosphate U	7	13	218	712	1.142929	0.45672	1

ptake and U tilization							
Lactose and Galactose U ptake and U tilization	6	11	219	715	1.15209	0.46469	1
Pyruvate me tabolism I: a naplerotic r eactions, PE P	4	7	221	721	1.186158	0.484692	1
Methionine Biosynthesis	11	24	214	697	1.023607	0.533265	1
Serine-glyoxylate c ycle	4	9	221	719	1.001047	0.593891	1
Serine Biosy nthesis	2	4	223	726	1.085118	0.595622	1
D-gluconate a nd ketogluc onates meta bolism	2	4	223	726	1.085118	0.595622	1
NAD and NA DP cofactor biosynthesis global	3	7	222	722	0.975693	0.622094	1
Pyruvate Ala nine Serine I nterconversi ons	3	7	222	722	0.975693	0.622094	1
Biotin biosy nthesis	1	2	224	729	1.084738	0.658369	1
Fermentatio ns: Mixed ac id	7	19	218	706	0.872035	0.691076	1
Pyruvate me tabolism II: acetyl-	5	14	220	713	0.853009	0.700466	1

CoA, acetogenesis from pyruvate							
5-FCL-like protein	5	15	220	712	0.809265	0.736678	1
Glycine and Serine Utilization	5	16	220	711	0.769685	0.769612	1
Glutamine, Glutamate, Aspartate and Asparagine Biosynthesis	5	16	220	711	0.769685	0.769612	1
Pyridoxin (Vitamin B6) Biosynthesis	2	7	223	723	0.720714	0.771935	1
Isoprenoinds for Quinones	1	4	224	727	0.649381	0.800706	1
Polyprenyl Diphosphate Biosynthesis	1	5	224	726	0.540492	0.847858	1
Chitin and N-acetylglucosamine utilization	1	5	224	726	0.540492	0.847858	1
Citrate Metabolism, Transport, and Regulation	1	5	224	726	0.540492	0.847858	1
De Novo Pyrimidine Synthesis	3	13	222	716	0.605028	0.861321	1
Purine conversions	6	23	219	703	0.664422	0.868378	1
Thiamin biosynthesis	1	7	224	724	0.404295	0.91142	1

Coenzyme A Biosynthesis	1	7	224	724	0.404295	0.91142	1
Fatty Acid Biosynthesis FASII	3	16	222	713	0.50744	0.921079	1
Histidine Biosynthesis	1	8	224	723	0.358903	0.932444	1
Fermentations: Lactate	2	16	223	714	0.356025	0.967711	1
pyrimidine conversions	1	11	224	720	0.268101	0.970091	1
Riboflavin, FMN and FAD metabolism	1	12	224	719	0.247143	0.97722	1
Folate Biosynthesis	1	13	224	718	0.229178	0.982655	1
De Novo Purine Biosynthesis	1	13	224	718	0.229178	0.982655	1
Threonine and Homoserine Biosynthesis	1	15	224	716	0.199983	0.989955	1
Heme and Siroheme Biosynthesis	0	2	225	730	0	1	1
riboflavin to FAD	0	6	225	726	0	1	1
Pyridoxin(Vitamin B6) Degradation Pathway	0	1	225	731	0	1	1
Folate biosynthesis cluster	0	6	225	726	0	1	1
One-carbon metabolism by te	0	5	225	727	0	1	1

trahydropterines							
Glycine Biosynthesis	0	1	225	731	0	1	1
Dehydrogenase complexes	0	3	225	729	0	1	1
Lipoic acid metabolism	0	3	225	729	0	1	1
Xanthine Metabolism in Bacteria	0	2	225	730	0	1	1
Purine Utilization	0	4	225	728	0	1	1
GMP synthase	0	2	225	730	0	1	1
Mevalonate Branch of Isoprenoid Biosynthesis	0	5	225	727	0	1	1
Isoprenoid Biosynthesis: Interconversions	0	2	225	730	0	1	1
Glutamine synthetases	0	1	225	731	0	1	1
Glutamate dehydrogenases	0	1	225	731	0	1	1
Proline Synthesis	0	4	225	728	0	1	1
Arginine Deiminase Pathway	0	3	225	729	0	1	1
S-methylmethionine	0	2	225	730	0	1	1
Lysine Biosynthesis DAP	0	10	225	722	0	1	1

Pathway, GJO scratch							
Lysine Biosynthesis DAP Pathway	0	10	225	722	0	1	1
Phenylalanine and Tyrosine Branches from Chorismate	0	1	225	731	0	1	1
Alanine biosynthesis	0	3	225	729	0	1	1
Glycerate metabolism	0	5	225	727	0	1	1
Dihydroxyacetone kinases	0	5	225	727	0	1	1
Glycolate, glyoxylate interconversions	0	4	225	728	0	1	1
Alpha-acetolactate operon	0	2	225	730	0	1	1
Acetoin, butanediol metabolism	0	2	225	730	0	1	1
VC0266	0	1	225	731	0	1	1

Table S1.6. Subsystem enrichment analysis of prevalent species. (F) *Providencia rettgeri*.

	Present		Absent				
Subsystems	**Accessory**	**Core**	**Accessory**	**Core**	**Odds ratio**	**Fisher's p-value**	**FDR corrected p-value**
Acetyl-CoA fermentation to Butyrate	11	7	73	851	7.095209	1.22E-05	0.00128
De Novo Purine Biosynthesis	8	3	76	858	8.173156	8.90E-05	0.004671
Fatty acid metabolism cluster	9	9	75	851	5.654388	0.000263	0.009221
Mannose Metabolism	6	1	78	862	9.424761	0.000408	0.010721
Carbon storage regulator	3	0	81	866	10.63303	0.010877	0.190355
Aromatic amino acid interconversions with aryl acids	3	0	81	866	10.63303	0.010877	0.190355
Indole-pyruvate oxidoreductase complex	2	0	82	867	10.51584	0.040977	0.614653
Butanol Biosynthesis	3	7	81	859	3.175453	0.099063	1
Serine Biosynthesis	2	5	82	862	2.99842	0.184446	1
Proline, 4-hydroxyproline uptake a	2	7	82	860	2.327669	0.251982	1

nd utilizatio n							
Glutamine, Glutamate, Aspartate a nd Asparagi ne Biosynth esis	3	14	81	852	1.854595	0.255283	1
Glycerol and Glycerol-3- phosphate U ptake and U tilization	2	10	82	857	1.740602	0.353784	1
Tryptophan synthesis	2	11	82	856	1.605029	0.386842	1
Flavodoxin	1	4	83	864	2.079783	0.426019	1
Chorismate: Intermediat e for synthe sis of Trypto phan, PAPA antibiotics, PABA, 3- hydroxyanth ranilate and more.	2	13	82	854	1.388066	0.450625	1
Glycolate, gl yoxylate int erconversio ns	1	5	83	863	1.73168	0.476931	1
Menaquinon e and Phyllo quinone Bio synthesis	1	7	83	861	1.296293	0.565744	1
Menaquinon e and Phyllo quinone Bio	1	7	83	861	1.296293	0.565744	1

synthesis -- gjo							
Purine Utilization	1	7	83	861	1.296293	0.565744	1
Urease subunits	1	7	83	861	1.296293	0.565744	1
Urea decomposition	1	7	83	861	1.296293	0.565744	1
Glyoxylate bypass	1	7	83	861	1.296293	0.565744	1
Pyruvate metabolism I: anaplerotic reactions, PEP	1	7	83	861	1.296293	0.565744	1
Biotin biosynthesis	1	9	83	859	1.03491	0.639626	1
Coenzyme A Biosynthesis	1	9	83	859	1.03491	0.639626	1
Glycine and Serine Utilization	2	21	82	846	0.897225	0.663533	1
Threonine and Homoserine Biosynthesis	1	10	83	858	0.939816	0.671761	1
Heme and Siroheme Biosynthesis	1	11	83	857	0.860556	0.701062	1
Common Pathway For Synthesis of Aromatic Compounds (DAHP synthase to chorismate)	1	11	83	857	0.860556	0.701062	1

Arginine and Ornithine Degradation	1	12	83	856	0.793477	0.727775	1
Pyruvate metabolism II: acetyl-CoA, acetogenesis from pyruvate	1	13	83	855	0.736033	0.752128	1
Chorismate Synthesis	1	13	83	855	0.736033	0.752128	1
TCA Cycle	1	14	83	854	0.68619	0.774325	1
5-FCL-like protein	1	17	83	851	0.569882	0.829794	1
Methionine Degradation	1	17	83	851	0.569882	0.829794	1
Cysteine Biosynthesis	1	17	83	851	0.569882	0.829794	1
pyrimidine conversions	1	20	83	848	0.486795	0.871752	1
Purine conversions	1	25	83	843	0.390908	0.920153	1
Methionine Biosynthesis	1	30	83	838	0.32594	0.950422	1
Biotin biosynthesis Experimental	0	4	84	865	0	1	1
Thiamin biosynthesis	0	9	84	860	0	1	1
Pyridoxin (Vitamin B6) Biosynthesis	0	11	84	858	0	1	1
Nonmevalonate Branch of Isoprenoid Biosynthesis	0	5	84	864	0	1	1

riboflavin to FAD	0	4	84	865	0	1	1
Riboflavin, FMN and FAD metabolism	0	8	84	861	0	1	1
NAD and NADP cofactor biosynthesis global	0	18	84	851	0	1	1
Deoxyribose and Deoxynucleoside Catabolism	0	11	84	858	0	1	1
Folate biosynthesis cluster	0	8	84	861	0	1	1
Folate Biosynthesis	0	12	84	857	0	1	1
Coenzyme A Biosynthesis cluster	0	3	84	866	0	1	1
One-carbon metabolism by tetrahydropterines	0	5	84	864	0	1	1
Fermentations: Mixed acid	0	10	84	859	0	1	1
Fermentations: Lactate	0	6	84	863	0	1	1
Glycine Biosynthesis	0	3	84	866	0	1	1
Glycine cleavage system	0	7	84	862	0	1	1
Dehydrogenase complexes	0	5	84	864	0	1	1

Lipoic acid metabolism	0	2	84	867	0	1	1
Creatine and Creatinine Degradation	0	2	84	867	0	1	1
Pentose phosphate pathway	0	8	84	861	0	1	1
Xanthine Metabolism in Bacteria	0	3	84	866	0	1	1
Purine nucleotide synthesis regulator	0	1	84	868	0	1	1
GMP synthase	0	2	84	867	0	1	1
Triacylglycerol metabolism	0	2	84	867	0	1	1
Fatty Acid Biosynthesis FASII	0	11	84	858	0	1	1
Acyl-CoA thioesterase II	0	1	84	868	0	1	1
Isoprenoinds for Quinones	0	4	84	865	0	1	1
Polyprenyl Diphosphate Biosynthesis	0	3	84	866	0	1	1
Isoprenoid Biosynthesis: Interconversions	0	1	84	868	0	1	1
Nitrate and nitrite ammonification	0	15	84	854	0	1	1

Denitrifying reductase gene clusters	0	4	84	865	0	1	1
Ammonia assimilation	0	6	84	863	0	1	1
Glutamine synthetases	0	1	84	868	0	1	1
Glutamate dehydrogenases	0	1	84	868	0	1	1
Proline Synthesis	0	5	84	864	0	1	1
Branched-Chain Amino Acid Biosynthesis	0	14	84	855	0	1	1
Histidine Degradation	0	9	84	860	0	1	1
Histidine Biosynthesis	0	8	84	861	0	1	1
Arginine Deiminase Pathway	0	3	84	866	0	1	1
Polyamine Metabolism	0	8	84	861	0	1	1
Urea carboxylase and Allophanate hydrolase cluster	0	4	84	865	0	1	1
Lysine Biosynthesis DAP Pathway, GJO scratch	0	13	84	856	0	1	1
Lysine Biosynthesis DAP Pathway	0	13	84	856	0	1	1
Threonine degradation	0	2	84	867	0	1	1

Leucine Biosynthesis	0	5	84	864	0	1	1
Alanine biosynthesis	0	8	84	861	0	1	1
Pyruvate Alanine Serine Interconversions	0	12	84	857	0	1	1
Acetolactate synthase subunits	0	5	84	864	0	1	1
Acetoin, butanediol metabolism	0	6	84	863	0	1	1
Phenylalanine and Tyrosine Branches from Chorismate	0	7	84	862	0	1	1
Aromatic amino acid degradation	0	10	84	859	0	1	1
Glycerate metabolism	0	6	84	863	0	1	1
D-galactarate, D-glucarate and D-glycerate catabolism - gjo	0	4	84	865	0	1	1
D-galactarate, D-glucarate and D-glycerate catabolism	0	4	84	865	0	1	1

Methylglyoxal Metabolism	0	7	84	862	0	1	1
Dihydroxyacetone kinases	0	2	84	867	0	1	1
D-gluconate and ketogluconates metabolism	0	5	84	864	0	1	1
Chitin and N-acetylglucosamine utilization	0	4	84	865	0	1	1
Lactose and Galactose Uptake and Utilization	0	6	84	863	0	1	1
Methylcitrate cycle	0	4	84	865	0	1	1
Citrate Metabolism, Transport, and Regulation	0	6	84	863	0	1	1
2-Ketogluconate Utilization	0	4	84	865	0	1	1
D-ribose utilization	0	3	84	866	0	1	1
L-ascorbate utilization (and related gene clusters)	0	3	84	866	0	1	1

D-galactonate catabolism	0	3	84	866	0	1	1
Xylose utilization	0	2	84	867	0	1	1

Table S1.6. Subsystem enrichment analysis of prevalent species. (G) *Tatumella* sp.

	Present		Absent				
Subsystems	**Accessory**	**Core**	**Accessory**	**Core**	**Odds ratio**	**Fisher's p-value**	**FDR corrected p-value**
Mannose Metabolism	3	1	31	748	17.87454	0.002303	0.207283
Conserved cluster around inner membrane protein gene yghQ, probably involved in polysaccharide biosynthesis	3	5	31	744	8.934928	0.009631	0.433413
Lactose utilization	1	0	33	751	22.43571	0.084696	1
Nitrate and nitrite ammonification	2	8	32	742	4.619457	0.091385	1
Heme and Siroheme Biosynthesis	2	13	32	737	3.06362	0.1648	1
Glutamine, Glutamate, Aspartate and Asparagine Biosynthesis	2	15	32	735	2.696915	0.196563	1
Biotin biosynthesis Experimental	1	3	33	748	5.639076	0.198828	1

Xanthine Metabolism in Bacteria	1	3	33	748	5.639076	0.198828	1
Acetolactate synthase subunits	1	4	33	747	4.510231	0.233706	1
Lactose and Galactose Uptake and Utilization	1	5	33	746	3.755621	0.267109	1
Purine Utilization	1	6	33	745	3.216886	0.299096	1
Biotin biosynthesis	1	7	33	744	2.812218	0.329727	1
Lysine degradation	1	7	33	744	2.812218	0.329727	1
Threonine and Homoserine Biosynthesis	1	9	33	742	2.245168	0.38714	1
Arginine Deiminase Pathway	1	9	33	742	2.245168	0.38714	1
Pyruvate Alanine Serine Interconversions	1	9	33	742	2.245168	0.38714	1
Acetoin, butanediol metabolism	1	9	33	742	2.245168	0.38714	1
Thiamin biosynthesis	1	10	33	741	2.038768	0.414026	1
Ammonia assimilation	1	11	33	740	1.866789	0.439767	1
Pyridoxin (Vitamin B6) Biosynthesis	1	13	33	738	1.596215	0.487997	1
TCA Cycle	1	13	33	738	1.596215	0.487997	1

Glycine and Serine Utilization	1	15	33	736	1.393244	0.532186	1
Branched-Chain Amino Acid Biosynthesis	1	15	33	736	1.393244	0.532186	1
Polyamine Metabolism	1	15	33	736	1.393244	0.532186	1
5-FCL-like protein	1	25	33	726	0.846316	0.703177	1
D-gluconate and ketogluconates metabolism	1	27	33	724	0.783759	0.729191	1
Arginine and Ornithine Degradation	1	28	33	723	0.755712	0.741354	1
Fatty acid metabolism cluster	0	3	34	749	0	1	1
Butanol Biosynthesis	0	10	34	742	0	1	1
Nonmevalonate Branch of Isoprenoid Biosynthesis	0	5	34	747	0	1	1
riboflavin to FAD	0	4	34	748	0	1	1
Riboflavin, FMN and FAD metabolism	0	8	34	744	0	1	1
Flavodoxin	0	4	34	748	0	1	1
Serine Biosynthesis	0	8	34	744	0	1	1

NAD and NADP cofactor biosynthesis global	0	13	34	739	0	1	1
Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilate and more.	0	14	34	738	0	1	1
Folate biosynthesis cluster	0	8	34	744	0	1	1
Folate Biosynthesis	0	12	34	740	0	1	1
Coenzyme A Biosynthesis cluster	0	3	34	749	0	1	1
Coenzyme A Biosynthesis	0	11	34	741	0	1	1
Purine conversions	0	20	34	732	0	1	1
Tryptophan synthesis	0	11	34	741	0	1	1
One-carbon metabolism by tetrahydropterines	0	8	34	744	0	1	1
pyrimidine conversions	0	14	34	738	0	1	1
Pyruvate metabolism II: acetyl-	0	9	34	743	0	1	1

CoA, acetogenesis from pyruvate							
Fermentations: Mixed acid	0	15	34	737	0	1	1
Fermentations: Lactate	0	5	34	747	0	1	1
Glycine Biosynthesis	0	1	34	751	0	1	1
De Novo Purine Biosynthesis	0	11	34	741	0	1	1
Methionine Biosynthesis	0	21	34	731	0	1	1
Glycine cleavage system	0	6	34	746	0	1	1
Methionine Degradation	0	9	34	743	0	1	1
Dehydrogenase complexes	0	5	34	747	0	1	1
Lipoic acid metabolism	0	1	34	751	0	1	1
Deoxyribose and Deoxynucleoside Catabolism	0	8	34	744	0	1	1
Pentose phosphate pathway	0	10	34	742	0	1	1
Purine nucleotide synthesis regulator	0	1	34	751	0	1	1
GMP synthase	0	2	34	750	0	1	1
Triacylglycerol metabolism	0	1	34	751	0	1	1

Fatty Acid Biosynthesis FASII	0	12	34	740	0	1	1
Acyl-CoA thioesterase II	0	1	34	751	0	1	1
Isoprenoinds for Quinones	0	4	34	748	0	1	1
Polyprenyl Diphosphate Biosynthesis	0	3	34	749	0	1	1
Isoprenoid Biosynthesis: Interconversions	0	1	34	751	0	1	1
Denitrifying reductase gene clusters	0	4	34	748	0	1	1
Glutamine synthetases	0	1	34	751	0	1	1
Histidine Biosynthesis	0	7	34	745	0	1	1
Putrescine utilization pathways	0	4	34	748	0	1	1
Proline, 4-hydroxyproline uptake and utilization	0	7	34	745	0	1	1
S-methylmethionine	0	2	34	750	0	1	1
Cysteine Biosynthesis	0	16	34	736	0	1	1
Lysine Biosynthesis DAP	0	12	34	740	0	1	1

Pathway, GJO scratch							
Lysine Biosynthesis DAP Pathway	0	12	34	740	0	1	1
Leucine Biosynthesis	0	6	34	746	0	1	1
Alanine biosynthesis	0	8	34	744	0	1	1
DAP (1,3-diaminopropane) production	0	2	34	750	0	1	1
Common Pathway For Synthesis of Aromatic Compounds (DAHP synthase to chorismate)	0	13	34	739	0	1	1
Phenylalanine and Tyrosine Branches from Chorismate	0	6	34	746	0	1	1
Proline Synthesis	0	4	34	748	0	1	1
Glycerate metabolism	0	4	34	748	0	1	1
Methylglyoxal Metabolism	0	7	34	745	0	1	1
Dihydroxyacetone kinases	0	1	34	751	0	1	1
Pyruvate metabolism I: anaplerotic r	0	6	34	746	0	1	1

eactions, PEP							
Glycolate, glyoxylate interconversions	0	1	34	751	0	1	1
Chitin and N-acetylglucosamine utilization	0	4	34	748	0	1	1
Citrate Metabolism, Transport, and Regulation	0	4	34	748	0	1	1
Alpha-acetolactate operon	0	3	34	749	0	1	1
Glycerol and Glycerol-3-phosphate Uptake and Utilization	0	11	34	741	0	1	1
D-ribose utilization	0	2	34	750	0	1	1
2-Ketogluconate Utilization	0	4	34	748	0	1	1

Table S1.6. Subsystem enrichment analysis of prevalent species. (H) Gene functions among top subsystems and distribution across pangenomes.

