Title:

Horizontal gene transfer-mediated bacterial strain variation affects host fitness

Running title:

bacterial strain variation and fly fitness

Authors:

Yun Wang, Franz Baumdicker, Sven Kuenzel, Fabian Staubach

Author affiliations:

Yun Wang, Albert Ludwigs University Freiburg, Biology I, Evolutionary Biology and Ecology, Freiburg, Germany

Franz Baumdicker, Cluster of Excellence 'Controlling Microbes to Fight Infections', University of Tübingen, University of Freiburg, Department of Mathematical Stochastics

Sven Kuenzel, Max Planck Institute for Evolutionary Biology, Ploen, Germany

Fabian Staubach, Albert Ludwigs University Freiburg, Biology I, Evolutionary Biology and Ecology, Freiburg, Germany

Corresponding author:

Fabian Staubach, fabian.staubach@biologie.uni-freiburg.de

Abstract

How microbes affect host fitness and environmental adaptation has become a fundamental research question in evolutionary biology. We tested for associations of bacterial genomic variation and *Drosophila melanogaster* offspring number in a microbial Genome Wide Association Study (GWAS). Leveraging strain variation in the genus *Gluconobacter*, a genus of bacteria that are commonly associated with *Drosophila* under natural conditions, we pinpoint the thiamine biosynthesis pathway (TBP) as contributing to differences in fitness conferred to the fly host. By tracing the evolutionary history of TBP genes in *Gluconobacter*, we find that TBP genes were most likely lost and reacquired by horizontal gene transfer (HGT). We suggest that HGT might contribute to microbiome flexibility and speculate that it can also more generally contribute to host adaptation.

Introduction

Microbes are important drivers of host phenotype and evolution [1]. Benefits derived from microorganisms can facilitate the occupation of new ecological niches [2–5] and microbial effects on host phenotypes and fitness can spur adaptive processes [6–14]. Changes in the effects of microbes on host fitness can alter interactions along the parasitism mutualism continuum [6, 15–18], thus affecting the evolutionay trajectories of the partners. The importance of microbes in evolution and health of higher organisms has sparked a search for the molecular underpinnings of how microbes affect host phenotype.

In this search, microbial Genome Wide Association Studies (GWAS) are an important tool [19–23]. The principle behind a microbial GWAS is to establish a link between traits and genetic variation of microbes by the means of GWAS. By testing for association between host traits and microbial genomic variation, Chaston et al. [24] introduced a particularly helpful approach to unravel how microbes affect host phenotypes [22, 24]. The authors measured host

phenotypes, here from *Drosophila melanogaster* that were mono-associated with several microbial isolates. Differences in host phenotype were then associated with the presence and absence of genes in the microbial isolates. By applying this approach, it was found that genes that play a role in glucose oxidation in bacteria affect *D. melanogaster* triglyceride levels [24] and that bacterial methionine and B vitamins are important for starvation resistance [25] as well as life span [26].

For targetting host phenotypes with microbial GWAS, model systems that allow the generation of axenic hosts that can successively be associated with individual microbial isolates are particularly useful [24]. One such model system is *D. melanogaster* and its bacterial microbiome. Techniques for the generation of gnotobiotic flies are readily available and standardized measurements of phenotypes exist. Affected phenotypes include the life history traits development time, fecundity, and life span as well as size of the adults [14, 26–33]. These traits are directly related to fitness, emphasizing the potential importance of microbes in host evolution and adaptation. Microbes often affect fitness related traits by provisioning nutrients. These nutrients include vitamins, amino acids, lipids, and trace elements [24, 34–39]. Nutrient provisioning is a recurring theme in macrobe–microbe interactions that are adaptive for the host [40–42].

The acquisition of nutrients from microbes need not rely on microbes that live inside the host. Instead, nutrients can also be acquired by harvesting or preying upon microbes that live outside the fly and subsequent digestion [35, 43–45]. Furthermore, bacteria have been identified that affect *D. melanogaster* phenotypes by increasing the ability for nutrient uptake [46] or metabolizing components of the food substrate, and thus modulating its nutrient content [24]. Interestingly, the metabolic potential to produce nutrients that affect fly fitness differ between closely related microbes and so do the effects on fly phenotype and fitness [26, 29, 32, 47–51]. These findings contribute to the notion that microbial variation at taxonomically low levels is not

only important for human [15], mouse [52], and plant [18] hosts, but also for *Drosophila* [53].

Because variation between closely related bacteria is important for the interaction of the host and its microbiome, it is also important to consider closely related microbes in studies that aim at elucidating the molecular underpinnings of host-microbe interaction. At the same time, using the pan-genome of closely related microbes in GWAS might offer particular power to the approach: on a similar genomic background, microbial genes that affect the host are easier to identify. For studies that are aimed at better understanding host-microbe interaction in an evolutionary context, it is also important to consider microbes that are associated with the host under natural conditions and if possible, to measure evolutionary relevant host phenotypes in a natural or near natural environment. Finally, tracing the evolutionary changes of the genomic elements that affect host fitness can help us to gain deeper insights into how host-microbe interaction evolves.

We aimed our study at better understanding whether and how fly fitness is affected by its natural microbiome by a microbial GWAS. In order to increase the power of the approach and consider variation at low taxonomic levels, we concentrated on variation within a taxonomically restricted group of bacteria. Therefore, we focused our study on *Gluconobacter,* a bacterial genus that is commonly associated with *D. melanogaster* under natural conditions [54–57]. We assessed fitness parameters on grape juice based fly food as a near natural food source. In order to better understand how microbial effects on host fitness evolve, we traced the evolutionary events that lead to changes in bacteria-mediated host fitness.

Materials and Methods

Fitness assays

Canton-S stocks were kept at 25 °C on food prepared following the Bloomington *Drosophila* Stock Center 'Cornmeal Molasses and Yeast Medium' (532 ml water, 40 ml molasses, 6.6 g

yeast. 32.6 g cornmeal, 3.2 g agar, 2.2 ml propionic acid and 7.6 ml Tegosept). To generate axenic flies, embryos were collected and washed in PBS, dechorionated in 50% bleach for 2-3 mins, and rinsed in sterile PBS for 1 min. Embryos were placed in sterilized food bottles under a sterile workbench and maintained at 25 °C under a 12:12 light cycle in axenic condition for 3 weeks during which the flies had time to hatch and mate. One axenic female from these bottles was used per vial in the fitness assay. For the fitness assay, bacterial cultures were grown in liquid YPD medium for 48-72 hours and normalized to the same optical density (OD_{600} = 0.6). 150 µl of OD normalized medium were added directly on 10ml sterile grape juice food (667 ml water, 333 ml Jacoby white grape juice, 8 g yeast, 50 g cornmeal, 10 g agar, 3 ml propionic acid). Axenic females were transferred to the vial immediately afterwards. We prepared two control treatments. First, we added sterile YPD medium to the food as axenic control. Second, we used conventionally reared flies homogenized in YPD as inoculum. We tracked the number of pupae per vial for 16 consecutive days after infection. On the last day, flies were counted, collected and weighed. All offspring were weighed together in one Eppendorf tube for each replicate and weight per fly was calculated. All fitness related measurements were done blind. That means the vials were given random numbers and only after the measurements were taken, the bacterial strain ID was connected to the result. For the thiamine supplement experiment, we added 1 µg/ml thiamine to the food described above. That concentration has proven effective for phenotypic rescue in [37]. All statistical analyses were performed in R and can be found in supplementary script S1.