Species	Metabolic functions	Metabolic distribution among prevalent taxa						
Gk	**Subsystem: D-gluconate and ketogluconates metabolism (OR = 3.2, p = 0.0002, adjusted p = 0.01)**							
	Ortholog functions	Ath	Gc	Gk	LAb	LApl	PRr	T
	2-Keto-D-gluconate dehydrogenase (EC 1.1.99.4), membrane-bound, flavoprotein							
	2-Keto-D-gluconate dehydrogenase (EC 1.1.99.4), membrane-bound, cytochrome c							
	2-Keto-D-gluconate dehydrogenase (EC 1.1.99.4), membrane-bound,							

	gamma subunit							
	Gluconate 2-dehydrogenase (EC 1.1.99.3), membrane-bound, gamma subunit							
	Gluconate 2-dehydrogenase (EC 1.1.99.3), membrane-bound, flavoprotein							
	Glucose 1-dehydrogenase (EC 1.1.1.47)							
LAb	**Subsystem: Citrate metabolism, transport, and regulation (OR = 17.8, p = 0.0001, adjusted p = 0.009)**							
	Ortholog functions	Ath	Gc	Gk	LAb	LApl	PRr	T
	[Citrate [pro-3S]-lyase] ligase (EC 6.2.1.22)							
	Apo-citrate lyase phosphoribosyl-dephospho-CoA							

	transferase (EC 2.7.7.61)							
	Citrate lyase alpha chain (EC 4.1.3.6)							
	Citrate lyase beta chain (EC 4.1.3.6)							
	Citrate lyase transcriptional regulator CitI							
	Subsystem: Trehalose uptake and utilization (OR = 3.5, p = 0.0002, adjusted p = 0.01)							
	Ortholog functions	Ath	Gc	Gk	LAb	LApl	PRr	T
LApl	Trehalose-6-phosphate hydrolase (EC 3.2.1.93)							
	PTS system, trehalose-specific IIA component (EC 2.7.1.69)							
	PTS system, trehalose-specific IIB component (EC 2.7.1.69)							
	PTS system, trehalose-							

	specific IIC component (EC 2.7.1.69)							
	Trehalose operon transcriptional repressor							
	Beta-phosphoglucomutase (EC 5.4.2.6)							
	PTS system, glucose-specific IIB component (EC 2.7.1.69)							
	Trehalose 6-phosphate phosphorylase (EC 2.4.1.216)							
	Trehalose phosphorylase (EC 2.4.1.64)							
PRr	**Subsystem: Mannose metabolism (OR = 9.4, p = 0.0004, adjusted p = 0.01)**							
	Ortholog functions	Ath	Gc	Gk	LAb	LApl	PRr	T
	Phosphomannomutase (EC 5.4.2.8)							
	Mannose−6−phosphate i							

	somerase (EC 5.3.1.8)							
	Subsystem: De Novo Purine Biosynthesis (OR = 8.2, p = 8.9e-5, adjusted p = 0.005)							
	Ortholog functions	Ath	Gc	Gk	LAb	LApl	PRr	T
	IMP cyclohydrolase (EC 3.5.4.10)							
	Phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3)							
	Phosphoribosylamine--glycine ligase (EC 6.3.4.13)							
	Amidophosphoribosyltransferase (EC 2.4.2.14)							
	Phosphoribosylaminoimidazole-succinocarboxamide synthase (EC 6.3.2.6)							
	Phosphoribosylformylgly							

	cinamidine cyclo-ligase (EC 6.3.3.1)							
	Phosphoribosylformylglycinamidine synthase, synthetase subunit (EC 6.3.5.3)							
	Phosphoribosylformylglycinamidine synthase, glutamine amidotransferase subunit (EC 6.3.5.3)							
	Subsystem: Acetyl-CoA fermentation to Butyrate (OR = 7.1, p = 1.2e-5, adjusted p = 0.001)							
	Ortholog functions	Ath	Gc	Gk	LAb	LApl	PRr	T
	3-hydroxybutyryl-CoA epimerase (EC 5.1.2.3)							
	Enoyl-CoA hydratase (EC 4.2.1.17)							

	3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35)							
	Electron transfer flavoprotein-ubiquinone oxidoreductase (EC 1.5.5.1)							
	Subsystem: Fatty acid metabolism cluster (OR = 5.7, p = 0.0003, adjusted p = 0.009)							
	Ortholog functions	Ath	Gc	Gk	LAb	LApl	PRr	T
	3-hydroxybutyryl-CoA epimerase (EC 5.1.2.3)							
	Enoyl-CoA hydratase (EC 4.2.1.17)							
	3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35)							

APPENDIX B

SUPPLEMENTAL MATERIAL FOR CHAPTER 3

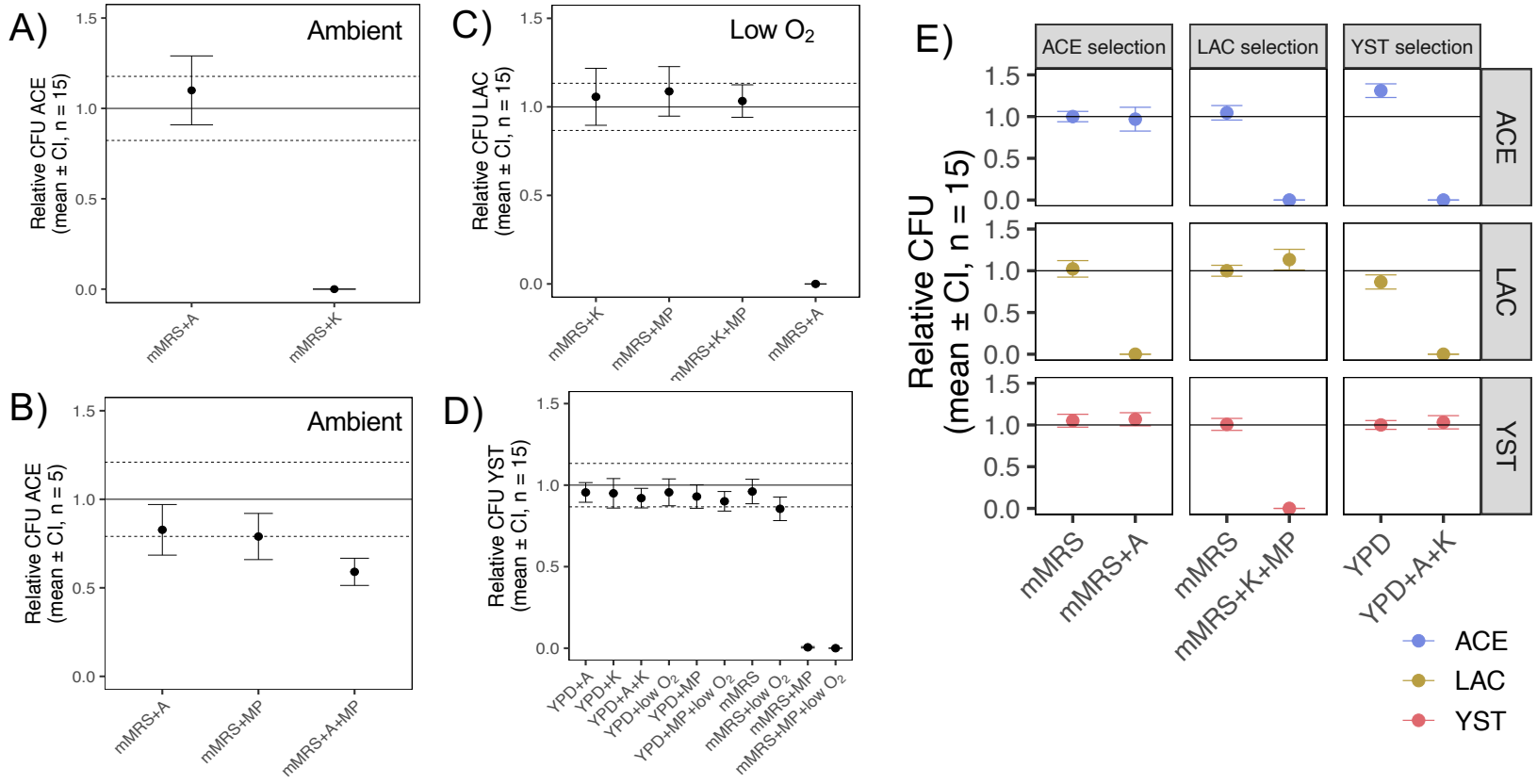


Figure S2.1. Method development for microbial selection media. Assessment of antimicrobials selected to inhibit (A & B) ACE, (C) LAC, and (D) YST using mono-cultures. (E) Validation of media and antimicrobials used to select for ACE, LAC, and YST using tri-culture community to calculate microbial abundance. Data represents five replicates from three independent experimental replicates, except for panel (B), which had one experimental replicate. CFU was normalized to CFU of microbe grown under optimal conditions (see Table 1). The mean and 95% confidence intervals of positive control are shown as lines for each panel (solid and dashed, respectively). Non-overlapping confidence intervals indicate significant difference between treatments. K = kanamycin. A = ampicillin. MP = methylparaben.

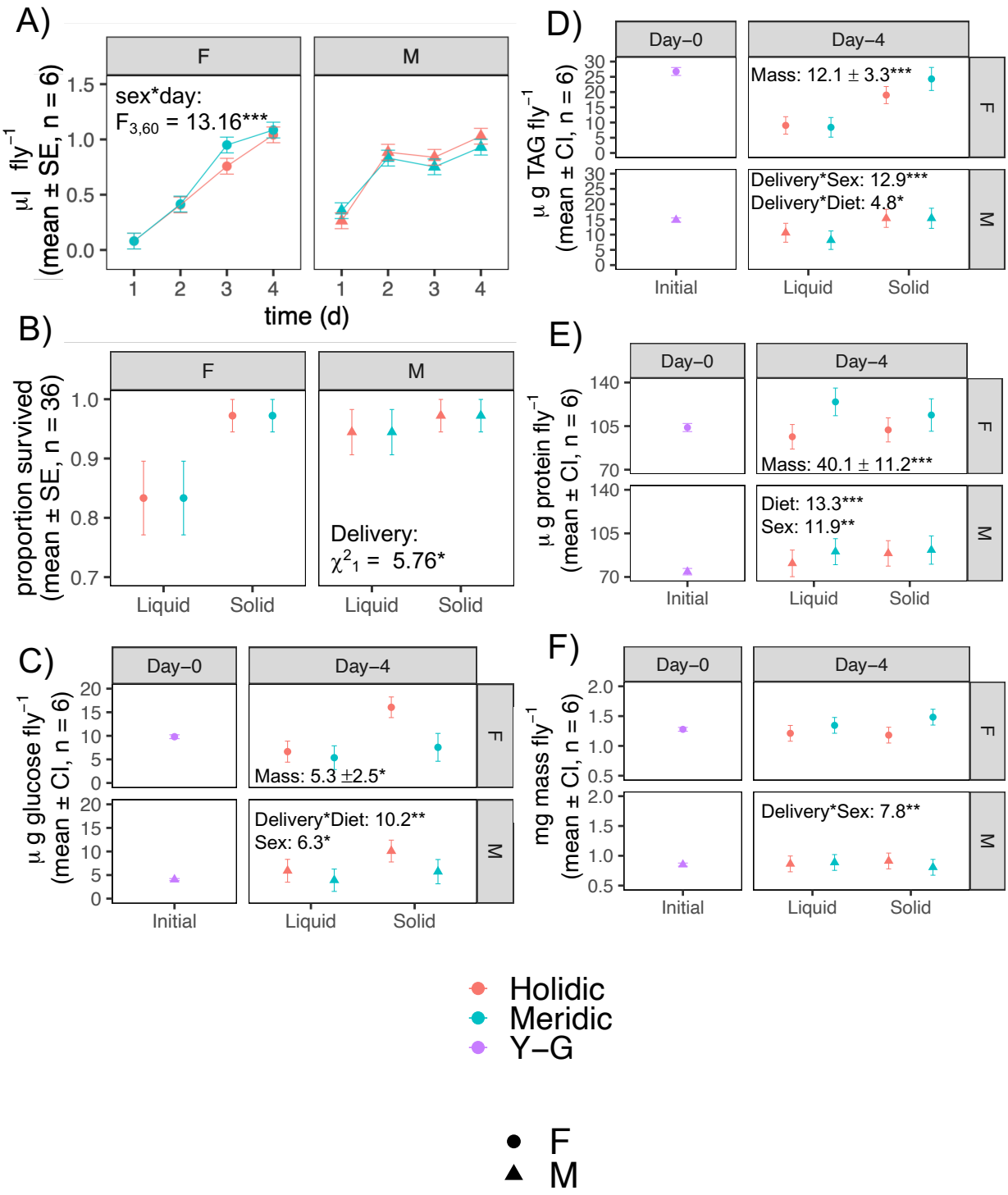

A)
F
M
sex*day:
$F_{3,60}$ = 13.16***
µl fly^{-1} (mean ± SE, n = 6)
time (d)
B)
proportion survived (mean ± SE, n = 36)
Delivery:
χ^2_1 = 5.76*
Liquid
Solid
C)
Day–0
Day–4
µ g glucose fly^{-1} (mean ± CI, n = 6)
Mass: 5.3 ±2.5*
Delivery*Diet: 10.2**
Sex: 6.3*
Initial
D)
µ g TAG fly^{-1} (mean ± CI, n = 6)
Mass: 12.1 ± 3.3***
Delivery*Sex: 12.9***
Delivery*Diet: 4.8*
E)
µ g protein fly^{-1} (mean ± CI, n = 6)
Mass: 40.1 ± 11.2***
Diet: 13.3***
Sex: 11.9**
F)
mg mass fly^{-1} (mean ± CI, n = 6)
Delivery*Sex: 7.8**
Holidic
Meridic
Y–G

Figure S2.2. Method development for capillary feeding experiments. Comparison of diet (holidic v. meridic) and delivery method (liquid or solid) on fly (A) feeding rate, (B) survival, nutritional indices ((C) glucose, (D) TAG, and (E) protein content), and fly weight for female (F) and male (M) flies. All data are displayed as estimated marginal mean and SE or CI, except for the raw mean and SE are shown for survival data. For nutritional indices and fly weight, day-0 raw mean and CI are shown for starting reference. The significant effect terms from models are displayed for each panel (significant main effects are not included if interaction term is significant). Panels C-F, the residual degrees of freedom are 39 and the fly weight covariate slope and SE are included under the day-4 label. Note: for solid medium, there is no direct measure of fly feeding as for the CAFÉ method. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

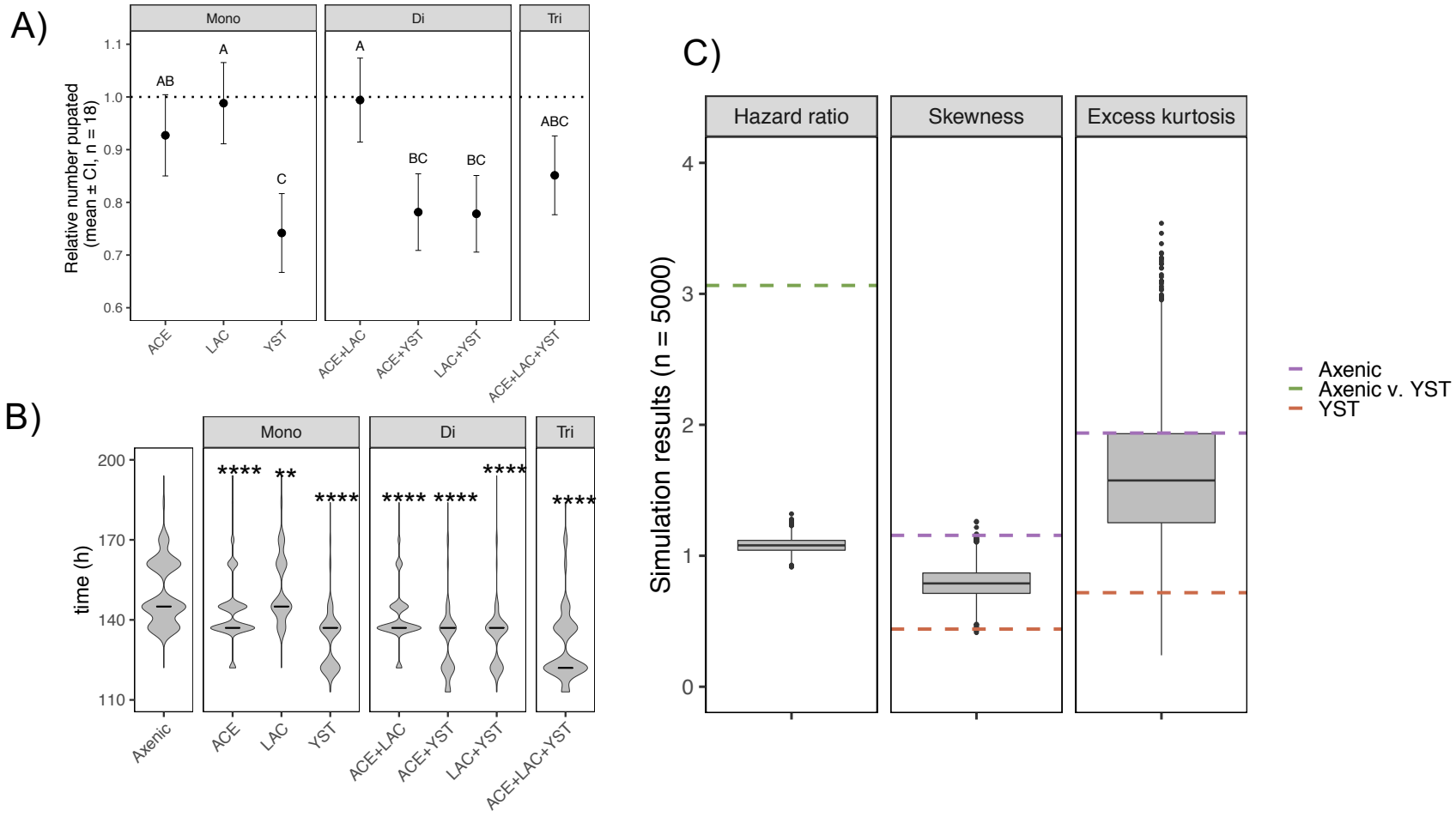

A)
Mono
Di
Tri
Relative number pupated
(mean ± CI, n = 18)
1.1
1.0
0.9
0.8
0.7
0.6
AB
A
C
A
BC
BC
ABC
ACE
LAC
YST
ACE+LAC
ACE+YST
LAC+YST
ACE+LAC+YST
B)
Mono
Di
Tri
time (h)
200
170
140
110
****
**
****
****
****
****
****
Axenic
ACE
LAC
YST
ACE+LAC
ACE+YST
LAC+YST
ACE+LAC+YST
C)
Hazard ratio
Skewness
Excess kurtosis
Simulation results (n = 5000)
4
3
2
1
0
Axenic
Axenic v. YST
YST

Figure S2.3. Analyses of *Drosophila* development time. (A) Relative number of insects pupated, (B) violin plot for time to pupation, and (C) summary statistics of simulated YST time to eclosion. In panel (A), the estimated marginal mean and 95% confidence intervals from ANOVA analysis are plotted, along with letter rankings from *post hoc* Tukey tests. The dashed line is the average value for axenic flies; confidence intervals that do not overlap with this line indicate a significant difference from axenic insects. Probability density function and median time (black bar) are shown for the time to pupation in panel (B). Asterisks indicated significance for Kolmogorov-Smirnov test ** < 0.01 and **** < 0.0001. Statistics for ANOVA treatment effect: $F_{6,110} = 7.40$, $p = 1.15 \times 10^{-6}$, $R^2 = 0.29$. Distribution statistics are in Table S2.1C. For panel (C), hazard ratio from Cox regression comparing simulated YST eclosion time to axenic data and measures of distribution shape (skewness and excess kurtosis) are shown from all 5,000 simulations. The dashed lines indicate results from observed data distributions to compare how the simulations performed against the observed data. For Cox regression analyses, only 22% of simulations had significant results. The variation in hazard ratios was low across simulations (relative standard deviation [RSD] = 5.2%), and the median time for all YST simulations was equal to the axenic flies (observed axenic and YST median time, 242 and 233 h, respectively) with increased measures of skewness and kurtosis. Simulated YST data indicates that if all flies reached eclosion, it is likely that this would result in an eclosion rate similar to the observed axenic fly data. Mono = mono-association. Di = di-association. Tri = tri-association.

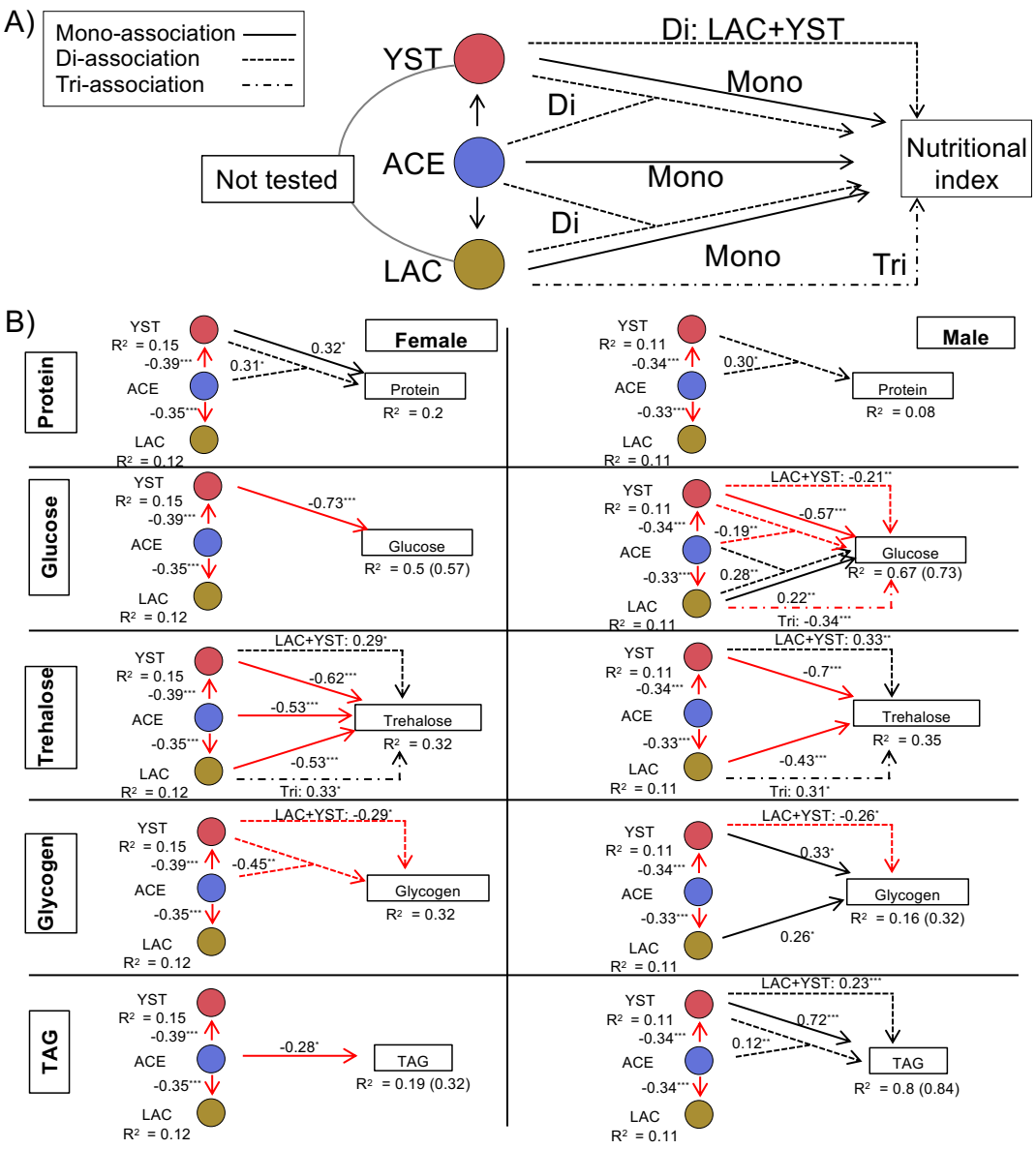

A)
Mono-association
Di-association
Tri-association
YST
ACE
LAC
Not tested
Di: LAC+YST
Mono
Di
Mono
Di
Mono
Tri
Nutritional index
B)
Female
Male
Protein
Glucose
Trehalose
Glycogen
TAG
YST
ACE
LAC
Protein
Glucose
Trehalose
Glycogen
TAG
Tri: 0.33*
Tri: -0.34***
Tri: 0.31*

Figure S2.4. Nutritional indices. (A) Generalized path analysis model tested to understand among-microbe interactions and their effect on a given nutritional index and (B) association of among-microbe interactions with nutritional indices. Most models did not include a random effect for experimental replicate ($p > 0.05$), expect for some of the linear models associating a nutritional index to among-microbe interactions (e.g. TAG ~ ACE*LAC*YST). These models include glucose content (both sexes), glycogen content (males only), and TAG content (both sexes). For Fig. 2.4 and S2.4B, both the marginal and conditional R^2 are reported for the nutritional index response variable. For panel B, left- and right-side are female and male flies, respectively. Red and black arrows indicate negative and positive associations, respectively. The standardized coefficient for each significant association and marginal R^2 values for all response variables are shown, with conditional R^2 values shown in parentheses if needed. Full description of statistical tests is in Table S2.1E.

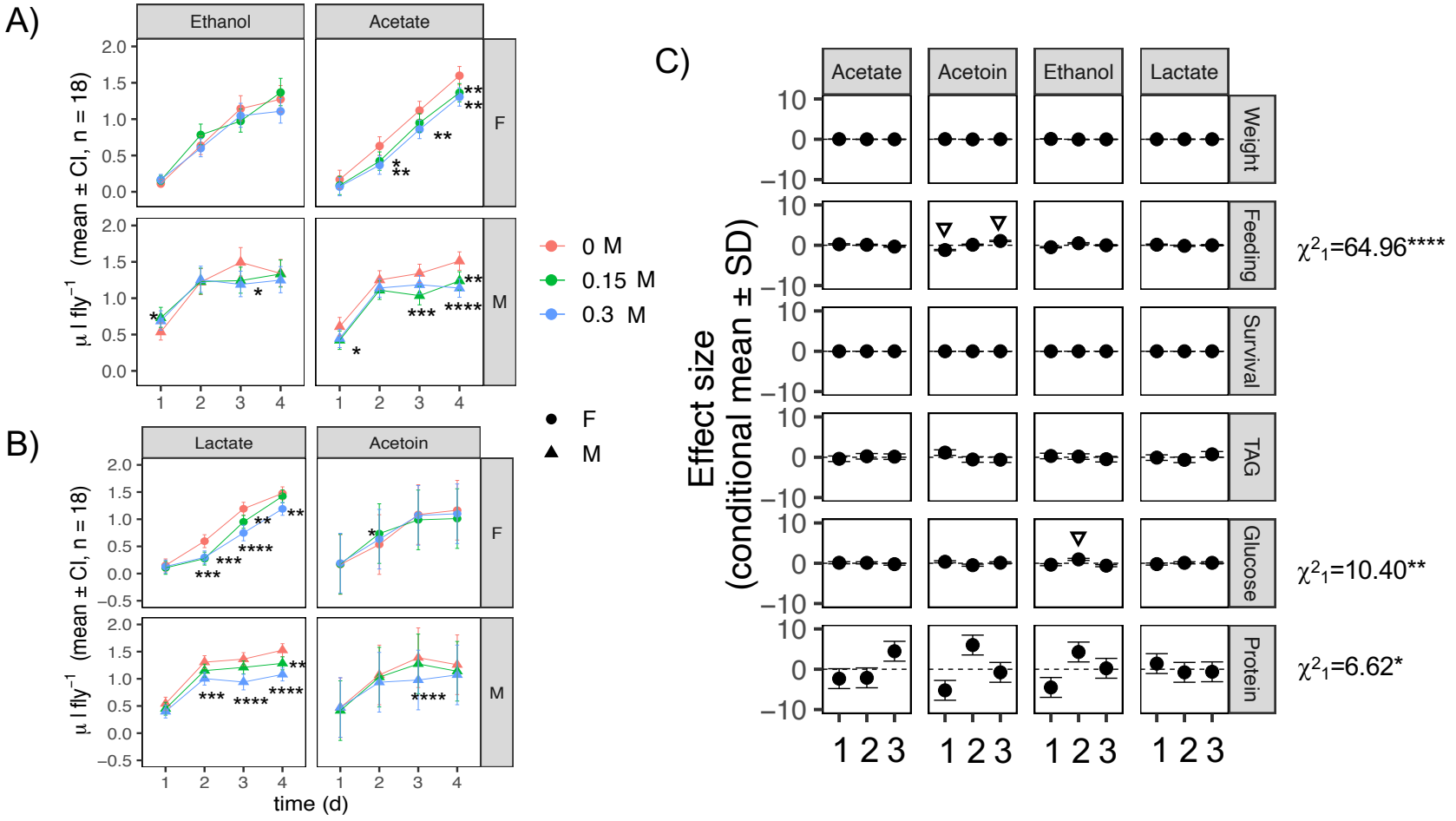

A)
Ethanol
Acetate
F
M
µl fly^{-1} (mean ± CI, n = 18)
0 M
0.15 M
0.3 M
B)
Lactate
Acetoin
time (d)
C)
Acetate
Acetoin
Ethanol
Lactate
Weight
Feeding
Survival
TAG
Glucose
Protein
Effect size
(conditional mean ± SD)
χ^2_1=64.96****
χ^2_1=10.40**
χ^2_1=6.62*

Figure S2.5. Performance and nutritional indices of *Drosophila* in CAFE assays. Feeding rates of flies administered (A) ethanol, (A) acetate, (B) lactate or (B) acetoin. (C) Comparison of variation between independent experimental replicates for the measures: fly weight, total diet volume consumed by day-4 (Feeding), survival at day-4 (Survival), and nutritional indices for TAG, glucose, and protein content. For panels A and B, the estimated marginal mean and confidence interval are plotted from ANOVA model. *Post hoc* Dunnett's test was used to compare treatments (0.15 and 0.3 M) to control diet for each day and sex (significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Effect test results in Table S1F. For panel C, experimental replicates are labeled as 1, 2, and 3 for each metabolite. Data are displayed as estimated conditional mean and standard deviation for experimental replicate random effect from each model. The dashed line indicates the grand mean. Significant effect terms from each model are shown next to measure label (significance: * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$). Open-triangle indicates experimental replicates that differ from grand mean as determined by non-overlapping 95% confidence intervals from simulated posterior distributions (note: no differences were found for protein content in the *post hoc* analysis). F = female. M = male. See Dyrad (doi: 10.5061/dryad.ngf1vhhrj) for extended version of this figure with data displayed from all associated measures in CAFE assays.

Table S2.1A. Effect of co-associations on microbial abundance and biomass

Analysis categories	**ACE**	**LAC**	**YST**	**Total abundance**	**Total biomass**
Tukey test					
ACE-F	B	-	-	B	C
ACE-M	ab	-	-	ab	bc
LAC-F	-	A	-	B	BC
LAC-M	-	a	-	bc	bc
YST-F	-	-	A	CD	A
YST-M	-	-	a	d	a
ACE+LAC-F	A	D	-	A	B
ACE+LAC-M	a	d	-	a	b
ACE+YST-F	C	-	B	D	D
ACE+YST-M	c	-	b	cd	c
LAC+YST-F	-	B	A	BC	A
LAC+YST-M	-	b	a	bcd	a
ACE+LAC+YST-F	C	C	B	CD	CD
ACE+LAC+YST-M	bc	c	b	bcd	c
ANOVA Effect test					
Treatment	**$F_{3,60.} = 31.58$, $p = 2.2x10^{-12}$**	**$F_{3,60.} = 231.69$, $p < 2.2 \times 10^{-16}$**	**$F_{3,64.} = 198.01$, $p < 2.2 \times 10^{-16}$**	**$F_{6,108.04} = 17.23$, $p = 6.80 \times 10^{-14}$**	**$F_{6,106.73} = 48.20$, $p < 2.2 \times 10^{-16}$**
Sex	**$F_{1,62} = 5.51$, $p = 0.0221$**	$F_{1,62} = 3.53$, $p = 0.0650$	**$F_{1,66} = 9.60$, $p = 0.0029$**	**$F_{1,110} = 33.23$, $p = 7.59 \times 10^{-8}$**	**$F_{1,107.43} = 40.74$, $p = 4.52 \times 10^{-9}$**
Interaction	$F_{3,62} = 2.74$, $p = 0.0506$	$F_{3,62} = 1.12$, $p = 0.3469$	**$F_{3,66} = 13.25$, $p = 7.2 \times 10^{-7}$**	**$F_{6,110} = 2.55$, $p = 0.0236$**	**$F_{6,107.50} = 4.21$, $p = 0.0008$**
Analysis of Deviance					
Experimental replicate (Vial)	**$X^2_2 = 29.71$, $p = 3.5 \times 10^{-7}$**	$X^2_2 = 0$, $p = 1$	**$X^2_2 = 20.92$, $p = 2.9 \times 10^{-5}$**	**$X^2_2 = 23.34$, $p = 8.53 \times 10^{-6}$**	**$X^2_2 = 30.59$, $p = 2.27 \times 10^{-7}$**
R^2					
Marginal	0.404	0.843	0.848	0.398	0.632
Conditional	0.682	0.843	0.922	0.632	0.797

The *post hoc* Tukey test was implemented to assess differences in microbial abundance between treatments for each sex. Uppercase letters are used to represent the ranking of female samples and lowercase letters are used to distinguish the rankings for male samples. A hyphen indicates treatment-sex combinations that were not used in species-specific analyses. P-values below the threshold are bolded.

Table S2.1B. Mixed effect Cox regression model results for influence of microbiota treatment on development time

i) Time to pupation: Axenic reference				
Treatment	Coefficient ± SE[a]	Hazard Ratio	z-value	p-value
ACE	**0.383 ± 0.090**	**1.466**	**4.26**	**< 0.0001**
LAC	-0.067 ± 0.091	0.935	-0.74	0.46
YST	**1.231 ± 0.089**	**3.426**	**13.87**	**< 0.0001**
ACE+LAC	**0.618 ± 0.091**	**1.856**	**6.79**	**< 0.0001**
ACE+YST	**1.298 ± 0.087**	**3.663**	**14.91**	**< 0.0001**
LAC+YST	**1.150 ± 0.087**	**3.157**	**13.28**	**< 0.0001**
ACE+LAC+YST	**1.142 ± 0.088**	**3.132**	**12.91**	**< 0.0001**
Analysis of deviance for Experiment(vial): **$\chi^2_2 = 653.13, p < 2.2 \times 10^{-16}$**				
ii) Time to eclosion: Axenic reference				
Treatment	Coefficient ± SE[a]	Hazard Ratio	z-value	p-value
ACE	**0.411 ± 0.085**	**1.509**	**4.81**	**< 0.0001**
LAC	-0.014 ± 0.086	0.987	-0.16	0.87
YST	**1.120 ± 0.084**	**3.064**	**13.27**	**< 0.0001**
ACE+LAC	**0.602 ± 0.086**	**1.826**	**6.97**	**< 0.0001**
ACE+YST	**1.378 ± 0.083**	**3.968**	**16.61**	**< 0.0001**
LAC+YST	**1.141 ± 0.082**	**3.129**	**13.87**	**< 0.0001**
ACE+LAC+YST	**1.270 ± 0.083**	**3.559**	**15.26**	**< 0.0001**
Analysis of deviance for Experiment(vial): **$\chi^2_2 = 367.93, p < 2.2 \times 10^{-16}$**				

[a]standard error
P-values below the threshold are bolded.

Table S2.1C. Microbial treatment effect on pupation and eclosion time distribution shape

Treatment	Pupation		Eclosion	
	Skewness	***Kurtosis***	***Skewness***	***Kurtosis***
Axenic	**skew = 0.63154, z = 8.16090, p = 3.324 x 10^{-16}**	**kurt = 2.5632, z = -3.8453, p = 0.0001204**	**skew = 1.1559, z = 13.3350, p < 2.2 x 10^{-16}**	**kurt = 4.9370, z = 7.5279, p = 5.155 x 10^{-14}**
ACE	**skew = 1.4526, z = 13.9400, p < 2.2 x 10^{-16}**	**kurt = 6.5036, z = 9.1271, p < 2.2 x 10^{-16}**	**skew = 1.3758, z = 13.5040, p < 2.2 x 10^{-16}**	**kurt = 5.1519, z = 7.1994, p = 6.05 x 10^{-13}**
LAC	**skew = 0.75575, z = 8.77260, p < 2.2 x 10^{-16}**	kurt = 2.8148, z = -1.2101, p = 0.2263	**skew = 0.995, z = 10.804, p < 2.2 x 10^{-16}**	**kurt = 3.7046, z = 3.5312, p = 0.0004137**
YST	**skew = 1.0856, z = 10.5910, p < 2.2 x 10^{-16}**	**kurt = 5.3998, z = 7.0930, p = 1.313 x 10^{-12}**	**skew = 0.44015, z = 4.91990, p = 8.66 x 10^{-7}**	**kurt = 3.7181, z = 3.2819, p = 0.001031**
ACE + LAC	**skew = 1.077, z = 11.327, p < 2.2 x 10^{-16}**	**kurt = 5.3551, z = 7.5245, p = 5.292 x 10^{-14}**	**skew = 1.7202, z = 15.5100, p < 2.2 x 10^{-16}**	**kurt = 6.8955, z = 9.5624, p < 2.2 x 10^{-16}**
ACE + YST	**skew = 0.82543, z = 9.04410, p < 2.2 x 10^{-16}**	**kurt = 4.2327, z = 5.0514, p = 4.386 x 10^{-7}**	**skew = 0.63815, z = 7.17800, p = 7.073 x 10^{-13}**	kurt = 3.08720, z = 0.63387, p = 0.5262
LAC + YST	**skew = 1.1518, z = 11.6320, p < 2.2 x 10^{-16}**	**kurt = 6.4887, z = 8.9211, p < 2.2 x 10^{-16}**	**skew = 0.51589, z = 5.99600, p = 2.023 x 10^{-9}**	**kurt = 4.4441, z = 5.5318, p = 3.17 x 10^{-8}**
ACE + LAC + YST	**skew = 1.3636, z = 13.3130, p < 2.2 x 10^{-16}**	**kurt = 4.9863, z = 6.8362, p = 8.132 x 10^{-12}**	**skew = 0.35846, z = 4.29100, p = 1.779 x 10^{-5}**	kurt = 2.7491, z = -1.6526, p = 0.09842
Kolmogorov-Smirnov tests	**Pupation**		**Eclosion**	
	D statistic	***p-value***	***D statistic***	***p-value***
ACE - Axenic	**0.2736**	**< 2.2 x 10^{-16}**	**0.2347**	**< 2.2 x 10^{-16}**
LAC - Axenic	**0.08423**	**0.0010**	0.0311	0.6761
YST - Axenic	**0.5896**	**< 2.2 x 10^{-16}**	**0.5428**	**< 2.2 x 10^{-16}**
ACE + LAC - Axenic	**0.3861**	**< 2.2 x 10^{-16}**	**0.3156**	**< 2.2 x 10^{-16}**
ACE + YST - Axenic	**0.6047**	**< 2.2 x 10^{-16}**	**0.6781**	**< 2.2 x 10^{-16}**
LAC + YST - Axenic	**0.6119**	**< 2.2 x 10^{-16}**	**0.5388**	**< 2.2 x 10^{-16}**
ACE + LAC + YST - Axenic	**0.5740**	**< 2.2 x 10^{-16}**	**0.6081**	**< 2.2 x 10^{-16}**

P-values below the threshold are bolded.

Table S2.1D. Effect of microbial treatment and sex on fly weight and nutritional indices

Analysis categories	**Weight**	**Protein**	**TAG**	**Glucose**	**Trehalose**	**Glycogen**
ANOVA Effect test						
Treatment	**$F_{7,122.53}$ = 16.75, p = 2.3 x 10^{-15}**	**$F_{7,124.11}$ = 5.61, p = 1.2 x 10^{-5}**	**$F_{7,124.02}$ = 62.41, p < 2.2 x 10^{-16}**	**$F_{7,124.05}$ = 68.53, p < 2.2 x 10^{-16}**	**$F_{7,124.22}$ = 12.46, p = 5.1 x 10^{-12}**	**$F_{7,124.04}$ = 8.64, p = 1.3 x 10^{-8}**
Sex	**$F_{1,124.02}$ = 4933.70, p < 2.2 x 10^{-16}**	**$F_{1,126}$ = 19.69, p = 2.0 x 10^{-5}**	**$F_{1,126}$ = 43.22, p = 1.2 x 10^{-9}**	**$F_{1,126}$ = 836.85, p < 2.2 x 10^{-16}**	**$F_{1,1126}$= 68.70, p = 1.5 x 10^{-13}**	**$F_{1,126}$ = 376.79, p < 2.2 x 10^{-16}**
Interaction	**$F_{3,124.09}$ = 8.11, p = 4.2 x 10^{-8}**	$F_{7,126}$ = 1.24, p = 0.2848	**$F_{7,126}$ = 36.86, p < 2.2 x 10^{-16}**	**$F_{7,126}$ = 10.08, p = 5.8 x 10^{-10}**	**$F_{7,126}$ = 3.10, p = 0.0047**	**$F_{6,126}$ = 7.19, p = 3.1 x 10^{-7}**
Analysis of Deviance						
Experimental replicate (Vial)	X^2_2 = 0.72, p = 0.6974	X^2_2 = 3.02, p = 0.221	**X^2_2 = 23.25, p = 8.9 x 10^{-6}**	**X^2_2 = 631.42, p < 2.2 x 10^{-16}**	**X^2_2 = 94.21, p < 2.2 x 10^{-16}**	**X^2_2 = 20.35, p = 3.8 x 10^{-5}**
R^2						
Marginal	0.947	0.211	0.702	0.812	0.394	0.578
Conditional	0.952	0.304	0.747	0.881	0.825	0.704

P-values below the threshold are bolded.