Bacterial loads and contamination control

Fly offspring from the fitness assays were stored in PBS/glycerol mixture at -80 °C for later contamination control and the counting of colony forming units (CFUs). 3-6 replicates per bacterial isolate were picked for CFU counting. For CFU counting, samples of 3 to 5 offspring

were homogenized with a pestle in 300 µl of PBS. The homogenates were plated on YPD agar medium. Plates were incubated for 48 hours. CFU counts were done visually or with the OpenCFU software [58], Table S4). Plates for CFU counting were also inspected for colony morphology and colony color that could indicate potential contamination, with negative results. All homogenates were plated on antibiotic YPD agar medium (with 100 µg/ml kanamycin or ampicillin) for assessing yeast contamination. No yeast colonies were observed except in the control replicates in which conventional lab microbiota were used. To further assess potential bacterial contaminants during our experiment, we quantified the relative abundance of target isolates on fly offspring using 16S rRNA gene sequencing. In short, DNA was extracted from pools of 3-5 offspring for 3-6 replicates per bacterial isolate after the experiment, including the replicates with the highest and lowest offspring number. The V4 regions of the bacterial 16S rRNA gene were amplified and sequenced on an illumina MiSeq sequencer following [56, 59]. Sequencing data were analyzed using mothur [60] (See supplementary script S2 for all commands executed). The relative abundance of target 16S rRNA gene sequences for mono-associated isolates was calculated. The average relative abundance of target 16S sequences was over 88% (Figure S8A) in the initial experiment. Only in (6 out of 66) replicates the relative abundance was below 75%, including 3 cases of *P. sneebia* that showed very low bacterial loads. For the thiamine treatment the target bacteria were significantly enriched in the microbial community (Figure S8B).

Bacterial isolates, genome sequencing and assembly:

We sequenced, assembled, and annotated draft genomes of eleven bacteria and added genome data for six bacteria from public databases (Table S1). Nine strains were isolated from wild-caught *Drosophila* collected in the San Francisco Bay Area (California, USA). Isolates were cultured in YPD for standard phenol-chloroform DNA extraction. Bacterial genomes were

151 sequenced using Illumina MiSeq technology and assembled with the A5 MiSeq assembler [61].
152 Completeness and contamination were assessed with checkM v1.1.2 [62], using standard
153 settings. Assembly statistics were generated with QUAST v5.0.2 [63]. Annotation was performed
154 with prokka v1.1 [64] or imported from GenBank. Average nucleotide identity (ANI) was
155 computed with fastANI (v0.1.2). New isolates were taxonomically classified, using GTDBtk
156 (v0.1.4) [65]. FastANI and GTDBtk were run on the kbase web interface [66].
157
158 Pan-genome clustering and phylogenetic trees
159 Genomes were analyzed using the panX analysis pipeline [60] with standard parameters (script
160 S3). Genes were grouped into 11269 clusters of homologous sequences, including clusters with
161 a single gene. Thereby the presence and absence of each gene cluster in the 17 genomes was
162 estimated. Based on the alignments of all single-copy genes that are present in all 17 genomes,
163 panX reconstructs a phylogenetic tree. For the phylogeny, FastTree 2 [67] and RaxML [68] are
164 applied to all variable positions from these alignments. For the phylogeny of thiE, nucleotide
165 sequences were aligned using MUSCLE v3.8.425 [99] and the tree was built using MrBayes
166 3.2.6 [70] as incorporated in the Geneious software suit v1.1 (Biomatters ltd.).
167
168 Microbial pan-genome-wide association study
169 We calculated the gene presence absence association score (PA score) between each
170 predicted cluster of homologous genes and fly offspring number. I.e. if D_g is the difference
171 between the mean fly offspring for strains with and without gene g, σ is the standard deviation of
172 fly offspring and n_g is the number gene gains and losses as inferred from the phylogeny. The
173 association score is given by $\sqrt{n_g}\frac{D_g}{\sigma}$. Three alternative association scores from treeWAS [21]

and the corresponding model based p-values were calculated. Association scores based on the presence and absence of genes are prone to false positives because genome wide linkage results in strongly correlated presence and absence of genes. Panx and treeWAS reduce this effect by taking the reconstructed ancestral gene gain and loss events into account.

Results

We performed a microbial GWAS for the number of offspring produced by females that were mono-associated with 17 bacterial isolates from genera that co-occur with *Drosophila melanogaster* in its natural environment. *Gluconobacter* was represented by 13 isolates. Two additional isolates were from the genus *Acetobacter*. Species from this genus can benefit *Drosophila* development [28]. One isolate was *Commensalibacter intestini* that might have a probiotic function in *D. melanogaster* [71] and is enriched in flies over substrate in wild-caught flies [57]. As an outgroup and to get a baseline for the fitness effect of an ingested pathogen, we added *Providencia sneebia,* that is highly pathogenic when entering the hemolyph [72]. All bacterial genomes analyzed were >99% complete with the exception of *P. sneebia* (>96%, Table S1). The mean number of offspring varied significantly between flies mono-associated with different isolates (P = 4.2 x 10^{-15} , Kruskal-Wallis-Test, Figure 1A) up to a 2.8-fold difference between *Gluconobacter morbifer* and *Gluconobacter sp.* P5H9d. Differences between bacterial strains also explained a significant proportion of variation in offspring number when we accounted for bacterial loads per fly (P = 1.4 x 10^{-4}, linear model), suggesting that not only bacterial biomass affects fly fitness. Presence-absence patterns of 11269 genes were tested for association with the number of offspring that mono-associated females produced using the PA method [73]. Associations were confirmed by permutation tests and TreeWAS [21] (Figure S1, Table 1, Table S2). The six highest PA scores depended strongly on presence-absence patterns between the closely related strains P1C6b, DSM2003, DSM2343, and DSM3504

(mean ANI = 95.5%) in the branch that includes *G. morbifer* (Figure S1 and accompanying text).

The bacterial thiamine biosynthesis pathway is associated with increased offspring number

Five of the six bacterial genes that were most strongly associated with offspring number were part of the thiamine biosynthesis pathway (TBP, Table 1). Females reared on bacteria carrying a complete TBP (TBP+) produced more offspring (P = 0.0038, Mann-Whitney-Test on strain medians, n = 17, Figure 1A), suggesting that bacterial thiamine production might increase the number of offspring.

Because high numbers of *Drosophila* offspring on a confined resource like a *Drosophila* vial can lead to crowding effects including smaller adults and reduced individual fitness, we weighed the adult flies at the end of the experiment. Weight did not differ significantly between the offspring of females reared on TBP+ and TBP- strains (P = 0.55, Mann-Whitney-Test on strain medians, n = 17, Figure S2), providing no evidence for larval crowding or reduced adult size.

The increase in the number of adult offspring for females reared on TBP+ strains could result from a shortened development time or an increased number of eggs laid by the females. Due to the lack of evidence that bacterial thiamine increases egg laying [37], we expected that time to puparium formation would be shorter for flies associated with TBP+ bacteria. Indeed, pupariation time was significantly reduced for offspring of females reared on TBP+ strains (P = 0.015, Mann-Whitney-Test on strain medians, n = 17, Figure S3). Significance of all p-values was confirmed in a linear model framework that accounts for bacterial load and also in a phylogenetic ANOVA (Script S1).

In order to further explore a potential role of thiamine in increasing fly offspring number, we supplemented the diet of females that were mono-associated with seven bacterial strains (*G. oxydans* DSM2343, *G. oxydans* DSM2003, *G. sp.* DSM3504, *G. sp.* P1C6_b, *G. cerevisiae*

DSM27644, *G. morbifer* G707 and *C. intestini* A911) with thiamine. This setup covered the *G. morbifer* branch of the phylogeny and all acetic acid bacteria that are missing key enzymes of the TBP (TBP-). The relative number of offspring for the three TBP- strains increased under thiamine treatment (P = 0.025, linear mixed effects model, Figure 1B), supporting a role of bacterial thiamine production in the number of offspring that flies produced. Furthermore, with thiamine added, neither average offspring weight nor the time to pupariation differed between TBP+ and TBP- associated flies (P = 1 both cases, Mann-Whitney-Test on strain medians, n = 7, Figure S4).

Thiamine biosynthesis genes were most likely lost and reacquired by horizontal gene transfer as an operon on the branch that includes *G. morbifer*

In order to better understand the evolutionary history of the TBP (Figure 2A) in *Gluconobacter*, we analyzed the synteny of the underlying loci in a phylogenetic framework. The strains in the upper two panels of Figure 2B possess all genes required for thiamine biosynthesis. A closer inspection of TBP genes on the *G. morbifer* branch (II in Figure 2C) revealed that two strains are TBP-, while the four other strains are TBP+. Inspection of the TBP gene loci revealed that all strains on branch II are missing the operon like structure thiOSG (Figure 2C) at the locus that is syntenic with branch I. The same pattern was found for thiC and thiD (Figure S5). thiOSG (Figure 2C), thiC, and thiD (Figure S6) are present in the closely related bacteria *Gluconobacter samuiensis* and *Neokomagateaa tanensis* at syntenic loci, suggesting deletion on branch II. The strains with an intact operon on branch II carried a TBP operon at different loci (Figure 2D), suggesting insertion.

Analyzing the sequences of the inserted genes in a phylogenetic framework, we found that the inserted genes form a distant clade. For example thiE1, the copy that remained at the locus shown in Figure 2C, followed the phylogeny based on the core genome, while the

249 potentially newly acquired copy thiE2 that is part of the operon thiCOSGEFD formed a distant
250 clade (Figure 2E), supporting HGT. Within this clade, the phylogeny of thiE2 is again congruent
251 with the core genome phylogeny, consistent with a single reacquisition event of thiCOSGEFD.
252 The same phylogenetic patterns were found for the other TBP genes that were shared across
253 branches (thiCOSGD, Figure S7), further supporting a single HGT of thiCOSGEFD to the *G.*
254 *morbifer* branch. Because TBP genes can occur on plasmids [74], we blast searched the
255 plasmids of the strains for which the plasmids were resolved for TBP genes, finding no evidence
256 for TBP genes (data not shown). In order to identify a potential donor of the operon, we blast
257 searched the sequence of the entire operon against the ncbi non-redundant nucleotide
258 database (nr). The best matching non-*Gluconobacter* sequences were from Rhodobacteraceae,
259 a phylogenetically distant bacterial family (Table S3). A single reacquisition event of the
260 essential TBP genes, as suggested by the concordance of the inserted operon with the core
261 gene phylogeny, implies that the TBP- strain DSM3504 lost the operon again in an independent
262 event, as depicted in Figure 2D (left).
263
264 **Discussion**
265 Microbial GWAS for host traits can benefit from strain level variation
266 We applied a microbial GWAS approach that associates bacterial genes with host phenotype
267 focusing on the genus *Gluconobacter*. Microbial GWAS approaches can be particularly
268 powerful, when pan-genomic variation of closely related bacterial strains can be leveraged, as
269 has been shown for e.g. virulence genes [75]. We showed that genetic variation between the
270 strains P1C6b, DSM2003, DSM2343, and DSM3504 (mean ANI = 95.5%) empowered us to
271 pinpoint the TBP (Figure S1). Variation between bacteria that have ANI > 95% is considered
272 strain level variation [76]. The only gene that had a higher association score for offspring
273 number than the TBP genes was a transposase. Transposases more frequently produce rare

presence-absence patterns because they are mobile and not linked as strongly to the rest of the genome as are non-mobile genetic elements. Therefore, we suspect that the high association score is an artefact of its mobility although we can not exclude an effect of the transposon on the number of fly offspring.

Variation between closely related microbes is important for host phenotypes

We observed significant variation of phenotypes between flies that were associated with closely related microbial strains. This supports the notion that strain level variation is important to consider when studying host-microbe interaction in animals, humans, and plants alike (e.g. [15, 18, 52, 77, 78]. In particular, in *D. melanogaster* evidence for the importance of variation between closely related bacteria is accumulating for life history of the host [26, 29, 47–51, 79]. Unawareness of strain level variation in bacterial effects on the host might have led to perceived inconsistency between studies [53].