Table S2.1E. Structural equation model output for among-microbe interactions and their influence on *Drosophila* nutritional indices.

i) Protein content – Female samples (Fisher's C = 1.295, df = 2, p = 0.523)					
Response	Predictor	Unstandardized coefficient	SE	Standardized coefficient	P value
Protein	ACE	-0.4200	2.0154	-0.0307	0.8352
Protein	LAC	-1.7197	2.0957	-0.0951	0.4134
Protein	**YST**	**5.7954**	**2.2936**	**0.3239**	**0.0127**
Protein	ACE*LAC	-0.2194	1.8894	-0.0179	0.9077
Protein	**ACE*YST**	**3.5765**	**1.7824**	**0.3138**	**0.0469**
Protein	LAC*YST	0.4653	0.9844	0.0626	0.6373
Protein	ACE*LAC*YST	-1.3972	1.5684	-0.1615	0.3747
YST	**ACE**	**-0.2952**	**0.0614**	**-0.3862**	**< 0.0001**
LAC	**ACE**	**-0.2628**	**0.0617**	**-0.3475**	**< 0.0001**
ii) Protein content – Male samples (Fisher's C = 1.045, df = 2, p = 0.593)					
Response	Predictor	Unstandardized coefficient	SE	Standardized coefficient	P value
Protein	ACE	-1.9905	2.8024	-0.1138	0.4788
Protein	LAC	-0.4113	3.0130	-0.0175	0.8916
Protein	YST	2.0101	3.3455	0.0817	0.5490
Protein	ACE*LAC	1.9296	2.8328	0.1265	0.4970
Protein	**ACE*YST**	**4.2620**	**2.0236**	**0.2953**	**0.0372**
Protein	LAC*YST	1.0465	1.4651	0.1030	0.4764
Protein	ACE*LAC*YST	-1.7050	2.3303	-0.1291	0.4657
YST	**ACE**	**-0.2398**	**0.0582**	**-0.3374**	**0.0001**
LAC	**ACE**	**-0.2486**	**0.0611**	**-0.3338**	**0.0001**
iii) Glucose content – Female samples (Fisher's C = 1.295, df = 2, p = 0.523)					
Response	Predictor	Unstandardized coefficient	SE	Standardized coefficient	P value
Glucose	ACE	-0.1082	0.0682	-0.1741	0.1154
Glucose	LAC	0.0148	0.0709	0.0180	0.8350
Glucose	**YST**	**-0.5910**	**0.0777**	**-0.7272**	**< 0.0001**
Glucose	ACE*LAC	0.0764	0.0639	0.1372	0.2346
Glucose	ACE*YST	-0.0795	0.0604	-0.1536	0.1910
Glucose	LAC*YST	-0.0056	0.0333	-0.0166	0.8668
Glucose	ACE*LAC*YST	0.0133	0.0533	0.0338	0.8037
YST	**ACE**	**-0.2952**	**0.0614**	**-0.3862**	**< 0.0001**
LAC	**ACE**	**-0.2628**	**0.0617**	**-0.3475**	**< 0.0001**
iv) Glucose content – Male samples (Fisher's C = 1.045, df = 2, p = 0.593)					
Response	Predictor	Unstandardized coefficient	SE	Standardized coefficient	P value
Glucose	ACE	-0.0504	0.0462	-0.0957	0.2777
Glucose	**LAC**	**0.1550**	**0.0496**	**0.2191**	**0.0022**
Glucose	**YST**	**-0.4220**	**0.0551**	**-0.5693**	**< 0.0001**
Glucose	**ACE*LAC**	**0.1298**	**0.0467**	**0.2825**	**0.0063**
Glucose	**ACE*YST**	**-0.0822**	**0.0333**	**-0.1892**	**0.0150**
Glucose	**LAC*YST**	**-0.0648**	**0.0241**	**-0.2117**	**0.0083**
Glucose	**ACE*LAC*YST**	**-0.1350**	**0.0385**	**-0.3393**	**0.0006**
YST	**ACE**	**-0.2398**	**0.0582**	**-0.3374**	**0.0001**
LAC	**ACE**	**-0.2486**	**0.0611**	**-0.3338**	**0.0001**

v) Trehalose content – Female samples (Fisher's C = 1.295, df = 2, p = 0.523)					
Response	Predictor	Unstandardized coefficient	SE	Standardized coefficient	P value
Trehalose	**ACE**	**-0.2915**	**0.0754**	**-0.5255**	**0.0002**
Trehalose	**LAC**	**-0.3887**	**0.0784**	**-0.5299**	**< 0.0001**
Trehalose	**YST**	**-0.4483**	**0.0858**	**-0.6178**	**< 0.0001**
Trehalose	ACE*LAC	0.1031	0.0707	0.2075	0.1470
Trehalose	ACE*YST	-0.1020	0.0667	-0.2207	0.1286
Trehalose	**LAC*YST**	**0.0884**	**0.0368**	**0.2934**	**0.0178**
Trehalose	**ACE*LAC*YST**	**0.1168**	**0.0587**	**0.3329**	**0.0487**
YST	**ACE**	**-0.2952**	**0.0614**	**-0.3862**	**< 0.0001**
LAC	**ACE**	**-0.2628**	**0.0617**	**-0.3475**	**< 0.0001**
vi) Trehalose content – Male samples (Fisher's C = 1.045, df = 2, p = 0.593)					
Response	Predictor	Unstandardized coefficient	SE	Standardized coefficient	P value
Trehalose	ACE	-0.1052	0.0871	-0.1629	0.2297
Trehalose	**LAC**	**-0.3726**	**0.0937**	**-0.4297**	**0.0001**
Trehalose	**YST**	**-0.6365**	**0.1040**	**-0.7005**	**< 0.0001**
Trehalose	ACE*LAC	-0.0388	0.0881	-0.0690	0.6600
Trehalose	ACE*YST	-0.0661	0.0629	-0.1241	0.2952
Trehalose	**LAC*YST**	**0.1239**	**0.0456**	**0.3302**	**0.0075**
Trehalose	**ACE*LAC*YST**	**0.1495**	**0.0725**	**0.3064**	**0.0412**
YST	**ACE**	**-0.2398**	**0.0582**	**-0.3374**	**0.0001**
LAC	**ACE**	**-0.2486**	**0.0611**	**-0.3338**	**0.0001**
vii) Glycogen content – Female samples (Fisher's C = 1.295, df = 2, p = 0.523)					
Response	Predictor	Unstandardized coefficient	SE	Standardized coefficient	P value
Glycogen	ACE	-0.0724	0.1345	-0.0729	0.5914
Glycogen	LAC	0.1846	0.1399	0.1406	0.1893
Glycogen	YST	-0.1517	0.1531	-0.1168	0.3236
Glycogen	ACE*LAC	0.1448	0.1261	0.1628	0.2530
Glycogen	**ACE*YST**	**-0.3761**	**0.1190**	**-0.4547**	**0.0020**
Glycogen	**LAC*YST**	**-0.1548**	**0.0657**	**-0.2870**	**0.0200**
Glycogen	ACE*LAC*YST	-0.0003	0.1047	-0.0004	0.9979
YST	**ACE**	**-0.2952**	**0.0614**	**-0.3862**	**< 0.0001**
LAC	**ACE**	**-0.2628**	**0.0617**	**-0.3475**	**< 0.0001**
viii) Glycogen content – Male samples (Fisher's C = 1.045, df = 2, p = 0.593)					
Response	Predictor	Unstandardized coefficient	SE	Standardized coefficient	P value
Glycogen	**ACE**	**0.3215**	**0.1371**	**0.3303**	**0.0206**
Glycogen	**LAC**	**0.3444**	**0.1470**	**0.2635**	**0.0207**
Glycogen	YST	-0.1050	0.1634	-0.0766	0.5219
Glycogen	ACE*LAC	-0.0384	0.1384	-0.0453	0.7816
Glycogen	ACE*YST	0.0056	0.0988	0.0069	0.9553
Glycogen	**LAC*YST**	**-0.1444**	**0.0715**	**-0.2553**	**0.0456**
Glycogen	ACE*LAC*YST	-0.0484	0.1141	-0.0659	0.6718
YST	**ACE**	**-0.2398**	**0.0582**	**-0.3374**	**0.0001**
LAC	**ACE**	**-0.2486**	**0.0611**	**-0.3338**	**0.0001**

ix) TAG content – Female samples (Fisher's C = 1.295, df = 2, p = 0.523)					
Response	Predictor	Unstandardized coefficient	SE	Standardized coefficient	P value
TAG	**ACE**	**-0.8619**	**0.4261**	**-0.2779**	**0.0453**
TAG	LAC	-0.1071	0.4429	-0.0261	0.8094
TAG	YST	0.0214	0.4852	0.0053	0.9648
TAG	ACE*LAC	-0.2136	0.3994	-0.0769	0.5938
TAG	ACE*YST	-0.0476	0.3776	-0.0184	0.8999
TAG	LAC*YST	0.3802	0.2082	0.2256	0.0702
TAG	ACE*LAC*YST	-0.0362	0.3327	-0.0185	0.9135
YST	**ACE**	**-0.2952**	**0.0614**	**-0.3862**	**< 0.0001**
LAC	**ACE**	**-0.2628**	**0.0617**	**-0.3475**	**< 0.0001**
x) TAG content – Male samples (Fisher's C = 0.988, df = 2, p = 0.61)					
Response	Predictor	Unstandardized coefficient	SE	Standardized coefficient	P value
TAG	ACE	0.0026	0.4246	0.0004	0.9952
TAG	LAC	-0.4702	0.4553	-0.0551	0.3037
TAG	**YST**	**6.3893**	**0.5064**	**0.7152**	**< 0.0001**
TAG	ACE*LAC	-0.7518	0.4287	-0.1352	0.0820
TAG	**ACE*YST**	**0.6317**	**0.3061**	**0.1201**	**0.0411**
TAG	**LAC*YST**	**0.8681**	**0.2215**	**0.2344**	**0.0001**
TAG	ACE*LAC*YST	0.6113	0.3535	0.1268	0.0862
YST	**ACE**	**-0.2398**	**0.0582**	**-0.3400**	**0.0001**
LAC	**ACE**	**-0.2486**	**0.0611**	**-0.3369**	**0.0001**

P-values below the threshold are bolded.

Table S2.1F. Effect of administered microbial fermentation products on *Drosophila* feeding, survival, weight, and nutritional indices

i) Metabolite effect on feeding by day				
ANOVA effect test (Type III ANOVA)	***Ethanol (sqrt-transformed)***	***Acetic acid***	***Lactic acid***	***Acetoin***
Concentration	$F_{2,100}$ = 1.46, p = 0.2362	**$F_{2,100}$ = 26.02, p = 8.0 x 10^{-10}**	**$F_{2,103}$ = , 41.41 p = 6.4 x 10^{-14}**	$F_{2,100}$ = 2.57, p = 0.0817
Sex	**$F_{1,100}$ = 204.93, p < 2.2 x 10^{-16}**	**$F_{1,100}$ = 114.84, p < 2.2 x 10^{-16}**	**$F_{1,102}$ = 138.09, p < 2.2 x 10^{-16}**	**$F_{1,100}$ = 50.98, p = 1.5 x 10^{-10}**
Day	**$F_{3,306}$ = 318.40, p < 2.2 x 10^{-16}**	**$F_{3,306}$ = 388.81, p < 2.2 x 10^{-16}**	**$F_{3,300}$ = 138.09, p < 2.2 x 10^{-16}**	**$F_{3,306}$ = 301.96, p < 2.2 x 10^{-16}**
Concentration*Sex	$F_{2,100}$ = 0.29, p = 0.7489	$F_{2,100}$ = 0.77, p = 0.4668	$F_{2,102}$ = 0.73, p = 0.4863	**$F_{2,100}$ = 3.33, p = 0.0397**
Concentration*Day	**$F_{6,306}$ = 4.06, p = 0.0006**	$F_{6,306}$ = 1.33, p = 0.2454	**$F_{6,300}$ = 4.05, p = 0.0006**	**$F_{6,306}$ = 2.94, p = 0.0084**
Sex*Day	**$F_{3,306}$ = 42.62, p < 2.2 x 10^{-16}**	**$F_{3,306}$ = 56.17, p < 2.2 x 10^{-16}**	**$F_{3,301}$ = 57.15, p < 2.2 x 10^{-16}**	**$F_{3,306}$ = 9.49, p = 5.2 x 10^{-6}**
Concentration*Sex*Day	$F_{6,306}$ = 1.13, p = 0.3462	$F_{6,306}$ = 0.96, p = 0.4561	$F_{6,300}$ = 1.02, p = 0.4108	**$F_{6,306}$ = 2.18, p = 0.0452**
Analysis of deviance				
Experimental replicate	**χ^2_1 = 6.855, p = 0.0088**	χ^2_1 = 2.054, p = 0.1519	χ^2_1 = 0.561, p = 0.4537	**χ^2_1 = 98.164, p < 2.2 x 10^{-16}**
R^2				
Marginal	0.741	0.765	0.782	0.503
Conditional	0.766	0.792	0.799	0.801
ii) Metabolite effect on final survival				
Effect test (Wald's χ^2)	***Ethanol***	***Acetic acid***	***Lactic acid***	***Acetoin***
Concentration	**χ^2_2 = 7.686, p = 0.0214**	**χ^2_2 = 11.859, p = 0.0027**	**χ^2_2 = 9.470, p = 0.0088**	χ^2_2 = 1.575, p = 0.4550
Sex	χ^2_1 = 2.148, p = 0.1428	χ^2_1 = 0.582, p = 0.4457	**χ^2_1 = 6.066, p = 0.0138**	χ^2_1 = 0.076, p = 0.7834
Concentration*Sex	χ^2_2 = 1.620, p = 0.4448	χ^2_2 = 5.382, p = 0.0678	χ^2_2 = 2.536, p = 0.2815	χ^2_2 = 3.752, p = 0.1532
Total food consumed	χ^2_1 = 1.807, p = 0.1789	χ^2_1 = 0.000, p = 0.9955	**χ^2_1 = 4.535, p = 0.0332**	χ^2_1 = 3.146, p = 0.0761
Analysis of deviance				
Experimental replicate	χ^2_1 = 0, p = 0.9999	χ^2_1 = 2.864, p = 0.091	**χ^2_1 = 6.235, p = 0.0125**	χ^2_1 = 0, p = 0.9992
iii) Metabolite effect on fly weight				
ANOVA effect test (Type II ANOVA)	***Ethanol (sqrt-transformed)***	***Acetic acid (sqrt-transformed)***	***Lactic acid***	***Acetoin***
Concentration	$F_{2,99}$ = 0.119, p = 0.8883	**$F_{2,100}$ = 6.102, p = 0.0032**	$F_{2,87}$ = 1.434, p = 0.2439	$F_{2,100}$ = 2.584, p = 0.0805
Sex	**$F_{1,101}$ = 39.105, p = 9.8 x 10^{-9}**	**$F_{1,101}$ = 63.300, p = 2.7 x 10^{-12}**	**$F_{1,88}$ = 9.188, p = 0.0032**	**$F_{1,96}$ = 83.452, p = 1.1 x 10^{-14}**
Concentration*Sex	$F_{2,99}$ = 0.892, p = 0.4130	**$F_{2,99}$ = 4.647, p = 0.0118**	$F_{2,86}$ = 0.460, p = 0.6331	$F_{2,100}$ = 0.761, p = 0.4699
Total food consumed	$F_{1,101}$ = 0.509, p	**$F_{1,101}$ = 5.060, p**	$F_{1,87}$ = 1.494,	$F_{1,38}$ = 2.764,

	= 0.4770	**= 0.0267**	p = 0.2248	p = 0.1046
Analysis of deviance				
Experimental replicate	**χ^2_1 = 5.182, p = 0.0228**	χ^2_1 = 0.636, p = 0.4253	χ^2_1 = 0, p = 1	χ^2_1 = 0.081, p = 0.7759
R^2				
Marginal	0.487	0.488	0.363	0.474
Conditional	0.562	0.510	0.371	0.564
iv) Metabolite effect on TAG content				
ANOVA effect test (Type II ANOVA)	***Ethanol***	***Acetic acid***	***Lactic acid***	***Acetoin***
Concentration	$F_{2,98.32}$ = 0.073, p = 0.9293	**$F_{2,99.32}$ = 5.822, p = 0.0041**	$F_{2,85.69}$ = 1.180, p = 0.3122	$F_{2,99.04}$ = 0.775, p = 0.4636
Sex	$F_{1,75.92}$ = 3.606, p = 0.0614	$F_{1,97.97}$ = 3.124, p = 0.0803	$F_{1,85.66}$ = 3.393, p = 0.0689	$F_{1,83.94}$ = 2.617, p = 0.1095
Concentration*Sex	$F_{2,98.02}$ = 0.128, p = 0.8805	$F_{2,98.31}$ = 0.786, p = 0.4587	$F_{2,85.01}$ = 2.976, p = 0.0563	$F_{2,99.04}$ = 0.459, p = 0.6335
Total food consumed	$F_{1,88.47}$ = 0.055, p = 0.8148	$F_{1,91.07}$ = 1.134, p = 0.2897	$F_{1,85.56}$ = 1.844, p = 0.1781	$F_{1,22.59}$ = 1.703, p = 0.2050
Weight fly^{-1}	**$F_{1,91.85}$ = 8.477, p = 0.0045**	**$F_{198.79}$ = 49.98, p = 2.2 x 10^{-10}**	**$F_{1,85.43}$ = 41.40, p = 7.0 x 10^{-9}**	**$F_{1,\ 99.90}$ = 33.19, p = 9.3 x 10^{-8}**
Analysis of deviance				
Experimental replicate	χ^2_1 = 0.089, p = 0.765	χ^2_1 = 0, p = 1	**χ^2_1 = 15.017, p = 0.0001**	χ^2_1 = 0.043, p = 0.8353
R^2				
Marginal	0.320	0.436	0.350	0.341
Conditional	0.338	0.441	0.524	0.426
v) Metabolite effect on glucose content				
ANOVA effect test (Type II ANOVA)	***Ethanol***	***Acetic acid***	***Lactic acid***	***Acetoin***
Concentration	$F_{2,98.09}$ = 0.759, p = 0.4709	$F_{2,98.95}$ = 0.628, p = 0.5359	$F_{2,84.01}$ = 3.784, p = 0.0267	$F_{2,98.57}$ = 0.994, p = 0.3739
Sex	**$F_{1,99}$ = 42.031, p = 3.5 x 10^{-9}**	**$F_{1,100}$ = 28.913, p = 5.0 x 10^{-7}**	**$F_{1,83.9}$ = 32.465, p = 1.8 x 10^{-7}**	**$F_{1,98.6}$ = 15.587, p = 0.0001**
Concentration*Sex	$F_{2,98.00}$ = 2.178, p = 0.1188	$F_{2,98.16}$ = 0.024, p = 0.9760	$F_{2,85.13}$ = 0.872, p = 0.4218	**$F_{2,98.57}$ = 4.831, p = 0.0100**
Total food consumed	**$F_{1,100}$ = 4.219, p = 0.0426**	**$F_{1,99}$ = 11.685, p = 0.0009**	$F_{1,85.25}$ = 2.894, p = 0.0926	$F_{1,69.81}$ = 2.483, p = 0.1196
Weight fly^{-1}	$F_{1,99.94}$ = 3.501, p = 0.0643	**$F_{1,99.97}$ = 5.213, p = 0.0245**	**$F_{1,86.67}$ = 5.572, p = 0.0205**	**$F_{1,99.6}$ = 44.592, p = 1.4 x 10^{-9}**
Analysis of deviance				
Experimental replicate	**χ^2_1 = 7.610, p = 0.0058**	χ^2_1 = 0.556, p = 0.4558	χ^2_1 = 0, p = 1	**χ^2_1 = 12.286, p = 0.0005**
R^2				
Marginal	0.535	0.476	0.576	0.575
Conditional	0.612	0.499	0.576	0.708
vi) Metabolite effect on protein content				
ANOVA effect test (Type II ANOVA)	***Ethanol***	***Acetic acid***	***Lactic acid***	***Acetoin***
Concentration	$F_{2,98.09}$ = 2.223, p = 0.1137	**$F_{2,99.29}$ = 7.939, p = 0.0006**	$F_{2,85.5}$ = 2.874, p = 0.0619	$F_{2,99.44}$ = 0.763, p = 0.4690

Sex	**$F_{1,99.8}$ = 38.28, p = 1.4 x 10^{-8}**	**$F_{1,95.1}$ = 60.998, p = 7.5 x 10^{-12}**	**$F_{1,86}$ = 35.555, p = 5.3 x 10^{-8}**	$F_{1,54.03}$ = 3.839, p = 0.0553
Concentration*Sex	$F_{2,98.00}$ = 1.184, p = 0.3103	$F_{2,98.40}$ = 0.295, p = 0.7452	**$F_{2,85.09}$ = 5.598, p = 0.0052**	$F_{2,99.43}$ = 0.795, p = 0.4545
Total food consumed	$F_{1,100}$ = 3.117, p = 0.0805	$F_{1,83.09}$ = 0.085, p = 0.7714	$F_{1,86.60}$ = 1.804, p = 0.1826	$F_{1,2.67}$ = 0.500, p = 0.5362
Weight fly^{-1}	$F_{1,100}$ = 0.273, p = 0.6023	**$F_{1,96.5}$ = 38.171, p = 1.5 x 10^{-8}**	**$F_{1,87}$ = 87.384, p = 8.7 x 10^{-15}**	$F_{1,98.43}$ = 0.169, p = 0.6821
Analysis of deviance				
Experimental replicate	**χ^2_1 = 6.976, p = 0.00826**	χ^2_1 = 0, p = 1	χ^2_1 = 0, p = 1	χ^2_1 = 0, p = 1
R^2				
Marginal	0.633	0.814	0.799	0.105
Conditional	0.693	0.814	0.802	0.106

P-values below the threshold are bolded.