Loss of the TBP and regain by HGT in the context of the evolution of host-microbe interaction

Fly fitness was strongly associated with genes from the TBP. In [37], thiamine produced by *Acetobacter pomorum* is sufficient to rescue the development of *D. melanogaster* on a thiamine free diet. Furthermore, the authors found that thiamine affects development, but not egg laying nor longevity. This is consistent with our result that flies raised on TBP- strains took longer to develop into pupae (Figure S3). The acquisition of B vitamins like thiamine (B1) is a typical benefit that insects receive from microbes [25, 80] and falls into the greater context of nutrient provisioning by microbes, which is a recurring theme in the evolution of host-microbe interaction [40, 41].

By tracing the genes of the TBP across genomes and the phylogeny, we found that the

298 pathway to produce Thiamine-P was regained most likely via HGT (Figure 2D). As such, our
299 study exemplifies that individual events of HGT into a host-associated microbe can alter host
300 fitness outcomes. Other studies that show an effect on host fitness via HGT to a host-associated
301 microbe involve defensive compounds produced by microbial symbionts in plants [81] and
302 animals [17, 82]. In our study, the increase in host fitness with the reacquisition of the TBP is
303 most likely mediated via nutritional benefits. Only a few similar cases have been described so
304 far. The most prominent may be the acquisition of vitamin B7 (biotin) synthesis by bed bug-
305 associated *Wolbachia* [83]. Similarly, cat flea-associated *Wolbachia* seem to have gained the
306 ability to produce biotin via HGT from closely related strains [84]. In ticks, *pabA* (and possibly
307 *pabB*) required for the synthesis of folic acid, was acquired by a *Coxiella* like symbiont through
308 horizontal gene transfer (HGT) from an Alphaproteobacterium [85] and is thought to affect tick
309 fitness.

310 A difference between our- and these previous studies is that although *Gluconobacter* is
311 frequently associated with *D. melanogaster* under natural conditions [54–57], it is neither an
312 obligate symbiont nor is it restricted to the fly gut. Given occurence in the environment and the
313 opportunity for horizontal transfer of the bacterium between hosts, there is currently no evidence
314 that the fly host significantly affects *Gluconobacter* evolution. Further, taking into account
315 evidence that the abundance of mobile metabolic genes is governed by selection [86, 87], we
316 must assume that loss and gain of the TBP must first of all benefit the bacterium to persist in the
317 bacterial genome. Thiamine is considered essential for bacteria [88], and thus the TBP can only
318 be lost if enough thiamine is available in the environment. Fruit, the main food substrate of
319 *Drosophila* under natural conditions [89] and the basis for the food used in our study, is mostly
320 poor in thiamine [90]. However, other bacteria that are associated with *Drosophila*, for example
321 other strains of *Gluconobacter* (this study), *Acetobacter pomorum* or *Lactobacillus plantarum*,
322 can produce thiamine [37, 39]. Under these conditions, it might be beneficial for a community

member to lose TBP as a result of selection for reduced metabolic expenditure [91]. This is consistent with TBP dependent fitness effects on the host being a byproduct of selection on thiamine production in the microbe.

Our study suggests that HGT to host associated-microbes could quickly increase host fitness. An increase in microbe mediated host fitness should also increase selection pressure on the host to favor that particular microbe that provides an increased benefit [41, 92, 93]. Waterworth et al. [94] suggested that the acquisition of genes to produce a defensive compound via HGT was key to the domestication of a bacterial defensive symbiont in beetles. We speculate that similar scenarios might be plausible for nutritional benefits in *Drosophila* because (i) mechanisms of host selection work efficiently for environmentally acquired bacteria [95–98], (ii) stable, strain specific associations of *Drosophila* with mutualistic bacteria have been reported [50], and (iii) evidence for host selection in the fly is accumulating in the laboratory [45, 99] as well as under natural conditions [55, 57]. Because the result of HGT here provides a potential benefit to the host under thiamine poor conditions that are often encountered under natural conditions e.g. on thiamine poor fruit, our study contributes to a broader view of adaptation that can involve a flexible microbiome [4, 100].

Data availability

Raw sequence data for the bacterial genomes, the assemblies used in this study, and the 16S rRNA gene sequences will be available under PRJNA656529

Acknowledgements

We thank Ruth Hershberg (Technion, Haifa, Israel), Christian Voolstra (Uni Konstanz, Germany), Lena Waidele (Uni Freiburg, Freiburg, Germany), and John Baines (MPI for Evolutionary Biology, Ploen, Germany) for helpful comments on the manuscript. This work was

funded by the DFG (STA1154/4-1; Projektnummer 408908608 and BA5529/1-1; Projektnummer 405974812). The authors acknowledge support by the state of Baden-Württemberg through bwHPC.

Competing interest

The authors declare no competing interests.

References

1. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. *PNAS* 2013; **110**: 3229–3236.
2. Douglas AE. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. *Annual Review of Entomology* 1998; **43**: 17–37.
3. Moran NA. Symbiosis as an adaptive process and source of phenotypic complexity. *Proceedings of the National Academy of Sciences* 2007; **104**: 8627–8633.
4. Bang C, Dagan T, Deines P, Dubilier N, Duschl WJ, Fraune S, et al. Metaorganisms in extreme environments: do microbes play a role in organismal adaptation? *Zoology* 2018; **127**: 1–19.
5. Salem H, Kirsch R, Pauchet Y, Berasategui A, Fukumori K, Moriyama M, et al. Symbiont digestive range reflects host plant breadth in herbivorous beetles. Current Biology 2020; 30: 2875-2886.e4.
6. Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ. Adaptation via symbiosis: recent spread of a *Drosophila* defensive symbiont. *Science* 2010; **329**: 212–215.
7. Himler AG, Adachi-Hagimori T, Bergen JE, Kozuch A, Kelly SE, Tabashnik BE, et al. Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits

and female bias. *Science* 2011; **332**: 254–256.

8. Bordenstein SR, Theis KR. Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLOS Biology* 2015; **13**: e1002226.

9. Moran NA, Yun Y. Experimental replacement of an obligate insect symbiont. *Proceedings of the National Academy of Sciences* 2015; **112**: 2093–2096.

10. Waidele L, Korb J, Voolstra CR, Künzel S, Dedeine F, Staubach F. Differential ecological specificity of protist and bacterial microbiomes across a set of termite species. *Front Microbiol* 2017; **8**.

11. Waidele L, Korb J, Voolstra CR, Dedeine F, Staubach F. Ecological specificity of the metagenome in a set of lower termite species supports contribution of the microbiome to adaptation of the host. *Animal Microbiome* 2019; **1**: 13.

12. Behrman EL, Howick VM, Kapun M, Staubach F, Bergland AO, Petrov DA, et al. Rapid seasonal evolution in innate immunity of wild *Drosophila melanogaster*. *Proc R Soc B* 2018; **285**: 20172599.

13. Rudman SM, Greenblum S, Hughes RC, Rajpurohit S, Kiratli O, Lowder DB, et al. Microbiome composition shapes rapid genomic adaptation of *Drosophila melanogaster*. *PNAS* 2019; **116**: 20025–20032.

14. Walters AW, Hughes RC, Call TB, Walker CJ, Wilcox H, Petersen SC, et al. The microbiota influences the *Drosophila melanogaster* life history strategy. *Molecular Ecology* 2020; **29**: 639–653.

15. Frank C, Werber D, Cramer JP, Askar M, Faber M, an der Heiden M, et al. Epidemic profile of shiga-toxin–producing *Escherichia coli* O104:H4 outbreak in Germany. *New England Journal of Medicine* 2011; **365**: 1771–1780.

16. Sachs JL, Skophammer RG, Bansal N, Stajich JE. Evolutionary origins and diversification of proteobacterial mutualists. *Proceedings of the Royal Society of London B: Biological*

Sciences 2014; **281**: 20132146.

17. Flórez LV, Scherlach K, Miller IJ, Rodrigues A, Kwan JC, Hertweck C, et al. An antifungal polyketide associated with horizontally acquired genes supports symbiont-mediated defense in *Lagria villosa* beetles. *Nat Commun* 2018; **9**: 1–10.
18. Melnyk RA, Hossain SS, Haney CH. Convergent gain and loss of genomic islands drive lifestyle changes in plant-associated *Pseudomonas*. *The ISME Journal* 2019; **13**: 1575–1588.
19. Sheppard SK, Didelot X, Meric G, Torralbo A, Jolley KA, Kelly DJ, et al. Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in *Campylobacter*. *PNAS* 2013; **110**: 11923–11927.
20. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol* 2016; **17**: 238.
21. Collins C, Didelot X. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. *PLOS Computational Biology* 2018; **14**: e1005958.
22. Sexton CE, Smith HZ, Newell PD, Douglas AE, Chaston JM. MAGNAMWAR: an R package for genome-wide association studies of bacterial orthologs. *Bioinformatics* 2018; **34**: 1951–1952.
23. White KM, Matthews MK, Hughes RC, Sommer AJ, Griffitts JS, Newell PD, et al. A Metagenome-wide association study and arrayed mutant library confirm *Acetobacter* lipopolysaccharide genes are necessary for association with *Drosophila melanogaster*. *G3 (Bethesda)* 2018; **8**: 1119–1127.
24. Chaston JM, Newell PD, Douglas AE. Metagenome-wide association of microbial determinants of host phenotype in *Drosophila melanogaster*. *mBio* 2014; **5**: e01631-14.
25. Judd AM, Matthews MK, Hughes R, Veloz M, Sexton CE, Chaston JM. Bacterial

methionine metabolism genes influence *Drosophila melanogaster* starvation resistance. *Appl Environ Microbiol* 2018; **84**.

26. Matthews MK, Wilcox H, Hughes R, Veloz M, Hammer A, Banks B, et al. genetic influences of the microbiota on the life span of *Drosophila melanogaster*. *Appl Environ Microbiol* 2020; **86**.

27. Brummel T, Ching A, Seroude L, Simon AF, Benzer S. *Drosophila* lifespan enhancement by exogenous bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 2004; **101**: 12974–12979.

28. Shin SC, Kim S-H, You H, Kim B, Kim AC, Lee K-A, et al. *Drosophila* microbiome modulates host developmental and metabolic homeostasis via insulin signaling. *Science* 2011; **334**: 670–674.

29. Storelli G, Defaye A, Erkosar B, Hols P, Royet J, Leulier F. *Lactobacillus plantarum* promotes *Drosophila* systemic growth by modulating hormonal signals through tor-dependent nutrient sensing. *Cell Metabolism* 2011; **14**: 403–414.

30. Téfit MA, Leulier F. *Lactobacillus plantarum* favors the early emergence of fit and fertile adult *Drosophila* upon chronic undernutrition. *Journal of Experimental Biology* 2017; jeb.151522.

31. Fast D, Duggal A, Foley E. Monoassociation with *Lactobacillus plantarum* disrupts intestinal homeostasis in adult *Drosophila melanogaster*. *mBio* 2018; **9**: e01114-18.

32. Gould AL, Zhang V, Lamberti L, Jones EW, Obadia B, Korasidis N, et al. Microbiome interactions shape host fitness. *PNAS* 2018; 201809349.

33. Obata F, Fons CO, Gould AP. Early-life exposure to low-dose oxidants can increase longevity via microbiome remodelling in *Drosophila*. *Nat Commun* 2018; **9**: 975.

34. Dobson AJ, Chaston JM, Newell PD, Donahue L, Hermann SL, Sannino DR, et al. Host genetic determinants of microbiota-dependent nutrition revealed by genome-wide

analysis of *Drosophila melanogaster*. *Nat Commun* 2015; **6**: 6312.

35. Yamada R, Deshpande SA, Bruce KD, Mak EM, Ja WW. Microbes promote amino acid harvest to rescue undernutrition in *Drosophila*. *Cell Reports* 2015; **10**: 865–872.

36. Chaston JM, Dobson AJ, Newell PD, Douglas AE. Host genetic control of the microbiota mediates the *Drosophila* nutritional phenotype. *Appl Environ Microbiol* 2016; **82**: 671–679.

37. Sannino DR, Dobson AJ, Edwards K, Angert ER, Buchon N. The *Drosophila melanogaster* gut microbiota provisions thiamine to its host. *mBio* 2018; **9**: e00155-18.

38. Henriques SF, Dhakan DB, Serra L, Francisco AP, Carvalho-Santos Z, Baltazar C, et al. Metabolic cross-feeding in imbalanced diets allows gut microbes to improve reproduction and alter host behaviour. *Nat Commun* 2020; **11**: 1–15.

39. Consuegra J, Grenier T, Baa-Puyoulet P, Rahioui I, Akherraz H, Gervais H, et al. *Drosophila*-associated bacteria differentially shape the nutritional requirements of their host during juvenile growth. *PLOS Biology* 2020; **18**: e3000681.

40. Ankrah NYD, Douglas AE. Nutrient factories: metabolic function of beneficial microorganisms associated with insects. *Environmental Microbiology* 2018; **20**: 2002–2011.

41. Moran NA, Ochman H, Hammer TJ. Evolutionary and ecological consequences of gut microbial communities. *Annual Review of Ecology, Evolution, and Systematics* 2019; **50**: 451–475.

42. Johnson EL, Heaver SL, Waters JL, Kim BI, Bretin A, Goodman AL, et al. Sphingolipids produced by gut bacteria enter host metabolic pathways impacting ceramide levels. *Nature Communications* 2020; **11**: 2471.