Table S2.1G. Effect of microbiota member presence on SCFA content

Analysis categories	Acetic acid	Propionic acid	Butyric acid	Total
Tukey test				
Axenic-F	AB	AB	AB	AB
Axenic-M	a	a	b	a
ACE-F	B	B	B	B
ACE-M	a	a	b	a
LAC-F	AB	AB	AB	AB
LAC-M	a	a	b	a
YST-F	AB	A	A	A
YST-M	a	a	a	a
ACE+LAC-F	AB	AB	AB	AB
ACE+LAC-M	a	a	b	a
ACE+YST-F	A	AB	B	AB
ACE+YST-M	a	a	ab	a
LAC+YST-F	AB	AB	AB	AB
LAC+YST-M	a	a	ab	a
ACE+LAC+YST-F	AB	AB	AB	AB
ACE+LAC+YST-M	a	a	ab	a
ANOVA Effect test				
sex	$F_{1,39.34} = 0.0025$, $p = 0.9607$, $\omega^2 = -0.015$	$F_{1,39.33} = 0.0881$, $p = 0.7681$, $\omega^2 = -0.012$	$F_{1,39.35} = 0.0635$, $p = 0.8023$, $\omega^2 = -0.011$	$F_{1,39.29} = 0.0704$, $p = 0.7921$, $\omega^2 = -0.013$
ACE	$F_{1,37.91} = 0.3776$, $p = 0.5424$, $\omega^2 = -0.010$	$F_{1,39.91} = 0.1708$, $p = 0.6816$, $\omega^2 = -0.011$	*$F_{1,39.90} = 3.1695$, $p = 0.0826$, $\omega^2 = 0.025$*	$F_{1,39.93} = 0.0010$, $p = 0.9747$, $\omega^2 = -0.014$
LAC	$F_{1,37.91} = 0.8531$, $p = 0.3612$, $\omega^2 = -0.003$	$F_{1,39.91} = 1.0488$, $p = 0.3120$, $\omega^2 = 4.5e\text{-}06$	$F_{1,39.90} = 0.6764$, $p = 0.4157$, $\omega^2 = -0.004$	$F_{1,39.93} = 1.1419$, $p = 0.2917$, $\omega^2 = 0.001$
YST	***$F_{1,37.91} = 9.1564$, $p = 0.0043$, $\omega^2 = 0.125$***	***$F_{1,39.91} = 8.3138$, $p = 0.0063$, $\omega^2 = 0.093$***	***$F_{1,39.90} = 18.024$, $p = 0.0001$, $\omega^2 = 0.197$***	***$F_{1,39.93} = 10.511$, $p = 0.0024$, $\omega^2 = 0.136$***
ACE*sex	$F_{1,39.38} = 1.5957$, $p = 0.2139$, $\omega^2 = 0.008$	***$F_{1,39.37} = 14.659$, $p = 0.0005$, $\omega^2 = 0.178$***	$F_{1,39.40} = 1.5576$, $p = 0.2194$, $\omega^2 = 0.007$	***$F_{1,39.32} = 6.9104$, $p = 0.0122$, $\omega^2 = 0.085$***
LAC*sex	$F_{1,39.38} = 0.0064$, $p = 0.9365$, $\omega^2 = -0.015$	$F_{1,39.37} = 0.0027$, $p = 0.9590$, $\omega^2 = -0.013$	*$F_{1,39.40} = 2.0504$, $p = 0.1601$, $\omega^2 = 0.012$*	$F_{1,39.32} = 0.0068$, $p = 0.9345$, $\omega^2 = -0.014$
YST*sex	$F_{1,39.38} = 1.3405$, $p = 0.2539$, $\omega^2 = 0.004$	$F_{1,39.37} = 1.3654$, $p = 0.2496$, $\omega^2 = 0.005$	***$F_{1,39.40} = 7.9497$, $p = 0.0075$, $\omega^2 = 0.082$***	$F_{1,39.32} = 0.0125$, $p = 0.9117$, $\omega^2 = -0.014$
ACE*LAC	*$F_{1,39.92} = 1.9286$, $p = 0.1726$, $\omega^2 = 0.014$*	***$F_{1,39.93} = 5.7471$, $p = 0.0213$, $\omega^2 = 0.062$***	$F_{1,39.92} = 0.8445$, $p = 0.3636$, $\omega^2 = -0.002$	*$F_{1,39.94} = 3.7050$, $p = 0.0614$, $\omega^2 = 0.039$*
ACE*YST	*$F_{1,39.92} = 3.2895$, $p = 0.0772$, $\omega^2 = 0.035$*	$F_{1,39.93} = 0.1624$, $p = 0.6891$, $\omega^2 = -0.011$	$F_{1,39.92} = 1.8160$, $p = 0.1854$, $\omega^2 = 0.009$	$F_{1,37.95} = 1.1702$, $p = 0.2858$, $\omega^2 = 0.003$
LAC*YST	$F_{1,39.92} = 1.0515$, $p = 0.3113$, $\omega^2 = 0.001$	*$F_{1,39.93} = 2.8748$, $p = 0.0978$, $\omega^2 = 0.024$*	$F_{1,39.92} = 1.7766$, $p = 0.1901$, $\omega^2 = 0.009$	*$F_{1,39.94} = 2.6080$, $p = 0.1142$, $\omega^2 = 0.023$*
ACE*LAC*sex	$F_{1,39.40} = 0.1534$,	$F_{1,39.40} = 1.3125$,	***$F_{1,39.42} = 4.2188$,***	$F_{1,39.35} = 0.0017$,

	p = 0.6975, ω^2 = - 0.013	p = 0.2589, ω^2 = 0.004	***p = 0.0467,*** ***ω^2 = 0.038***	p= 0.9671, ω^2 = -0.014
ACE*YST*sex	$F_{1,39.40}$ = 0.4364, p = 0.5127, ω^2 = - 0.008	$F_{1,39.40}$ = 0.0809, p = 0.7776, ω^2 = -0.012	$F_{1,39.42}$ = 0.5096, p = 0.4795, ω^2 = -0.006	$F_{1,39.35}$ = 0.1985, p= 0.6584, ω^2 = -0.011
LAC*YST*sex	$F_{1,39.40}$ = 0.3572, p = 0.5535, ω^2 = - 0.010	$F_{1,39.40}$ = 0.4912, p = 0.4875, ω^2 = -0.007	$F_{1,39.42}$ = 0.0219, p = 0.8830, ω^2 = -0.012	$F_{1,39.35}$ = 0.4627, p= 0.5004, ω^2 = -0.008
ACE*LAC*YST	$F_{1,39.93}$ = 0.4910, p = 0.4875, ω^2 = - 0.008	$F_{1,39.93}$ = 0.0063, p = 0.9370, ω^2 = -0.013	$F_{1,39.93}$ = 1.5815, p = 0.2158, ω^2 = 0.007	$F_{1,39.95}$ = 0.0252, p= 0.8746, ω^2 = -0.014
ACE*LAC*YST*sex	*$F_{1,39.42}$ = 3.6614, p = 0.0630, ω^2 = 0.041*	$F_{1,39.41}$ = 0.1656, p = 0.6863, ω^2 = -0.011	$F_{1,39.44}$ = 0.1699, p = 0.6824, ω^2 = -0.010	*$F_{1,39.36}$ = 2.0081, p= 0.1643, ω^2 = 0.015*
Analysis of Deviance				
Vial replicate	**X^2_1 = 14.40, p = 0.0001**	**X^2_1 = 13.20, p = 0.0003**	**X^2_1 = 12.88, p = 0.0003**	**X^2_1 = 19.23, p = 1.16 x 10^{-5}**
R^2				
Marginal	0.244	0.274	0.341	0.262
Conditional	0.586	0.649	0.662	0.695

The *post hoc* Tukey test was implemented to assess differences in SCFA content between treatments for each sex. Uppercase letters are used to represent the ranking of female samples and lowercase letters are used to distinguish the rankings for male samples. Significant p-values were bolded and influential effect sizes were italicized. P-values below the threshold are bolded.

Table S2.1H. Moderation effect of individual SCFA content and microbiota members on individual nutritional indices

i) TAG ~ Acetate		
ANOVA Effect test	***Female***	***Male***
ACE	$F_{1,29.11} = 7.954$, p = 0.0086	**$F_{1,30.06} = 55.596$, p = 2.6 x 10^{-8}**
LAC	$F_{1,29.02} = 0.166$, p = 0.6864	$F_{1,30.01} = 0.225$, p = 0.6390
YST	$F_{1,29.22} = 4.527$, p = 0.0419	**$F_{1,30.16} = 602.769$, p < 2.2 x 10^{-16}**
[Acetate]	$F_{1,30.22} = 0.455$, p = 0.5049	$F_{1,31.80} = 1.755$, p = 0.1948
ACE*[Acetate]	$F_{1,29.16} = 0.647$, p = 0.4276	**$F_{1,30.01} = 23.248$, p = 3.9 x 10^{-5}**
LAC*[Acetate]	$F_{1,29.36} = 1.212$, p = 0.2798	$F_{1,31.12} = 0.170$, p = 0.6831
YST*[Acetate]	$F_{1,29.22} = 0.271$, p = 0.6064	$F_{1,30.28} = 0.416$, p = 0.5240
ACE*LAC	$F_{1,29.03} = 0.146$, p = 0.7048	**$F_{1,30.06} = 12.789$, p = 0.0012**
ACE*YST	$F_{1,29.27} = 0.055$, p = 0.8166	**$F_{1,30.09} = 23.741$, p = 3.3 x 10^{-5}**
LAC*YST	$F_{1,29.07} = 1.623$, p = 0.2127	**$F_{1,30.42} = 21.530$, p = 6.2 x 10^{-5}**
ACE*LAC*[Acetate]	$F_{1,29.00} = 0.580$, p = 0.4524	$F_{1,30.86} = 4.630$, p = 0.0394
ACE*YST*[Acetate]	$F_{1,29.53} = 0.080$, p = 0.7797	**$F_{1,31.03} = 17.940$, p = 0.0002**
LAC*YST*[Acetate]	$F_{1,29.41} = 0.914$, p = 0.3468	$F_{1,30.10} = 1.793$, p = 0.1906
ACE*LAC*YST	$F_{1,29.06} = 0.727$, p = 0.4010	$F_{1,30.28} = 4.781$, p = 0.0366
ACE*LAC*YST*[Acetate]	$F_{1,29.56} = 2.103$, p = 0.1576	$F_{1,30.33} = 0.691$, p = 0.4125
Analysis of Deviance		
Experimental replicate	**$X^2_1 = 11.45$, p = 0.0007**	**$X^2_1 = 5.73$, p = 0.0167**
R^2		
Marginal	0.233	0.929
Conditional	0.532	0.946
ii) Glucose ~ Acetate		
ANOVA Effect test	***Female (ln-transformed)***	***Male (ln-transformed)***
ACE	$F_{1,29.42} = 2.988$, p = 0.0944	$F_{1,30.20} = 2.148$, p = 0.1531
LAC	$F_{1,29.08} = 2.958$, p = 0.0961	$F_{1,30.04} = 5.118$, p = 0.0311
YST	**$F_{1,29.75} = 69.513$, p = 2.8 x 10^{-9}**	**$F_{1,30.47} = 169.154$, p = 5.7 x 10^{-14}**
[Acetate]	$F_{1,30.63} = 0.792$, p = 0.3803	$F_{1,28.74} = 1.078$, p = 0.3079
ACE*[Acetate]	$F_{1,29.83} = 1.189$, p = 0.2844	$F_{1,30.06} = 0.018$, p = 0.8938
LAC*[Acetate]	$F_{1,30.24} = 0.445$, p = 0.5099	$F_{1,31.89} = 3.185$, p = 0.0839
YST*[Acetate]	$F_{1,30.06} = 0.033$, p = 0.8572	$F_{1,30.75} = 1.929$, p = 0.1748
ACE*LAC	$F_{1,29.19} = 2.767$, p = 0.1069	$F_{1,30.23} = 2.611$, p = 0.1165
ACE*YST	$F_{1,30.02} = 11.299$, p = 0.0021	$F_{1,30.31} = 4.688$, p = 0.0384
LAC*YST	$F_{1,29.38} = 0.859$, p = 0.3617	$F_{1,31.18} = 1.460$, p = 0.2361
ACE*LAC*[Acetate]	$F_{1,29.04} = 1.613$, p = 0.2142	$F_{1,31.90} = 0.027$, p = 0.8707
ACE*YST*[Acetate]	$F_{1,30.62} = 0.487$, p = 0.4907	$F_{1,32.00} = 1.281$, p = 0.2661
LAC*YST*[Acetate]	$F_{1,30.59} = 0.452$, p = 0.5062	$F_{1,30.32} = 0.012$, p = 0.9141
ACE*LAC*YST	$F_{1,29.28} = 0.135$, p = 0.7161	$F_{1,30.83} = 1.206$, p = 0.2806
ACE*LAC*YST*[Acetate]	$F_{1,30.78} = 2.505$, p = 0.1237	$F_{1,30.93} = 1.241$, p = 0.2739
Analysis of Deviance		
Experimental replicate	$X^2_1 = 0.034$, p = 0.854	**$X^2_1 = 35.47$, p = 2.6 x 10^{-9}**
R^2		
Marginal	0.684	0.814
Conditional	0.703	0.819

iii) Protein ~ Acetate		
ANOVA Effect test	***Female***	***Male***
ACE	$F_{1,29.65} = 0.191$, $p = 0.6650$	$F_{1,30.27} = 1.102$, $p = 0.3021$
LAC	$F_{1,29.16} = 0.003$, $p = 0.9564$	$F_{1,30.07} = 0.183$, $p = 0.6722$
YST	$F_{1,30.08} = 8.476$, $p = 0.0067$	$F_{1,30.61} = 2.782$, $p = 0.1055$
[Acetate]	$F_{1,27.86} = 1.729$, $p = 0.1992$	$F_{1,24.45} = 0.202$, $p = 0.6572$
ACE*[Acetate]	$F_{1,30.43} = 0.831$, $p = 0.3691$	$F_{1,30.09} = 0.378$, $p = 0.5434$
LAC*[Acetate]	$F_{1,30.68} = 0.772$, $p = 0.3864$	$F_{1,30.96} = 0.001$, $p = 0.9784$
YST*[Acetate]	$F_{1,30.65} = 3.378$, $p = 0.0758$	$F_{1,30.92} = 0.137$, $p = 0.7135$
ACE*LAC	$F_{1,29.37} = 0.206$, $p = 0.6537$	$F_{1,30.33} = 1.088$, $p = 0.3053$
ACE*YST	$F_{1,30.51} = 1.043$, $p = 0.3153$	$F_{1,30.43} = 1.471$, $p = 0.2345$
LAC*YST	$F_{1,29.69} = 0.768$, $p = 0.3880$	$F_{1,31.44} = 0.229$, $p = 0.6353$
ACE*LAC*[Acetate]	$F_{1,29.10} = 0.530$, $p = 0.4723$	$F_{1,32.00} = 1.453$, $p = 0.2368$
ACE*YST*[Acetate]	$F_{1,30.96} = 1.741$, $p = 0.1967$	$F_{1,31.69} = 0.113$, $p = 0.7395$
LAC*YST*[Acetate]	$F_{1,31.00} = 0.903$, $p = 0.3495$	$F_{1,30.41} = 0.052$, $p = 0.8218$
ACE*LAC*YST	$F_{1,29.47} = 0.040$, $p = 0.8429$	$F_{1,31.04} = 0.225$, $p = 0.6384$
ACE*LAC*YST*[Acetate]	$F_{1,30.99} = 0.374$, $p = 0.5452$	$F_{1,31.13} = 0.924$, $p = 0.3440$
Analysis of Deviance		
Experimental replicate	$X^2_1 = 0$, $p = 1$	$X^2_1 = 0$, $p = 1$
R^2		
Marginal	0.346	0.201
Conditional	0.346	0.201
iv) Glycogen ~ Acetate		
ANOVA Effect test	***Female***	***Male***
ACE	$F_{1,29.65} = 2.034$, $p = 0.1643$	$F_{1,30.06} = 5.552$, $p = 0.0252$
LAC	$F_{1,29.16} = 1.837$, $p = 0.1857$	$F_{1,30.01} = 0.206$, $p = 0.6533$
YST	**$F_{1,30.08} = 13.687$, $p = 0.0009$**	$F_{1,30.18} = 1.836$, $p = 0.1855$
[Acetate]	$F_{1,27.86} = 0.483$, $p = 0.4929$	$F_{1,31.88} = 0.374$, $p = 0.5454$
ACE*[Acetate]	$F_{1,30.43} = 0.191$, $p = 0.6650$	$F_{1,30.02} = 0.212$, $p = 0.6488$
LAC*[Acetate]	$F_{1,30.68} = 0.252$, $p = 0.6194$	$F_{1,31.21} = 2.913$, $p = 0.0978$
YST*[Acetate]	$F_{1,30.65} = 0.339$, $p = 0.5645$	$F_{1,\ 30.31} = 0.014$, $p = 0.9053$
ACE*LAC	$F_{1,29.37} = 0.192$, $p = 0.6647$	$F_{1,30.07} = 1.364$, $p = 0.2520$
ACE*YST	$F_{1,30.51} = 3.616$, $p = 0.0667$	$F_{1,30.10} = 1.797$, $p = 0.1901$
LAC*YST	$F_{1,29.69} = 2.921$, $p = 0.0978$	$F_{1,30.46} = 2.026$, $p = 0.1648$
ACE*LAC*[Acetate]	$F_{1,29.10} = 0.199$, $p = 0.6586$	$F_{1,30.94} = 0.167$, $p = 0.6859$
ACE*YST*[Acetate]	$F_{1,30.96} = 0.447$, $p = 0.5089$	$F_{1,\ 31.12} = 0.364$, $p = 0.5508$
LAC*YST*[Acetate]	$F_{1,31.00} = 1.343$, $p = 0.2553$	$F_{1,30.11} = 0.716$, $p = 0.4040$
ACE*LAC*YST	$F_{1,29.47} = 0.001$, $p = 0.9824$	$F_{1,30.31} = 0.613$, $p = 0.4398$
ACE*LAC*YST*[Acetate]	$F_{1,30.99} = 0.691$, $p = 0.4121$	$F_{1,30.36} = 2.029$, $p = 0.1645$
Analysis of Deviance		
Experimental replicate	$X^2_1 = 0$, $p = 1$	**$X^2_1 = 3.96$, $p = 0.0466$**
R^2		
Marginal	0.446	0.281
Conditional	0.446	0.430

v) Trehalose ~ Acetate		
ANOVA Effect test	***Female (ln-transformed)***	***Male (ln-transformed)***
ACE	$F_{1,29.39} = 0.816$, p = 0.3738	$F_{1,30.27} = 0.040$, p = 0.8436
LAC	$F_{1,29.08} = 1.384$, p = 0.2489	$F_{1,30.07} = 0.194$, p = 0.6632
YST	**$F_{1,29.72} = 25.518$, p = 2.1 x 10^{-5}**	**$F_{1,30.61} = 14.018$, p = 0.0008**
[Acetate]	$F_{1,30.75} = 9.706$, p = 0.0040	$F_{1,24.45} = 0.255$, p = 0.6183
ACE*[Acetate]	$F_{1,29.77} = 7.690$, p = 0.0095	$F_{1,30.09} = 0.316$, p = 0.5784
LAC*[Acetate]	$F_{1,30.19} = 9.719$, p = 0.0040	$F_{1,30.96} = 0.152$, p = 0.6990
YST*[Acetate]	$F_{1,29.99} = 0.499$, p = 0.4854	$F_{1,30.92} = 0.065$, p = 0.8006
ACE*LAC	$F_{1,29.17} = 3.069$, p = 0.0903	$F_{1,30.33} = 1.165$, p = 0.2890
ACE*YST	$F_{1,29.97} = 6.471$, p = 0.0164	$F_{1,30.43} = 0.599$, p = 0.4450
LAC*YST	$F_{1,29.35} = 1.799$, p = 0.1902	$F_{1,31.44} = 7.366$, p = 0.0107
ACE*LAC*[Acetate]	$F_{1,29.04} = 2.895$, p = 0.0995	$F_{1,32.00} = 0.356$, p = 0.5551
ACE*YST*[Acetate]	$F_{1,30.57} = 1.088$, p = 0.3051	$F_{1,31.69} = 0.092$, p = 0.7635
LAC*YST*[Acetate]	$F_{1,30.52} = 0.407$, p = 0.5285	$F_{1,30.41} = 0.107$, p = 0.7456
ACE*LAC*YST	$F_{1,29.26} = 0.129$, p = 0.7219	$F_{1,31.04} = 0.442$, p = 0.5110
ACE*LAC*YST*[Acetate]	$F_{1,30.72} = 0.005$, p = 0.9420	$F_{1,31.13} = 1.733$, p = 0.1977
Analysis of Deviance		
Experimental replicate	$X^2_1 = 0.25$, p = 0.6163	$X^2_1 = 0$, p = 1
R^2		
Marginal	0.585	0.394
Conditional	0.614	0.394
vi) TAG ~ Propionate		
ANOVA Effect test	***Female***	***Male (sqrt-transformed)***
ACE	$F_{1,29.03} = 6.032$, p = 0.0203	**$F_{1,30.06} = 30.458$, p = 5.3 x 10^{-6}**
LAC	$F_{1,29.11} = 0.060$, p = 0.8077	$F_{1,30.02} = 3.124$, p = 0.0873
YST	$F_{1,29.10} = 2.672$, p = 0.1129	**$F_{1,30.26} = 397.131$, p > 2.2 x 10^{-16}**
[Propionate]	$F_{1,30.25} = 1.034$, p = 0.3173	$F_{1,31.33} = 0.054$, p = 0.8172
ACE*[Propionate]	$F_{1,29.46} = 0.599$, p = 0.4451	$F_{1,30.26} = 7.695$, p = 0.0094
LAC*[Propionate]	$F_{1,29.27} = 0.001$, p = 0.9716	$F_{1,30.75} = 0.010$, p = 0.9207
YST*[Propionate]	$F_{1,29.81} = 0.536$, p = 0.4699	$F_{1,31.17} = 1.032$, p = 0.3174
ACE*LAC	$F_{1,29.16} = 0.0002$, p = 0.9877	$F_{1,30.02} = 7.213$, p = 0.0117
ACE*YST	$F_{1,29.28} = 0.775$, p = 0.3857	$F_{1,30.13} = 3.546$, p = 0.0694
LAC*YST	$F_{1,29.03} = 1.431$, p = 0.2413	$F_{1,30.26} = 11.174$, p = 0.0022
ACE*LAC*[Propionate]	$F_{1,29.79} = 1.349$, p = 0.2548	$F_{1,30.70} = 3.355$, p = 0.0767
ACE*YST*[Propionate]	$F_{1,29.46} = 0.033$, p = 0.8582	$F_{1,30.97} = 1.043$, p = 0.3151
LAC*YST*[Propionate]	$F_{1,30.63} = 0.551$, p = 0.4637	$F_{1,31.32} = 0.037$, p = 0.8484
ACE*LAC*YST	$F_{1,29.34} = 0.514$, p = 0.4790	$F_{1,30.40} = 1.043$, p = 0.3153
ACE*LAC*YST*[Propionate]	$F_{1,29.28} = 0.040$, p = 0.8425	$F_{1,30.06} = 0.104$, p = 0.7495
Analysis of Deviance		
Experimental replicate	**$X^2_1 = 5.47$, p = 0.0193**	**$X^2_1 = 6.20$, p = 0.0128**
R^2		
Marginal	0.244	0.898
Conditional	0.429	0.924