43. Inamine H, Ellner SP, Newell PD, Luo Y, Buchon N, Douglas AE. Spatiotemporally heterogeneous population dynamics of gut bacteria inferred from fecal time series data.

mBio 2018; **9**: e01453-17.

44. Keebaugh ES, Yamada R, Obadia B, Ludington WB, Ja WW. Microbial quantity impacts drosophila nutrition, development, and lifespan. *iScience* 2018; **4**: 247–259.

45. Storelli G, Strigini M, Grenier T, Bozonnet L, Schwarzer M, Daniel C, et al. *Drosophila* perpetuates nutritional mutualism by promoting the fitness of its intestinal symbiont *Lactobacillus plantarum*. *Cell Metabolism* 2018; **27**: 362-377.e8.

46. Matos RC, Schwarzer M, Gervais H, Courtin P, Joncour P, Gillet B, et al. D-Alanylation of teichoic acids contributes to *Lactobacillus plantarum*-mediated *Drosophila* growth during chronic undernutrition. *Nature Microbiology* 2017; **2**: 1635–1647.

47. Newell PD, Douglas AE. Interspecies interactions determine the impact of the gut microbiota on nutrient allocation in *Drosophila melanogaster*. *Appl Environ Microbiol* 2014; **80**: 788–796.

48. Newell PD, Chaston JM, Wang Y, Winans NJ, Sannino DR, Wong ACN, et al. In vivo function and comparative genomic analyses of the *Drosophila* gut microbiota identify candidate symbiosis factors. *Front Microbiol* 2014; **5**.

49. Winans NJ, Walter A, Chouaia B, Chaston JM, Douglas AE, Newell PD. A genomic investigation of ecological differentiation between free-living and *Drosophila*-associated bacteria. *Mol Ecol* 2017; **26**: 4536–4550.

50. Pais IS, Valente RS, Sporniak M, Teixeira L. *Drosophila melanogaster* establishes a species-specific mutualistic interaction with stable gut-colonizing bacteria. *PLOS Biology* 2018; **16**: e2005710.

51. Lee J, Han G, Kim JW, Jeon CO, Hyun S. Taxon-specific effects of *Lactobacillus* on *Drosophila* host development. *Microb Ecol* 2020; **79**: 241–251.

52. Schwarzer M, Makki K, Storelli G, Machuca-Gayet I, Srutkova D, Hermanova P, et al. *Lactobacillus plantarum* strain maintains growth of infant mice during chronic

undernutrition. *Science* 2016; **351**: 854–857.

53. Douglas AE. Contradictory results in microbiome science exemplified by recent *Drosophila* research. *mBio* 2018; **9**: e01758-18.

54. Staubach F, Baines JF, Künzel S, Bik EM, Petrov DA. Host species and environmental effects on bacterial communities associated with *Drosophila* in the laboratory and in the natural environment. *PLoS ONE* 2013; **8**: e70749.

55. Adair KL, Wilson M, Bost A, Douglas AE. Microbial community assembly in wild populations of the fruit fly *Drosophila melanogaster*. *The ISME Journal* 2018; 1.

56. Wang Y, Staubach F. Individual variation of natural *D. melanogaster*-associated bacterial communities. *FEMS Microbiol Lett* 2018; **365**.

57. Wang Y, Kapun M, Waidele L, Kuenzel S, Bergland AO, Staubach F. Common structuring principles of the *Drosophila melanogaster* microbiome on a continental scale and between host and substrate. *Environmental Microbiology Reports* 2020; **12**: 220–228.

58. Geissmann Q. OpenCFU, a new free and open-source software to count cell colonies and other circular objects. *PLoS One* 2013; **8**.

59. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013; **79**: 5112–5120.

60. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* 2009; **75**: 7537–7541.

61. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 2015; **31**: 587–589.

62. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the

quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015; gr.186072.114.

63. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013; **29**: 1072–1075.

64. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**: 2068–2069.

65. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 2020; **36**: 1925–1927.

66. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology* 2018; **36**: 566–569.

67. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS ONE* 2010; **5**: e9490.

68. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; **30**: 1312–1313.

69. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 2004; **5**: 113.

70. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 2012; **61**: 539–542.

71. Ryu J-H, Kim S-H, Lee H-Y, Bai JY, Nam Y-D, Bae J-W, et al. Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in *Drosophila*. *Science* 2008; **319**: 777–782.

72. Galac MR, Lazzaro BP. Comparative pathology of bacteria in the genus *Providencia* to a natural host, *Drosophila melanogaster*. *Microbes and Infection* 2011; **13**: 673–683.

548 73. Ding W, Baumdicker F, Neher RA. panX: pan-genome analysis and exploration. *Nucleic*
549 *Acids Res* 2018; **46**: e5–e5.
550 74. Karunakaran R, Ebert K, Harvey S, Leonard ME, Ramachandran V, Poole PS. Thiamine
551 is synthesized by a salvage pathway in *Rhizobium leguminosarum* bv. *viciae* strain 3841.
552 *J Bacteriol* 2006; **188**: 6661–6668.
553 75. Gori A, Harrison OB, Mlia E, Nishihara Y, Chan JM, Msefula J, et al. Pan-GWAS of
554 *Streptococcus agalactiae* highlights lineage-specific genes associated with virulence and
555 niche adaptation. *mBio* 2020; **11**.
556 76. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI
557 analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature*
558 *Communications* 2018; **9**: 5114.
559 77. Brouwer MSM, Roberts AP, Hussain H, Williams RJ, Allan E, Mullany P. Horizontal gene
560 transfer converts non-toxigenic *Clostridium difficile* strains into toxin producers. *Nat*
561 *Commun* 2013; **4**: 2601.
562 78. Van Rossum T, Ferretti P, Maistrenko OM, Bork P. Diversity within species: interpreting
563 strains in microbiomes. *Nature Reviews Microbiology* 2020; 1–16.
564 79. Obadia B, Güvener ZT, Zhang V, Ceja-Navarro JA, Brodie EL, Ja WW, et al. Probabilistic
565 invasion underlies natural gut microbiome stability. *Current Biology* 2017; **27**: 1999-
566 2006.e8.
567 80. Douglas AE. The B vitamin nutrition of insects: the contributions of diet, microbiome and
568 horizontally acquired genes. *Current Opinion in Insect Science* 2017; **23**: 65–69.
569 81. Pinto-Carbó M, Sieber S, Dessein S, Wicker T, Verstraete B, Gademann K, et al.
570 Evidence of horizontal gene transfer between obligate leaf nodule symbionts. *The ISME*
571 *Journal* 2016; **10**: 2092–2105.
572 82. Lopanik NB. Chemical defensive symbioses in the marine environment. *Functional*

Ecology 2014; **28**: 328–340.

83. Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T. Evolutionary origin of insect–*Wolbachia* nutritional mutualism. *PNAS* 2014; **111**: 10257–10262.

84. Driscoll TP, Verhoeve VI, Brockway C, Shrewsberry DL, Plumer ML, Sevdalis SE, et al. Evolution of *Wolbachia* mutualism and reproductive parasitism: insight from two novel strains that co-infect cat fleas. *bioRxiv* 2020; 2020.06.01.128066.

85. Smith TA, Driscoll T, Gillespie JJ, Raghavan R. A *Coxiella*-Like endosymbiont is a potential vitamin source for the lone star tick. *Genome Biol Evol* 2015; **7**: 831–838.

86. Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 2011; **480**: 241–244.

87. Brito IL, Yilmaz S, Huang K, Xu L, Jupiter SD, Jenkins AP, et al. Mobile genes in the human microbiome are structured from global to individual scales. *Nature* 2016; **535**: 435–439.

88. Costliow ZA, Degnan PH. Thiamine acquisition strategies impact metabolism and competition in the gut microbe *Bacteroides thetaiotaomicron*. *mSystems* 2017; **2**.

89. Mansourian S, Enjin A, Jirle EV, Ramesh V, Rehermann G, Becher PG, et al. Wild African *Drosophila melanogaster* are seasonal specialists on marula fruit. *Current Biology* 2018; **28**: 3960-3968.e3.

90. Ross AC. Modern Nutrition in Health and Disease. 2014. Wolters Kluwer Health/Lippincott Williams & Wilkins.

91. Morris JJ. Black Queen evolution: the role of leakiness in structuring microbial communities. *Trends in Genetics* 2015; **31**: 475–482.

92. Bull JJ, Rice WR. Distinguishing mechanisms for the evolution of co-operation. *Journal of Theoretical Biology* 1991; **149**: 63–74.

93. Foster KR, Schluter J, Coyte KZ, Rakoff-Nahoum S. The evolution of the host microbiome as an ecosystem on a leash. *Nature* 2017; **548**: 43–51.

94. Waterworth SC, Flórez LV, Rees ER, Hertweck C, Kaltenpoth M, Kwan JC. Horizontal gene transfer to a defensive symbiont with a reduced genome in a multipartite beetle microbiome. *mBio* 2020; **11**.

95. Kiers ET, Rousseau RA, West SA, Denison RF. Host sanctions and the legume–rhizobium mutualism. *Nature* 2003; **425**: 78–81.

96. Kremer N, Philipp EER, Carpentier M-C, Brennan CA, Kraemer L, Altura MA, et al. Initial symbiont contact orchestrates host-organ-wide transcriptional changes that prime tissue colonization. *Cell Host & Microbe* 2013; **14**: 183–194.

97. Koehler S, Gaedeke R, Thompson C, Bongrand C, Visick KL, Ruby E, et al. The model squid–*Vibrio* symbiosis provides a window into the impact of strain- and species-level differences during the initial stages of symbiont engagement. *Environmental Microbiology* 2019; **21**: 3269–3283.

98. Wendlandt CE, Regus JU, Gano-Cohen KA, Hollowell AC, Quides KW, Lyu JY, et al. Host investment into symbiosis varies among genotypes of the legume *Acmispon strigosus*, but host sanctions are uniform. *New Phytologist* 2019; **221**: 446–458.

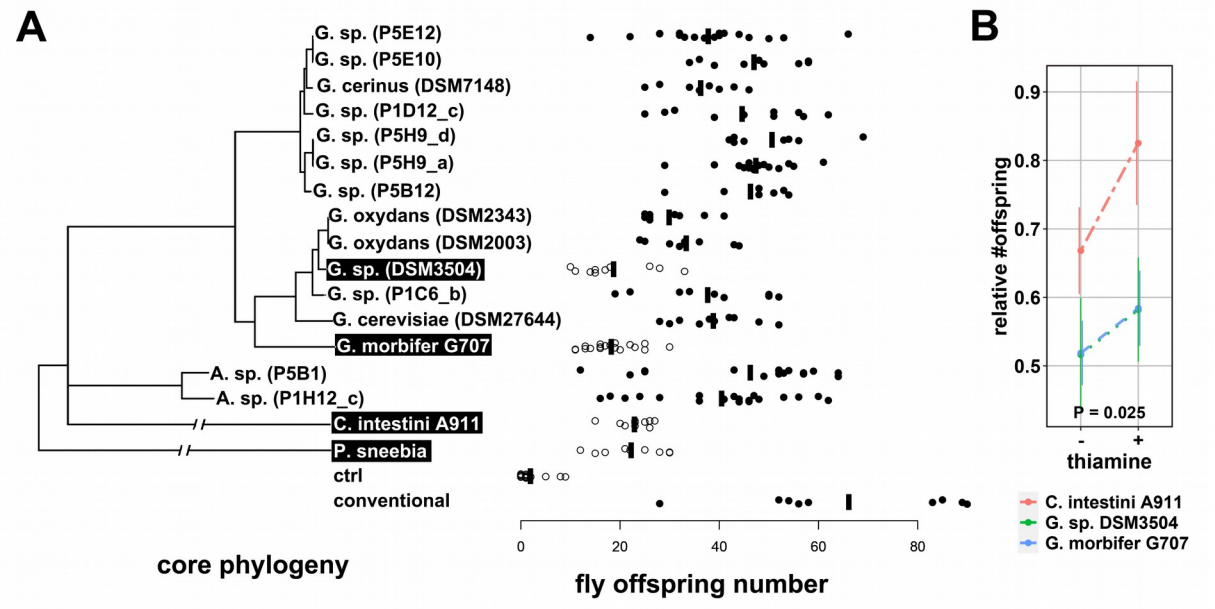
99. Adair KL, Bost A, Bueno E, Kaunisto S, Kortet R, Peters-Schulze G, et al. Host determinants of among-species variation in microbiome composition in drosophilid flies. *ISME J* 2020; **14**: 217–229.

100. Voolstra CR, Ziegler M. Adapting with Microbial Help: Microbiome flexibility facilitates rapid responses to environmental change. *BioEssays* 2020; **42**: 2000004.

622 **Table 1** List of the ten genes that were most strongly associated with offspring number
623 according to PA-association scores. All associations were confirmed by at least one of three
624 methods from treeWAS [21].