vii) Glucose ~ Propionate		
ANOVA Effect test	***Female***	***Male (sqrt-transformed)***
ACE	$F_{1,29.16} = 0.135$, p = 0.7159	$F_{1,30.08} = 0.648$, p = 0.4273
LAC	$F_{1,29.36} = 4.994$, p = 0.0332	$F_{1,30.03} = 8.688$, p = 0.0061
YST	**$F_{1,29.44} = 43.147$, p = 3.1 x 10^{-7}**	**$F_{1,30.35} = 137.028$, p = 8.8 x 10^{-13}**
[Propionate]	$F_{1,29.74} = 1.852$, p = 0.1837	$F_{1,31.65} = 0.813$, p = 0.3741
ACE*[Propionate]	$F_{1,30.79} = 0.002$, p = 0.9679	$F_{1,30.35} = 0.017$, p = 0.8968
LAC*[Propionate]	$F_{1,\ 30.04} = 0.816$, p = 0.3734	$F_{1,31.00} = 0.414$, p = 0.5249
YST*[Propionate]	$F_{1,30.98} = 2.409$, p = 0.1308	$F_{1,31.51} = 1.167$, p = 0.2882
ACE*LAC	$F_{1,29.56} = 4.224$, p = 0.0488	$F_{1,30.03} = 1.327$, p = 0.2585
ACE*YST	$F_{1,30.18} = 11.623$, p = 0.0019	$F_{1,30.18} = 2.251$, p = 0.1439
LAC*YST	$F_{1,29.11} = 2.322$, p = 0.1384	$F_{1,30.35} = 4.411$, p = 0.0441
ACE*LAC*[Propionate]	$F_{1,31.00} = 0.366$, p = 0.5497	$F_{1,30.93} = 0.210$, p = 0.6499
ACE*YST*[Propionate]	$F_{1,30.53} = 2.592$, p = 0.1177	$F_{1,31.29} = 0.063$, p = 0.8035
LAC*YST*[Propionate]	$F_{1,26.25} = 1.055$, p = 0.3137	$F_{1,31.69} = 0.000$, p = 0.9971
ACE*LAC*YST	$F_{1,30.34} = 0.417$, p = 0.5235	$F_{1,30.55} = 0.109$, p = 0.7435
ACE*LAC*YST*[Propionate]	$F_{1,30.34} = 1.193$, p = 0.2833	$F_{1,30.09} = 0.017$, p = 0.8983
Analysis of Deviance		
Experimental replicate	$X^2_1 = 0$, p = 1	$X^2_1 = 3.00$, p = 0.0834
R^2		
Marginal	0.628	0.768
Conditional	0.628	0.810
viii) Protein ~ Propionate		
ANOVA Effect test	***Female***	***Male***
ACE	$F_{1,29.16} = 0.388$, p = 0.5381	$F_{1,30.26} = 1.475$, p = 0.2339
LAC	$F_{1,29.36} = 0.054$, p = 0.8187	$F_{1,30.09} = 0.203$, p = 0.6557
YST	**$F_{1,29.44} = 14.851$, p = 0.0006**	$F_{1,30.95} = 3.981$, p = 0.0549
[Propionate]	$F_{1,29.74} = 0.654$, p = 0.4253	$F_{1,30.65} = 0.270$, p = 0.6070
ACE*[Propionate]	$F_{1,30.79} = 0.817$, p = 0.3730	$F_{1,31.08} = 1.252$, p = 0.2718
LAC*[Propionate]	$F_{1,30.04} = 2.740$, p = 0.1083	$F_{1,32.00} = 0.357$, p = 0.5543
YST*[Propionate]	$F_{1,30.98} = 11.055$, p = 0.0023	$F_{1,30.52} = 0.774$, p = 0.3859
ACE*LAC	$F_{1,29.56} = 0.412$, p = 0.5258	$F_{1,30.15} = 1.138$, p = 0.2946
ACE*YST	$F_{1,30.18} = 2.117$, p = 0.1560	$F_{1,30.61} = 0.471$, p = 0.4976
LAC*YST	$F_{1,29.11} = 1.518$, p = 0.2278	$F_{1,31.03} = 0.101$, p = 0.7522
ACE*LAC*[Propionate]	$F_{1,31.00} = 1.367$, p = 0.2513	$F_{1,32.00} = 2.610$, p = 0.1160
ACE*YST*[Propionate]	$F_{1,30.53} = 5.631$, p = 0.0241	$F_{1,30.68} = 0.788$, p = 0.3817
LAC*YST*[Propionate]	$F_{1,26.25} = 2.470$, p = 0.1280	$F_{1,26.45} = 0.300$, p = 0.5887
ACE*LAC*YST	$F_{1,30.34} = 0.530$, p = 0.4720	$F_{1,31.60} = 0.303$, p = 0.5861
ACE*LAC*YST*[Propionate]	$F_{1,30.34} = 1.519$, p = 0.2273	$F_{1,30.35} = 0.746$, p = 0.3945
Analysis of Deviance		
Experimental replicate	$X^2_1 = 0$, p = 1	$X^2_1 = 0$, p = 1
R^2		
Marginal	0.457	0.256
Conditional	0.457	0.256

ix) Glycogen ~ Propionate		
ANOVA Effect test	***Female***	***Male***
ACE	$F_{1,29.16} = 3.263$, $p = 0.0812$	$F_{1,30.05} = 6.715$, $p = 0.0146$
LAC	$F_{1,29.36} = 1.484$, $p = 0.2328$	$F_{1,30.02} = 0.217$, $p = 0.6447$
YST	**$F_{1,29.44} = 18.380$, $p = 0.0002$**	$F_{1,30.21} = 1.252$, $p = 0.2721$
[Propionate]	$F_{1,29.74} = 0.284$, $p = 0.5981$	$F_{1,31.14} = 1.170$, $p = 0.2877$
ACE*[Propionate]	$F_{1,30.79} = 0.145$, $p = 0.7062$	$F_{1,30.21} = 0.967$, $p = 0.3334$
LAC*[Propionate]	$F_{1,30.04} = 0.623$, $p = 0.4361$	$F_{1,30.62} = 3.203$, $p = 0.0834$
YST*[Propionate]	$F_{1,30.98} = 1.601$, $p = 0.2152$	$F_{1,30.98} = 0.071$, $p = 0.7918$
ACE*LAC	$F_{1,29.56} = 0.571$, $p = 0.4560$	$F_{1,30.02} = 1.667$, $p = 0.2065$
ACE*YST	$F_{1,30.18} = 3.613$, $p = 0.0669$	$F_{1,30.11} = 2.313$, $p = 0.1388$
LAC*YST	$F_{1,29.11} = 4.109$, $p = 0.0519$	$F_{1,30.21} = 2.701$, $p = 0.1107$
ACE*LAC*[Propionate]	$F_{1,31.00} = 0.325$, $p = 0.5730$	$F_{1,30.57} = 1.426$, $p = 0.2416$
ACE*YST*[Propionate]	$F_{1,30.53} = 0.340$, $p = 0.5643$	$F_{1,30.79} = 0.158$, $p = 0.6941$
LAC*YST*[Propionate]	$F_{1,26.25} = 0.736$, $p = 0.3989$	$F_{1,31.10} = 0.009$, $p = 0.9268$
ACE*LAC*YST	$F_{1,30.34} = 0.011$, $p = 0.9182$	$F_{1,30.33} = 0.189$, $p = 0.6667$
ACE*LAC*YST*[Propionate]	$F_{1,30.34} = 2.587$, $p = 0.1181$	$F_{1,30.05} = 1.934$, $p = 0.1746$
Analysis of Deviance		
Experimental replicate	$X^2_1 = 0$, $p = 1$	**$X^2_1 = 7.56$, $p = 0.0060$**
R^2		
Marginal	0.483	0.279
Conditional	0.483	0.500
x) Trehalose ~ Propionate		
ANOVA Effect test	***Female (ln-transformed)***	***Male (ln-transformed)***
ACE	$F_{1,29.07} = 0.007$, $p = 0.9320$	$F_{1,30.26} = 0.085$, $p = 0.7730$
LAC	$F_{1,29.20} = 1.861$, $p = 0.1829$	$F_{1,30.09} = 0.362$, $p = 0.5519$
YST	$F_{1,29.20} = 11.328$, $p = 0.0022$	$F_{1,30.95} = 11.648$, $p = 0.0018$
[Propionate]	$F_{1,30.93} = 2.714$, $p = 0.1096$	$F_{1,30.65} = 0.038$, $p = 0.8464$
ACE*[Propionate]	$F_{1,29.93} = 3.319$, $p = 0.0785$	$F_{1,31.08} = 0.131$, $p = 0.7198$
LAC*[Propionate]	$F_{1,29.52} = 0.341$, $p = 0.5639$	$F_{1,32.00} = 0.075$, $p = 0.7858$
YST*[Propionate]	$F_{1,30.40} = 2.145$, $p = 0.1533$	$F_{1,30.52} = 0.009$, $p = 0.9257$
ACE*LAC	$F_{1,29.30} = 2.671$, $p = 0.1129$	$F_{1,30.15} = 0.868$, $p = 0.3589$
ACE*YST	$F_{1,29.58} = 1.745$, $p = 0.1967$	$F_{1,30.61} = 0.939$, $p = 0.3401$
LAC*YST	$F_{1,29.05} = 4.912$, $p = 0.0347$	$F_{1,31.03} = 6.807$, $p = 0.0139$
ACE*LAC*[Propionate]	$F_{1,30.42} = 1.177$, $p = 0.2864$	$F_{1,32.00} = 1.260$, $p = 0.2700$
ACE*YST*[Propionate]	$F_{1,29.88} = 0.032$, $p = 0.8602$	$F_{1,30.68} = 0.013$, $p = 0.9117$
LAC*YST*[Propionate]	$F_{1,30.91} = 2.222$, $p = 0.1462$	$F_{1,26.45} = 0.703$, $p = 0.4092$
ACE*LAC*YST	$F_{1,29.68} = 0.838$, $p = 0.3674$	$F_{1,31.60} = 1.205$, $p = 0.2806$
ACE*LAC*YST*[Propionate]	$F_{1,29.60} = 0.007$, $p = 0.9363$	$F_{1,30.35} = 0.000$, $p = 0.9897$
Analysis of Deviance		
Experimental replicate	$X^2_1 = 0.61$, $p = 0.4345$	$X^2_1 = 0$, $p = 1$
R^2		
Marginal	0.429	0.399
Conditional	0.485	0.399

xi) TAG ~ Butyrate		
ANOVA Effect test	***Female***	***Male***
ACE	$F_{1,29.02} = 11.610$, p = 0.0019	**$F_{1,30.02} = 28.194$, p = 9.7 x 10^{-6}**
LAC	$F_{1,29.03} = 0.091$, p = 0.7650	$F_{1,31.05} = 0.468$, p = 0.4990
YST	$F_{1,29.05} = 7.143$, p = 0.0122	**$F_{1,31.72} = 168.330$, p = 3.1 x 10^{-14}**
[Butyrate]	$F_{1,29.98} = 1.113$, p = 0.2999	$F_{1,29.95} = 0.066$, p = 0.7994
ACE*[Butyrate]	$F_{1,29.12} = 1.895$, p = 0.1792	$F_{1,30.04} = 0.435$, p = 0.5146
LAC*[Butyrate]	$F_{1,29.01} = 0.155$, p = 0.6965	$F_{1,30.33} = 0.032$, p = 0.8584
YST*[Butyrate]	$F_{1,29.16} = 0.912$, p = 0.3476	$F_{1,31.04} = 1.134$, p = 0.2951
ACE*LAC	$F_{1,29.08} = 0.182$, p = 0.6727	$F_{1,30.23} = 4.330$, p = 0.0460
ACE*YST	$F_{1,29.01} = 1.257$, p = 0.2714	$F_{1,30.06} = 9.371$, p = 0.0046
LAC*YST	$F_{1,29.01} = 0.030$, p = 0.8637	$F_{1,30.08} = 2.292$, p = 0.1405
ACE*LAC*[Butyrate]	$F_{1,30.25} = 0.472$, p = 0.4973	$F_{1,31.78} = 1.770$, p = 0.1929
ACE*YST*[Butyrate]	$F_{1,29.02} = 6.957$, p = 0.0133	$F_{1,31.32} = 0.867$, p = 0.3588
LAC*YST*[Butyrate]	$F_{1,29.14} = 1.651$, p = 0.2090	$F_{1,30.51} = 2.824$, p = 0.1031
ACE*LAC*YST	$F_{1,29.16} = 0.004$, p = 0.9500	$F_{1,31.97} = 5.565$, p = 0.0246
ACE*LAC*YST*[Butyrate]	$F_{1,\ 29.15} = 0.236$, p = 0.6305	$F_{1,31.41} = 0.186$, p = 0.6695
Analysis of Deviance		
Experimental replicate	**$X^2_1 = 12.84$, p = 0.0003**	$X^2_1 = 0$, p = 0.7386
R^2		
Marginal	0.281	0.879
Conditional	0.590	0.887
xii) Glucose ~ Butyrate		
ANOVA Effect test	***Female***	***Male***
ACE	$F_{1,29.06} = 0.112$, p = 0.7404	$F_{1,30.01} = 0.200$, p = 0.6581
LAC	$F_{1,29.09} = 2.510$, p = 0.1239	$F_{1,30.73} = 10.487$, p = 0.0029
YST	**$F_{1,29.16} = 41.598$, p = 4.6 x 10^{-7}**	**$F_{1,31.34} = 67.071$, p = 2.8 x 10^{-9}**
[Butyrate]	$F_{1,30.99} = 0.186$, p = 0.6690	$F_{1,31.76} = 0.992$, p = 0.3268
ACE*[Butyrate]	$F_{1,29.42} = 0.709$, p = 0.4065	$F_{1,30.02} = 1.355$, p = 0.2536
LAC*[Butyrate]	$F_{1,29.05} = 1.072$, p = 0.3090	$F_{1,30.22} = 0.793$, p = 0.3804
YST*[Butyrate]	$F_{1,29.52} = 1.912$, p = 0.1772	$F_{1,31.99} = 3.458$, p = 0.0722
ACE*LAC	$F_{1,29.29} = 1.056$, p = 0.3126	$F_{1,30.16} = 0.001$, p = 0.9718
ACE*YST	$F_{1,29.04} = 10.692$, p = 0.0028	$F_{1,30.04} = 1.670$, p = 0.2061
LAC*YST	$F_{1,29.04} = 0.479$, p = 0.4944	$F_{1,30.05} = 1.884$, p = 0.1800
ACE*LAC*[Butyrate]	$F_{1,30.38} = 1.820$, p = 0.1873	$F_{1,31.94} = 0.248$, p = 0.6222
ACE*YST*[Butyrate]	$F_{1,29.08} = 0.000$, p = 0.9997	$F_{1,30.92} = 2.975$, p = 0.0945
LAC*YST*[Butyrate]	$F_{1,29.49} = 0.443$, p = 0.5108	$F_{1,30.35} = 2.719$, p = 0.1095
ACE*LAC*YST	$F_{1,29.54} = 0.016$, p = 0.8997	$F_{1,31.83} = 0.004$, p = 0.9522
ACE*LAC*YST*[Butyrate]	$F_{1,29.51} = 0.307$, p = 0.5839	$F_{1,30.94} = 0.466$, p = 0.5000
Analysis of Deviance		
Experimental replicate	$X^2_1 = 1.49$, p = 0.2217	$X^2_1 = 1.67$, p = 0.1965
R^2		
Marginal	0.612	0.759
Conditional	0.661	0.793

xiii) Protein ~ Butyrate		
ANOVA Effect test	***Female***	***Male***
ACE	$F_{1,29.16} = 0.140$, $p = 0.7111$	$F_{1,30.06} = 0.098$, $p = 0.7563$
LAC	$F_{1,29.23} = 0.001$, $p = 0.9729$	$F_{1,31.53} = 0.360$, $p = 0.5527$
YST	$F_{1,29.38} = 11.624$, $p = 0.0019$	$F_{1,32.00} = 0.328$, $p = 0.5706$
[Butyrate]	$F_{1,24.74} = 0.723$, $p = 0.4035$	$F_{1,22.88} = 2.488$, $p = 0.1285$
ACE*[Butyrate]	$F_{1,30.03} = 0.006$, $p = 0.9371$	$F_{1,30.08} = 0.751$, $p = 0.3931$
LAC*[Butyrate]	$F_{1,29.12} = 0.610$, $p = 0.4413$	$F_{1,30.52} = 1.053$, $p = 0.3129$
YST*[Butyrate]	$F_{1,30.17} = 2.140$, $p = 0.1539$	$F_{1,25.48} = 0.895$, $p = 0.3531$
ACE*LAC	$F_{1,29.73} = 0.017$, $p = 0.8960$	$F_{1,30.36} = 0.618$, $p = 0.4378$
ACE*YST	$F_{1,29.10} = 0.198$, $p = 0.6596$	$F_{1,30.12} = 1.896$, $p = 0.1787$
LAC*YST	$F_{1,29.12} = 0.609$, $p = 0.4416$	$F_{1,30.16} = 0.279$, $p = 0.6012$
ACE*LAC*[Butyrate]	$F_{1,14.72} = 0.616$, $p = 0.4450$	$F_{1,28.35} = 0.029$, $p = 0.8669$
ACE*YST*[Butyrate]	$F_{1,29.22} = 0.745$, $p = 0.3952$	$F_{1,31.85} = 0.003$, $p = 0.9545$
LAC*YST*[Butyrate]	$F_{1,30.25} = 0.008$, $p = 0.9290$	$F_{1,30.78} = 0.680$, $p = 0.4159$
ACE*LAC*YST	$F_{1,30.24} = 0.004$, $p = 0.9476$	$F_{1,30.21} = 0.134$, $p = 0.7168$
ACE*LAC*YST*[Butyrate]	$F_{1,30.17} = 0.052$, $p = 0.8217$	$F_{1,31.99} = 0.004$, $p = 0.9533$
Analysis of Deviance		
Experimental replicate	$X^2_1 = 0$, $p = 1$	$X^2_1 = 0$, $p = 1$
R^2		
Marginal	0.285	0.272
Conditional	0.285	0.272
xiv) Glycogen ~ Butyrate		
ANOVA Effect test	***Female***	***Male***
ACE	$F_{1,29.08} = 3.886$, $p = 0.0583$	$F_{1,30.01} = 6.741$, $p = 0.0144$
LAC	$F_{1,29.12} = 1.296$, $p = 0.2642$	$F_{1,30.37} = 0.024$, $p = 0.8788$
YST	**$F_{1,29.21} = 23.094$, $p = 4.3 \times 10^{-5}$**	$F_{1,30.76} = 0.816$, $p = 0.3735$
[Butyrate]	$F_{1,30.57} = 6.287$, $p = 0.0177$	$F_{1,31.73} = 2.098$, $p = 0.1574$
ACE*[Butyrate]	$F_{1,29.56} = 0.005$, $p = 0.9418$	$F_{1,30.01} = 1.882$, $p = 0.1803$
LAC*[Butyrate]	$F_{1,29.06} = 1.963$, $p = 0.1718$	$F_{1,30.10} = 0.098$, $p = 0.7569$
YST*[Butyrate]	$F_{1,29.68} = 2.146$, $p = 0.1534$	$F_{1,31.46} = 0.008$, $p = 0.9286$
ACE*LAC	$F_{1,29.39} = 3.780$, $p = 0.0615$	$F_{1,30.08} = 1.299$, $p = 0.2635$
ACE*YST	$F_{1,29.05} = 7.223$, $p = 0.0118$	$F_{1,30.02} = 2.084$, $p = 0.1593$
LAC*YST	$F_{1,29.06} = 8.440$, $p = 0.0070$	$F_{1,30.02} = 0.101$, $p = 0.7526$
ACE*LAC*[Butyrate]	$F_{1,28.41} = 1.487$, $p = 0.2327$	$F_{1,31.20} = 0.867$, $p = 0.3588$
ACE*YST*[Butyrate]	$F_{1,29.11} = 1.495$, $p = 0.2313$	$F_{1,30.46} = 0.100$, $p = 0.7543$
LAC*YST*[Butyrate]	$F_{1,29.65} = 1.465$, $p = 0.2357$	$F_{1,30.17} = 0.053$, $p = 0.8199$
ACE*LAC*YST	$F_{1,29.70} = 0.231$, $p = 0.6341$	$F_{1,31.07} = 0.048$, $p = 0.8273$
ACE*LAC*YST*[Butyrate]	$F_{1,29.66} = 0.060$, $p = 0.8088$	$F_{1,30.44} = 0.031$, $p = 0.8622$
Analysis of Deviance		
Experimental replicate	$X^2_1 = 0.46$, $p = 0.4963$	**$X^2_1 = 6.35$, $p = 0.0118$**
R^2		
Marginal	0.570	0.228
Conditional	0.604	0.470

xv) Trehalose ~ Butyrate		
ANOVA Effect test	***Female (sqrt-transformed)***	***Male (sqrt-transformed)***
ACE	$F_{1,29.16} = 0.000$, $p = 0.9847$	$F_{1,30.06} = 0.150$, $p = 0.7013$
LAC	$F_{1,29.23} = 1.376$, $p = 0.2503$	$F_{1,31.53} = 0.464$, $p = 0.5007$
YST	$F_{1,29.38} = 6.001$, $p = 0.0205$	$F_{1,32.00} = 2.776$, $p = 0.1055$
[Butyrate]	$F_{1,24.74} = 0.009$, $p = 0.9263$	$F_{1,22.88} = 0.283$, $p = 0.6000$
ACE*[Butyrate]	$F_{1,30.03} = 0.639$, $p = 0.4305$	$F_{1,30.08} = 0.025$, $p = 0.8766$
LAC*[Butyrate]	$F_{1,29.12} = 0.005$, $p = 0.9440$	$F_{1,30.52} = 1.568$, $p = 0.2201$
YST*[Butyrate]	$F_{1,30.17} = 0.310$, $p = 0.5817$	$F_{1,25.48} = 0.210$, $p = 0.6510$
ACE*LAC	$F_{1,29.73} = 4.091$, $p = 0.0522$	$F_{1,30.36} = 1.358$, $p = 0.2530$
ACE*YST	$F_{1,29.10} = 0.135$, $p = 0.7163$	$F_{1,30.12} = 0.363$, $p = 0.5515$
LAC*YST	$F_{1,29.12} = 5.802$, $p = 0.0226$	$F_{1,30.16} = 9.627$, $p = 0.0041$
ACE*LAC*[Butyrate]	$F_{1,14.72} = 0.285$, $p = 0.6017$	$F_{1,28.35} = 0.086$, $p = 0.7711$
ACE*YST*[Butyrate]	$F_{1,29.22} = 0.205$, $p = 0.6542$	$F_{1,31.85} = 0.184$, $p = 0.6707$
LAC*YST*[Butyrate]	$F_{1,30.25} = 0.027$, $p = 0.8699$	$F_{1,30.78} = 0.165$, $p = 0.6875$
ACE*LAC*YST	$F_{1,30.24} = 0.131$, $p = 0.7197$	$F_{1,30.21} = 0.014$, $p = 0.9070$
ACE*LAC*YST*[Butyrate]	$F_{1,30.17} = 0.143$, $p = 0.7078$	$F_{1,31.99} = 0.081$, $p = 0.7778$
Analysis of Deviance		
Experimental replicate	$X^2_1 = 0$, $p = 1$	$X^2_1 = 0$, $p = 1$
R^2		
Marginal	0.344	0.372
Conditional	0.344	0.372

P-values below the threshold are in bold.

APPENDIX C

SUPPLEMENTAL MATERIAL FOR CHAPTER 4

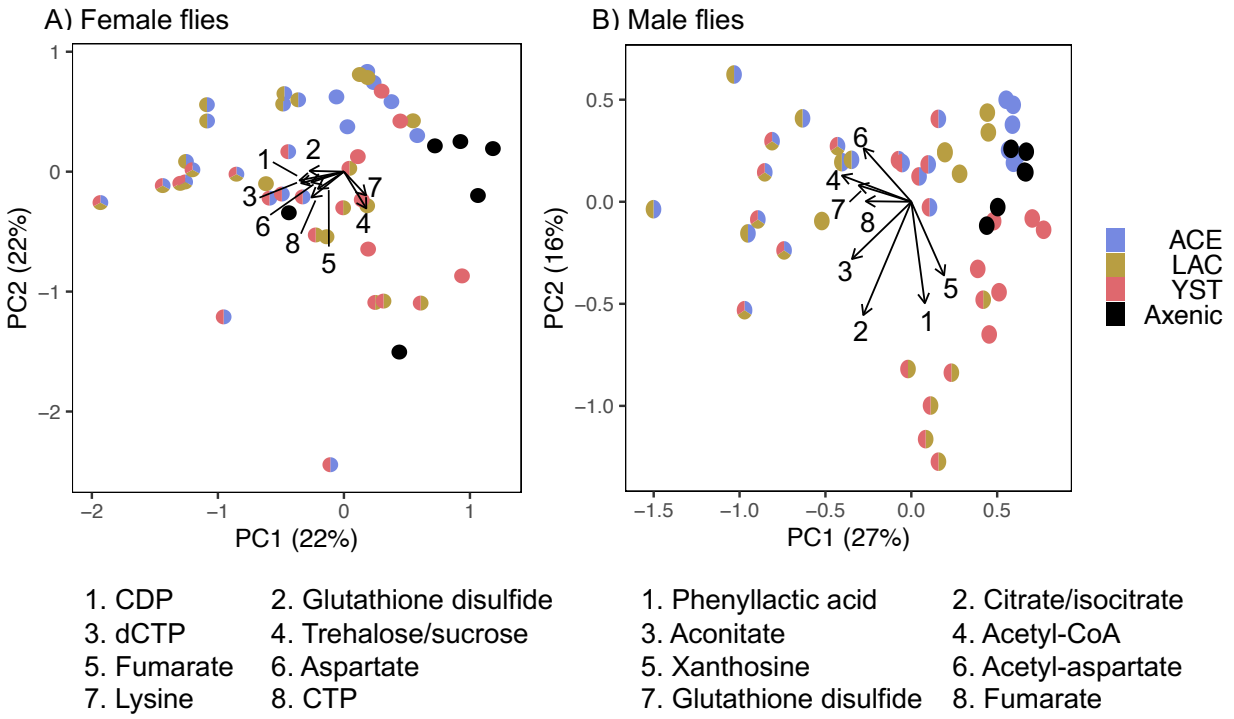


Figure S3.1. Metabolites associated with among-treatment differences. PLS-DA of (A) female and (B) male flies. Points are colored according to microbial treatment. Arrows represents the loadings with VIP scores > 1; only 8 of the 17 loadings are shown for female flies. The percent variation explained on each axis is displayed. Female PLS-DA: $R^2 = 0.17$, $Q^2 = 0.10$, RMSEE = 0.31, R^2 p = 0.001, Q^2 p = 0.001. Male PLS-DA: $R^2 = 0.23$, $Q^2 = 0.20$, RMSEE = 0.30, R^2 p = 0.001, Q^2 p = 0.001.

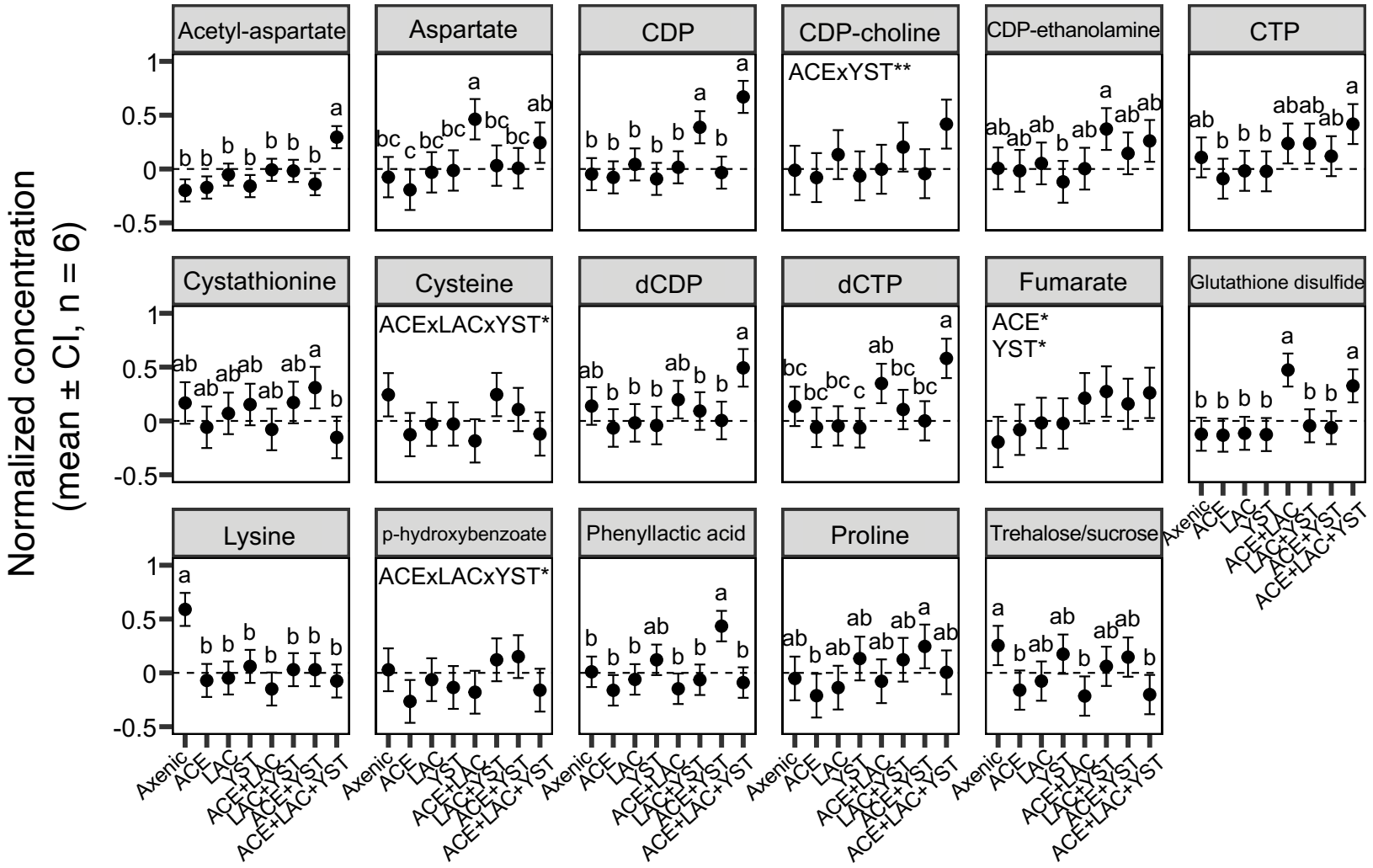

Normalized concentration
(mean ± CI, n = 6)
Acetyl-aspartate
Aspartate
CDP
CDP-choline
ACExYST**
CDP-ethanolamine
CTP
Cystathionine
Cysteine
ACExLACxYST*
dCDP
dCTP
Fumarate
ACE*
YST*
Glutathione disulfide
Lysine
p-hydroxybenzoate
ACExLACxYST*
Phenyllactic acid
Proline
Trehalose/sucrose
1
0.5
0
-0.5
Axenic
ACE
LAC
YST
ACE+LAC
ACE+YST
LAC+YST
ACE+LAC+YST

Figure S3.2. Effect of microorganisms and co-associations on metabolite concentration. (A) Female flies. Metabolites (range scaled concentrations) with q value < 0.2 and VIP score > 1 were analyzed by *post hoc* Tukey test results (letter rankings displayed). When there was not enough power to identify differences among-treatments, a full factorial ANOVA was implemented to identify treatment differences (only significant terms displayed). The estimated marginal means and 95% confidence intervals are displayed from ANOVA models. The dashed line indicates the grand mean for metabolite concentration and treatments with non-overlapping confidence intervals represent significant deviations from average. * $p < 0.05$; ** $p < 0.01$.

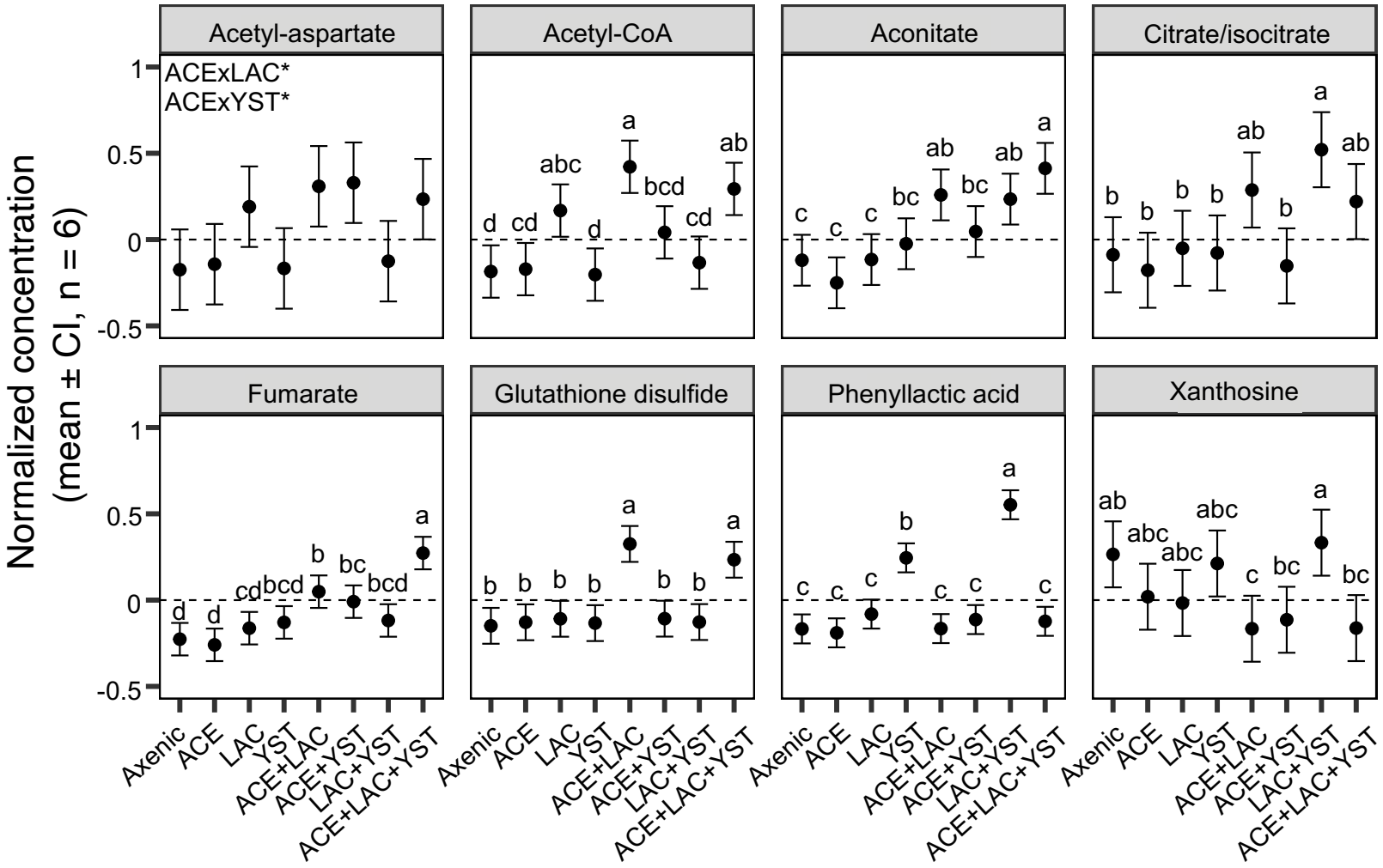

Acetyl-aspartate
Acetyl-CoA
Aconitate
Citrate/isocitrate
Fumarate
Glutathione disulfide
Phenyllactic acid
Xanthosine
ACExLAC*
ACExYST*
Normalized concentration
(mean ± CI, n = 6)
1
0.5
0
-0.5
Axenic
ACE
LAC
YST
ACE+LAC
ACE+YST
LAC+YST
ACE+LAC+YST

Figure S3.2. Effect of microorganisms and co-associations on metabolite concentration. (B) Male flies. Metabolites (range scaled concentrations) with q value < 0.2 and VIP score > 1 were analyzed by *post hoc* Tukey test results (letter rankings displayed). When there was not enough power to identify differences among-treatments, a full factorial ANOVA was implemented to identify treatment differences (only significant terms displayed). The estimated marginal means and 95% confidence intervals are displayed from ANOVA models. The dashed line indicates the grand mean for metabolite concentration and treatments with non-overlapping confidence intervals represent significant deviations from average. * $p < 0.05$.

Table S3.1.Sex differences in microbial-responsive metabolites. (A) ANOVA results of metabolite concentration by treatment.

(I) Female flies			
Metabolite	**F statistics**	**p value**	**q value**
CDP	13.8	6.20E-09	9.24E-07
Acetyl-aspartate	9.9	4.44E-07	2.53E-05
Glutathione disulfide	9.7	5.09E-07	2.53E-05
Lysine	9.2	9.82E-07	3.66E-05
Phenyllactic acid	7.8	6.36E-06	1.90E-04
dCTP	6.5	3.81E-05	9.46E-04
Xanthosine	5.6	1.48E-04	3.16E-03
Gluconate	5.5	1.76E-04	3.28E-03
N-Acetyl-glutamate	5.1	3.19E-04	5.04E-03
Glutathione	5.1	3.38E-04	5.04E-03
Aspartate	4.9	4.59E-04	6.22E-03
dCDP	4.5	8.64E-04	1.07E-02
Alanine/Sacrosine	4.2	1.54E-03	1.64E-02
Trehalose/Sucrose	4.2	1.54E-03	1.64E-02
O-Acetyl-serine	3.5	5.26E-03	0.1
CTP	3.4	6.41E-03	0.1
3-Phosphoglycerate	3.0	1.24E-02	0.1
Cysteine	2.8	1.72E-02	0.1
Ornithine	2.9	1.52E-02	0.1
2-Isopropylmalic acid	2.9	1.56E-02	0.1
CDP-ethanolamine	2.8	1.71E-02	0.1
Taurine	2.8	1.85E-02	0.1
Thiamine	2.7	2.00E-02	0.1
Hydroxyproline/Aminolevulinate	2.7	2.24E-02	0.1
Aconitate	2.7	2.31E-02	0.1
Cystathionine	2.7	2.16E-02	0.1
dTDP	2.6	2.55E-02	0.1
GMP	2.6	2.73E-02	0.1
2,3-Dihydroxybenzoic acid	2.5	3.10E-02	0.2
Proline	2.3	4.23E-02	0.2
Citraconic acid	2.3	4.16E-02	0.2
Ribose	2.4	3.82E-02	0.2
Tyrosine	2.3	4.20E-02	0.2

Pantothenate	2.4	4.12E-02	0.2
CDP-choline	2.3	4.23E-02	0.2
p-Hydroxybenzoate	2.3	4.37E-02	0.2
Fumarate	2.3	4.59E-02	0.2
Citrulline	2.2	0.1	0.2
Glutamate	2.1	0.1	0.2
Allantoate	2.1	0.1	0.2
Anthranilate	2.1	0.1	0.2
Lactate	2.1	0.1	0.3
Fructose-6-phosphate	2.0	0.1	0.3
Leucine/Isoleucine	2.0	0.1	0.3
Serine	2.0	0.1	0.3
Malate	2.0	0.1	0.3
UTP	2.0	0.1	0.3
Ser-Leu/Ile-Asp (SID)	1.9	0.1	0.3
Acetyl-CoA	1.9	0.1	0.3
CMP	1.9	0.1	0.3
Glycine	1.8	0.1	0.3
Glucose	1.8	0.1	0.3
GDP	1.8	0.1	0.3
Quinolinate	1.7	0.1	0.4
UDP-D-glucose	1.7	0.1	0.4
Hydroxyphenylpyruvate	1.7	0.1	0.4
Hydroxyisocaproic acid	1.7	0.1	0.4
Hypoxanthine	1.7	0.1	0.4
Threonine/Homoserine	1.6	0.2	0.4
Dihydroxy-acetone-phosphate	1.6	0.2	0.4
D-Sedoheptulose-1/7-phosphate	1.6	0.2	0.4
Ketoleucine	1.6	0.2	0.4
sn-Glycerol-3-phosphate	1.6	0.2	0.4
Acetyl-glycine	1.5	0.2	0.4
Xanthine	1.5	0.2	0.4
N-Acetyl-L-ornithine	1.5	0.2	0.4
Ser-Asp (SD)	1.6	0.2	0.4
UDP	1.5	0.2	0.4
2-Keto-isovalerate	1.5	0.2	0.4
Aminoadipic acid	1.5	0.2	0.4
Valine/5-Aminopentanoic acid	1.5	0.2	0.4
Histidinol	1.5	0.2	0.4

Glycerate	1.5	0.2	0.4
S-Adenosyl-L-homocysteine	1.5	0.2	0.4
4-Aminobutyrate	1.4	0.2	0.4
Glucarate	1.4	0.2	0.4
Glutamine	1.4	0.2	0.5
UMP	1.4	0.2	0.5
Oxaloacetate	1.4	0.2	0.5
GTP	1.4	0.2	0.5
Phenylpyruvate	1.3	0.3	0.5
Succinate/Methylmalonic acid	1.3	0.3	0.5
Pipecolic Acid	1.2	0.3	0.5
Methionine	1.2	0.3	0.5
Deoxyribose-phosphate	1.2	0.3	0.5
Leu-Asp (LD)	1.2	0.3	0.5
5-Phosphoribosyl-1-pyrophosphate	1.3	0.3	0.5
Thiamine pyrophosphate	1.2	0.3	0.5
FMN	1.2	0.3	0.5
FAD	1.2	0.3	0.5
Nicotinamide mononucleotide	1.1	0.4	0.6
Adenine	1.1	0.4	0.6
Pyrophosphate	1.1	0.4	0.6
Erythrose-4-phosphate	1.1	0.4	0.6
Delta-gluconolactone	1.0	0.4	0.6
Xanthosine-5-phosphate	1.1	0.4	0.6
3-S-Methylthiopropionate	1.0	0.4	0.7
IDP	1.0	0.4	0.7
Ribose-5-phosphate	0.9	0.5	0.7
Acetoacetate	0.9	0.5	0.7
Cytosine	0.8	0.6	0.7
Uracil	0.9	0.5	0.7
Maleic acid	0.8	0.6	0.7
Nicotinate	0.9	0.5	0.7
Pyroglutamic acid	0.8	0.6	0.7
Orotate	0.9	0.5	0.7
Uric acid	0.9	0.5	0.7
N-Carbamoyl-L-aspartate	0.9	0.5	0.7
2-Keto-D-gluconate	0.9	0.5	0.7
Biotin	0.8	0.6	0.7
S-Methyl-5'-thioadenosine	0.9	0.5	0.7