Gene Name	Annotation	PA score	treeWAS terminal p-value	treeWAS simultaneous p-value	treeWAS subsequent p-value
	transposase	4.62	0.000009	0.013	0.003
ThiO	putative thiamine biosynthesis oxidoreductase	4.34	0.0002	0.003	0.02
ThiG	thiazole synthase	4.34	0.0002	0.003	0.02
ThiS	thiamine biosynthesis protein	4.34	0.0002	0.003	0.02
ThiC	thiamine biosynthesis protein	4.34	0.0002	0.003	0.02
ThiD	phosphomethylpyrimidine kinase	4.34	0.0002	0.003	0.02
	ferric iron siderophore receptor	4.21	0.000009	0.11	0.001
	oxidoreductase	4.21	0.000009	0.11	0.001
	LysR family transcriptional regulator	4.21	0.000009	0.11	0.001
	Methyltransferase domain protein	4.06	0.0003	0.009	0.89

626
627
628
629
630
631
632



634 **Figure 1** (A) left: Bacterial tree based on 134 single-copy orthologs. Leaf labels of bacteria that
635 do not carry a complete thiamine biosynthesis pathway are on black background. Right: Number
636 of offspring produced by mono-associated CantonS females; vertical bars: median; ctrl: axenic
637 flies; conventional: flies reinfected with lab microbiota. (B) The effect of thiamine treatment (x-
638 axis) on the number of offspring in mono-associated females. Under thiamine treatment, the
639 number of offspring increased for the strains that do not possess a complete thiamine
640 biosynthesis pathway relative to strains that possess the complete pathway. P-value determined
641 by linear mixed effects model. Error bars represent the standard error of the mean.
642
643
644
645
646
647

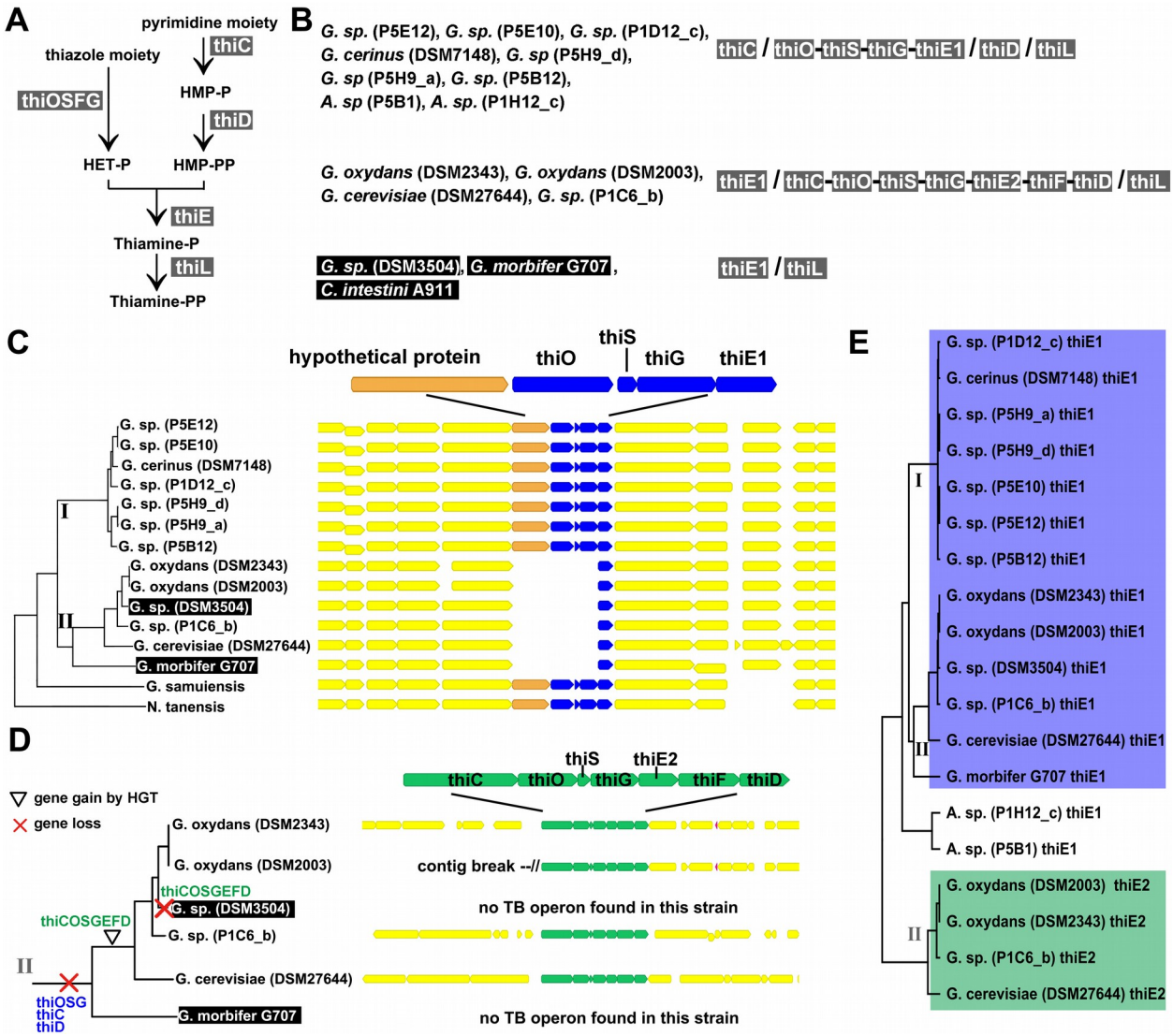


Figure 2 (A) The thiamine biosynthesis pathway in acetic acid bacteria (B) Overview of thiamine biosynthesis genes in the analyzed bacteria. Note that the function of thiF that appears to be missing in the strains of the upper row can be replaced by the function of the homolog MoeB (Rodionov et al., 2002) that we found in all strains analyzed. Genes forming an operon are separated by a hyphen. Genes from different loci are separated by slashes. (C) Synteny of the flanking regions of thiamine biosynthesis genes in *Gluconobacter*. thiOSG are missing on the *G. morbifer* branch (II) at this locus. Thiamine biosynthesis genes are in blue. (D) right: The complete pathway to synthesize Thiamine-P (green) forms an operon on the *G. morbifer* branch

(branch II); left: The phylogeny depicts the inferred evolutionary scenario on branch II. (E) Phylogeny of thiE. *G. oxydans* DSM2343, *G. oxydans* DSM2003, *G. sp.* P1C6_b, *G. cerevisiae* DSM27644 have two copies of thiE, thiE1 (blue) and thiE2 (green). The phylogeny of thiE1 (blue background) is congruent with the core genome phylogeny. ThiE2 (green background) forms a distinct clade that is more distant than thiE from *Acetobacter*, indicating HGT from a distant clade.