IMP	0.9	0.5	0.7
UDP-N-acetyl-glucosamine	0.8	0.6	0.7
Phosphoenolpyruvate	0.8	0.6	0.8
NADP+	0.8	0.6	0.8
Histidine	0.8	0.6	0.8
2-Aminooctanoic acid	0.7	0.6	0.8
ATP	0.7	0.6	0.8
NADPH	0.8	0.6	0.8
Kynurenic acid	0.7	0.6	0.8
Val-Asp (VD)	0.7	0.6	0.8
Octoluse bisphosphate	0.7	0.7	0.8
4-Phosphopantothenate	0.7	0.7	0.8
Cyclic-AMP	0.7	0.7	0.8
2-Oxo-4-methylthiobutanoate	0.6	0.7	0.8
Hydroxyphenylacetic acid	0.7	0.7	0.8
Indole-3-carboxylic acid	0.6	0.7	0.8
Glyceraldehdye-3-phosphate	0.7	0.7	0.8
3-Phospho-serine	0.6	0.7	0.8
Xanthurenic acid	0.6	0.8	0.8
Pro-Glu (PE)	0.7	0.7	0.8
6-Phospho-D-gluconate	0.7	0.7	0.8
dTMP	0.6	0.8	0.8
dAMP	0.6	0.7	0.8
AMP	0.6	0.7	0.8
dGDP	0.6	0.7	0.8
UDP-D-glucuronate	0.7	0.7	0.8
dGMP	0.6	0.8	0.8
Riboflavin	0.6	0.8	0.8
Citrate/isocitrate	0.6	0.8	0.8
Alpha-ketoglutarate	0.5	0.8	0.9
N-Acetyl-L-alanine	0.5	0.8	0.9
NADH	0.5	0.8	0.9
Fructose-1,6-bisphosphate	0.5	0.8	0.9
dCMP	0.5	0.9	0.9
Ribulose-5-phosphate	0.4	0.9	0.9
Pro-Asp (PD)	0.4	0.9	0.9
Sedoheptoluse bisphosphate	0.3	0.9	0.9
Asparagine	0.2	1.0	1.0
(ii) Male flies			

Metabolite	F statistics	p value	q value
Phenyllactic acid	**40.5**	**3.37E-16**	**5.02E-14**
Fumarate	**13.8**	**6.24E-09**	**3.78E-07**
Glutathione disulfide	**13.6**	**7.60E-09**	**3.78E-07**
NADPH	**12.8**	**1.66E-08**	**6.17E-07**
Acetyl-CoA	**10.4**	**2.35E-07**	**6.99E-06**
Aconitate	**9.7**	**5.34E-07**	**1.33E-05**
Aspartate	**5.9**	**9.05E-05**	**1.93E-03**
Citrate/isocitrate	**5.5**	**1.82E-04**	**3.39E-03**
UDP-D-glucose	**5.2**	**2.74E-04**	**4.54E-03**
Thiamine	**4.9**	**4.79E-04**	**7.14E-03**
3-Phosphoglycerate	**4.5**	**8.61E-04**	**1.17E-02**
Xanthosine	**4.4**	**1.04E-03**	**1.30E-02**
UMP	**4.0**	**1.99E-03**	**2.28E-02**
Acetyl-aspartate	**3.9**	**2.51E-03**	**2.67E-02**
Pyrophosphate	**3.7**	**3.50E-03**	**3.26E-02**
Nicotinamide mononucleotide	**3.7**	**3.41E-03**	**3.26E-02**
5-Phosphoribosyl-1-pyrophosphate	**3.7**	**3.87E-03**	**3.39E-02**
UDP	**3.5**	**5.33E-03**	**4.42E-02**
Hydroxyproline/Aminolevulinate	**3.4**	**6.49E-03**	**0.1**
Proline	**3.3**	**7.32E-03**	**0.1**
Lactate	**3.2**	**8.93E-03**	**0.1**
N-Acetyl-L-alanine	**3.0**	**1.15E-02**	**0.1**
ATP	**3.0**	**1.20E-02**	**0.1**
4-Phosphopantothenate	**2.9**	**1.47E-02**	**0.1**
Gluconate	**2.5**	**3.04E-02**	**0.2**
Citraconic acid	**2.5**	**3.28E-02**	**0.2**
Hydroxyisocaproic acid	2.4	3.68E-02	0.2
IDP	2.4	3.81E-02	0.2
Aminoadipic acid	2.4	4.03E-02	0.2
3-S-Methylthiopropionate	2.3	4.17E-02	0.2
O-Acetyl-serine	2.3	4.52E-02	0.2
Tyrosine	2.2	0.1	0.3
Glutamate	2.2	0.1	0.3
Threonine/Homoserine	2.1	0.1	0.3
Ribulose-5-phosphate	2.1	0.1	0.3
D-Sedoheptulose-1/7-phosphate	2.1	0.1	0.3
Glycerate	2.0	0.1	0.3

Malate	2.0	0.1	0.3
Allantoate	2.1	0.1	0.3
Acetoacetate	2.0	0.1	0.3
6-Phospho-D-gluconate	2.0	0.1	0.3
Pyroglutamic acid	1.9	0.1	0.3
AMP	1.9	0.1	0.3
UDP-N-acetyl-glucosamine	1.9	0.1	0.4
p-Hydroxybenzoate	1.8	0.1	0.4
N-Carbamoyl-L-aspartate	1.8	0.1	0.4
FMN	1.8	0.1	0.4
2,3-Dihydroxybenzoic acid	1.8	0.1	0.4
Ribose-5-phosphate	1.8	0.1	0.4
CMP	1.8	0.1	0.4
Phosphoenolpyruvate	1.7	0.1	0.4
Xanthine	1.7	0.1	0.4
Erythrose-4-phosphate	1.7	0.1	0.4
Riboflavin	1.7	0.1	0.4
Alanine/Sacrosine	1.6	0.2	0.4
Ornithine	1.6	0.2	0.4
Xanthosine-5-phosphate	1.6	0.2	0.4
Fructose-6-phosphate	1.6	0.2	0.4
CDP-choline	1.6	0.2	0.4
Maleic acid	1.5	0.2	0.4
Ketoleucine	1.5	0.2	0.4
2-Oxo-4-methylthiobutanoate	1.5	0.2	0.4
Histidine	1.5	0.2	0.4
Glucarate	1.5	0.2	0.4
NADP+	1.5	0.2	0.4
Pro-Glu (PE)	1.5	0.2	0.5
dGDP	1.5	0.2	0.5
Fructose-1,6-bisphosphate	1.4	0.2	0.5
Acetyl-glycine	1.4	0.2	0.5
Succinate/Methylmalonic acid	1.4	0.3	0.5
4-Aminobutyrate	1.3	0.3	0.6
UTP	1.3	0.3	0.6
dGMP	1.3	0.3	0.6
3-Phospho-serine	1.2	0.3	0.6
Lysine	1.2	0.3	0.7
sn-Glycerol-3-phosphate	1.2	0.3	0.7

Delta-gluconolactone	1.2	0.3	0.7
Kynurenic acid	1.2	0.3	0.7
Deoxyribose-phosphate	1.2	0.3	0.7
Glutamine	1.1	0.4	0.7
Uric acid	1.1	0.4	0.7
Citrulline	1.1	0.4	0.7
Glycine	1.1	0.4	0.7
Indole-3-carboxylic acid	1.1	0.4	0.7
Dihydroxy-acetone-phosphate	1.1	0.4	0.7
Ser-Asp (SD)	1.1	0.4	0.7
Trehalose/Sucrose	1.1	0.4	0.7
IMP	1.1	0.4	0.7
Taurine	1.1	0.4	0.7
Adenine	1.1	0.4	0.7
GMP	1.0	0.4	0.7
GDP	1.0	0.4	0.7
CDP-ethanolamine	1.0	0.4	0.7
2-Keto-isovalerate	1.0	0.4	0.7
Orotate	1.0	0.4	0.7
N-Acetyl-glutamate	1.0	0.4	0.7
CDP	1.0	0.4	0.7
Leucine/Isoleucine	0.9	0.5	0.7
Ribose	1.0	0.5	0.7
S-Methyl-5'-thioadenosine	1.0	0.5	0.7
Ser-Leu/Ile-Asp (SID)	1.0	0.5	0.7
Octoluse bisphosphate	0.9	0.5	0.7
Thiamine pyrophosphate	0.9	0.5	0.7
Histidinol	0.9	0.5	0.7
Glyceraldehdye-3-phosphate	0.9	0.5	0.7
Hydroxyphenylacetic acid	0.9	0.5	0.7
dTDP	0.9	0.5	0.7
Nicotinate	0.9	0.5	0.7
Hypoxanthine	0.9	0.5	0.7
2-Isopropylmalic acid	0.9	0.5	0.7
Serine	0.8	0.6	0.8
Cysteine	0.8	0.6	0.8
Oxaloacetate	0.8	0.6	0.8
Asparagine	0.8	0.6	0.8
2-Aminooctanoic acid	0.8	0.6	0.8

Quinolinate	0.8	0.6	0.8
2-Keto-D-gluconate	0.8	0.6	0.8
Pro-Asp (PD)	0.8	0.6	0.8
Leu-Asp (LD)	0.8	0.6	0.8
dCMP	0.8	0.6	0.8
S-Adenosyl-L-homocysteine	0.8	0.6	0.8
FAD	0.8	0.6	0.8
N-Acetyl-L-ornithine	0.7	0.7	0.8
Biotin	0.7	0.7	0.8
Glutathione	0.7	0.7	0.8
dTMP	0.7	0.6	0.8
Sedoheptoluse bisphosphate	0.7	0.7	0.8
Methionine	0.6	0.7	0.8
Cystathionine	0.6	0.7	0.8
dAMP	0.7	0.7	0.8
Uracil	0.6	0.7	0.8
Pipecolic Acid	0.6	0.7	0.8
Anthranilate	0.6	0.7	0.8
Glucose	0.6	0.7	0.8
Cytosine	0.6	0.8	0.8
Alpha-ketoglutarate	0.6	0.8	0.8
Val-Asp (VD)	0.6	0.8	0.8
CTP	0.6	0.8	0.8
GTP	0.5	0.8	0.9
NADH	0.5	0.8	0.9
UDP-D-glucuronate	0.5	0.8	0.9
Valine/5-Aminopentanoic acid	0.5	0.8	0.9
Xanthurenic acid	0.5	0.9	0.9
Pantothenate	0.4	0.9	0.9
Cyclic-AMP	0.4	0.9	0.9
dCDP	0.3	1.0	1.0
Phenylpyruvate	0.2	1.0	1.0
Hydroxyphenylpyruvate	0.1	1.0	1.0
dCTP	0.0	1.0	1.0

Treatment and residual degrees of freedom were 7 and 40, respectively. Metabolites with q values < 0.2 are shown in bold.

Table S3.1.Sex differences in microbial-responsive metabolites. (B) Metabolite classification and distribution by sex.

ClassyFire subclass	Metabolite	Female	Male
Alcohols and polyols	Pantothenate		
Alpha hydroxy acids and derivatives	Lactate		
Amino acids, peptides, and analogues	4-Phosphopantothenate		
Amino acids, peptides, and analogues	Acetyl-aspartate		
Amino acids, peptides, and analogues	Alanine/Sacrosine		
Amino acids, peptides, and analogues	Aspartate		
Amino acids, peptides, and analogues	Cystathionine		
Amino acids, peptides, and analogues	Cysteine		
Amino acids, peptides, and analogues	Glutathione		
Amino acids, peptides, and analogues	Glutathione disulfide		
Amino acids, peptides, and analogues	Hydroxyproline/Aminolevulinate		
Amino acids, peptides, and analogues	Lysine		
Amino acids, peptides, and analogues	N-Acetyl-glutamate		

Amino acids, peptides, and analogues	N-Acetyl-L-alanine		
Amino acids, peptides, and analogues	O-Acetyl-serine		
Amino acids, peptides, and analogues	Ornithine		
Amino acids, peptides, and analogues	Proline		
Amino acids, peptides, and analogues	Tyrosine		
Benzene and substituted derivatives	Phenyllactic acid		
Benzoic acids and derivatives	2,3-Dihydroxybenzoic acid		
Benzoic acids and derivatives	p-Hydroxybenzoate		
Carbohydrates and carbohydrate conjugates	3-Phosphoglycerate		
Carbohydrates and carbohydrate conjugates	5-Phosphoribosyl-1-pyrophosphate		
Carbohydrates and carbohydrate conjugates	CTP		
Carbohydrates and carbohydrate conjugates	Gluconate		
Carbohydrates and carbohydrate conjugates	Ribose		
Carbohydrates and carbohydrate conjugates	Trehalose/Sucrose		
Dicarboxylic acids and derivatives	Fumarate		

Fatty acids and conjugates	2-Isopropylmalic acid	[redacted]	
Fatty acids and conjugates	Citraconic acid		[redacted]
Fatty acyl thioesters	Acetyl-CoA		
Nicotinamide nucleotides	Nicotinamide mononucleotide		
Non-metal phosphates	Pyrophosphate		
Nucleosides, nucleotides, and analogues	NADPH		
Organic pyrophosphates	dCDP		
Organosulfonic acids and derivatives	Taurine		
Purine nucleosides	Xanthosine		
Purine ribonucleotides	ATP		
Purine ribonucleotides	GMP		
Pyrimidine deoxyribonucleotides	dCTP		
Pyrimidine deoxyribonucleotides	dTDP		
Pyrimidine nucleotide sugars	UDP-D-glucose		
Pyrimidine ribonucleotides	CDP		
Pyrimidine ribonucleotides	CDP-choline		
Pyrimidine ribonucleotides	CDP-ethanolamine		
Pyrimidine ribonucleotides	UDP		
Pyrimidine ribonucleotides	UMP		

Pyrimidines and pyrimidine derivatives	Thiamine	
Tricarboxylic acids and derivatives	Aconitate	
Tricarboxylic acids and derivatives	Citrate/isocitrate	

Black and white cells correspond to presence and absence, respectively.

Table S3.2. Multivariate association between metabolome and microbiome.

(a) Multivariate correlation with microbial abundance (CFU fly^{-1})				
Factor	Female		Male	
	Effect size (R^2)	p value	Effect size (R^2)	p value
Log_{10}(ACE + 1)	**0.37**	**0.001**	**0.33**	**0.001**
Log_{10}(LAC + 1)	0.01	0.841	**0.13**	**0.040**
Log_{10}(YST + 1)	0.04	0.392	0.13	0.056

(b) Full factorial PERMANOVA results (df = 1,40)				
Factor	Female		Male	
	F statistic (p)	Effect size (R^2)	F statistic (p)	Effect size (R^2)
ACE	**6.60 (0.001)**	**0.10**	**8.57 (0.001)**	**0.11**
LAC	**4.19 (0.001)**	**0.06**	**11.17 (0.001)**	**0.14**
YST	**4.85 (0.001)**	**0.07**	**6.28 (0.001)**	**0.08**
ACE*LAC	**2.87 (0.008)**	**0.04**	**3.93 (0.002)**	**0.05**
ACE*YST	**3.86 (0.001)**	**0.06**	**3.29 (0.004)**	**0.04**
LAC*YST	1.43 (0.149)	0.02	**2.50 (0.016)**	**0.03**
ACE*LAC*YST	**2.63 (0.006)**	**0.04**	1.30 (0.214)	0.02

Values in bold correspond to significant results.

Table S3.3. List of metabolites that differentiate by treatment.

Metabolite	Female	Male
Acetyl-aspartate	1.1	1.3
Acetyl-CoA	-	1.5
Aconitate	-	1.6
Aspartate	1.2	-
CDP	1.8	-
CDP-choline	1.1	-
CDP-ethanolamine	1.1	-
Citrate/isocitrate	-	1.8
CTP	1.1	-
Cystathionine	1.1	-
Cysteine	1.1	-
dCDP	1.1	-
dCTP	1.4	-
Fumarate	1.2	1.0
Glutathione disulfide	1.5	1.2
Lysine	1.2	-
p-Hydroxybenzoate	1.1	-
Phenyllactic acid	1.1	2.2
Proline	1.0	-
Trehalose/Sucrose	1.3	-
Xanthosine	-	1.4

VIP scores > 1 displayed for each sex; - indicates values < 1.

Table S3.4. Correlation between nutritional indices and top microbial-responsive metabolites

(i) Female flies					
Nutritional index	**Metabolite**	**Rho**	**S statistic**	**p value**	**q value**
Glucose	**Phenyllactic acid**	**-0.653284**	**26808**	**1.72E-06**	**8.78E-05**
Glucose	**Proline**	**-0.484428**	**24070**	**0.00074732**	**0.01905658**
Glucose	p-Hydroxybenzoate	-0.3150786	21324	0.03339784	0.42200769
Glucose	CTP	-0.2378662	20072	0.11135834	0.56792753
Glucose	CDP	-0.2138826	19683.1069	0.15349997	0.59713503
Glucose	Trehalose/Sucrose	-0.2085106	19596	0.16391942	0.59713503
Glucose	Cystathionine	-0.1834721	19190	0.2215247	0.70610999
Glucose	CDP-choline	-0.1675661	18932.0838	0.26566368	0.7466606
Glucose	Fumarate	-0.1467567	18594.6605	0.33043263	0.80247924
Glucose	Cysteine	-0.1086065	17976.0543	0.47247203	0.85657504
Glucose	CDP-ethanolamine	-0.1173605	18118	0.43605185	0.85657504
Glucose	dCTP	-0.0827756	17557.207	0.58444651	0.89984998
Glucose	Aspartate	-0.0781375	17482	0.60467639	0.89984998
Glucose	Lysine	0.05346901	15348	0.72340881	0.89984998
Glucose	dCDP	-0.0670367	17302	0.65707948	0.89984998
Glucose	Glutathione disulfide	0.03225409	15692	0.83112988	0.96104433
Glucose	Acetyl-aspartate	-0.0151047	16459.9231	0.92063689	0.96154404
Protein	CDP	0.34475315	10624.8276	0.01895994	0.32231905
Protein	CDP-choline	0.28375836	11613.8581	0.05599613	0.42200769
Protein	Acetyl-aspartate	0.22421851	12579.2968	0.13412816	0.57004467
Protein	p-Hydroxybenzoate	0.19346284	13078	0.19704707	0.66996003
Protein	CDP-ethanolamine	0.17804502	13328	0.23567602	0.70702805
Protein	CTP	0.14079556	13932	0.3495008	0.81020639
Protein	dCDP	0.13092815	14092	0.38456222	0.85272491
Protein	dCTP	0.09017733	14752.7745	0.55117732	0.89984998
Protein	Phenyllactic acid	0.08011101	14916	0.59555482	0.89984998
Protein	Cysteine	-0.0554442	17114.0277	0.71438467	0.89984998
Protein	Proline	0.06333642	15188	0.67493576	0.89984998
Protein	Glutathione disulfide	0.02176997	15862	0.88566831	0.96104433
Protein	Trehalose/Sucrose	-0.0237434	16600	0.87535388	0.96104433
Protein	Fumarate	-0.0234416	16595.1055	0.87710959	0.96104433
Protein	Cystathionine	-0.0326241	16744	0.82921784	0.96104433
Protein	Lysine	0.0120259	16020	0.93684041	0.96154404
Protein	Aspartate	-0.0009251	16230	0.99543747	0.99543747

TAG	Trehalose/Sucrose	0.27375887	11776	0.06589746	0.42200769
TAG	Phenyllactic acid	0.2657416	11906	0.07447195	0.42200769
TAG	p-Hydroxybenzoate	0.29805735	11382	0.04464034	0.42200769
TAG	Proline	0.27178538	11808	0.06793082	0.42200769
TAG	Cystathionine	0.22565526	12556	0.13137858	0.57004467
TAG	Cysteine	0.16331062	13566.9183	0.27816767	0.7466606
TAG	Fumarate	-0.1542827	18716.6944	0.30595587	0.78018746
TAG	CDP	-0.1129853	18047.0565	0.45469347	0.85657504
TAG	Acetyl-aspartate	-0.1190426	18145.2753	0.43071178	0.85657504
TAG	dCDP	0.10477953	14516	0.48707208	0.85657504
TAG	CDP-choline	-0.1071263	17952.0536	0.47856415	0.85657504
TAG	CTP	0.0986124	14616	0.51318246	0.87241019
TAG	Glutathione disulfide	-0.0686401	17328	0.6493999	0.89984998
TAG	dCTP	-0.0547726	17103.1369	0.71770057	0.89984998
TAG	Lysine	0.07332717	15026	0.62716155	0.89984998
TAG	CDP-ethanolamine	-0.0241135	16606	0.87342216	0.96104433
TAG	Aspartate	0.01091582	16038	0.94269024	0.96154404
(ii) Male flies					
Nutritional index	Metabolite	Rho	S statistic	p value	q value
Glucose	**Phenyllactic acid**	**-0.63**	**26431.58**	**2.71E-06**	**3.25E-05**
Glucose	**Aconitate**	**-0.38**	**22345.38**	**0.01**	**0.06**
Glucose	**Fumarate**	**-0.34**	**21804.17**	**0.02**	**0.09**
Glucose	Xanthosine	-0.13	18396.03	0.37	0.60
Glucose	Citrate/isocitrate	-0.11	17982.22	0.47	0.70
Glucose	Acetyl-aspartate	0.08	14873.14	0.58	0.70
Glucose	Glutathione disulfide	-0.09	17630.00	0.56	0.70
Glucose	Acetyl-CoA	0.07	15096.00	0.65	0.71
Protein	**Xanthosine**	**-0.46**	**23680.91**	**1.28E-03**	**0.01**
Protein	Fumarate	0.27	11888.87	0.07	0.25
Protein	Acetyl-CoA	0.24	12276.00	0.10	0.31
Protein	Acetyl-aspartate	0.23	12565.93	0.13	0.32
Protein	Glutathione disulfide	0.14	13876.00	0.34	0.58
Protein	Citrate/isocitrate	-0.10	17848.20	0.51	0.70
Protein	Aconitate	0.08	14949.92	0.61	0.70
Protein	Phenyllactic acid	-0.02	16499.04	0.91	0.91
TAG	**Phenyllactic acid**	**0.70**	**4823.24**	**5.26E-08**	**1.26E-06**
TAG	**Acetyl-CoA**	**-0.33**	**21518.00**	**0.03**	**0.11**

TAG	Aconitate	0.23	12482.77	0.12	0.32
TAG	Glutathione disulfide	-0.21	19598.00	0.16	0.36
TAG	Xanthosine	0.20	12953.50	0.18	0.36
TAG	Acetyl-aspartate	-0.18	19165.10	0.23	0.42
TAG	Fumarate	0.08	14956.96	0.61	0.70
TAG	Citrate/isocitrate	0.06	15255.88	0.70	0.73

Correlations with q values < 0.2 are indicated in bold.