

RESEARCH ARTICLE

Open Access

# Horizontal gene transfer-mediated bacterial strain variation affects host fitness in *Drosophila*



Yun Wang<sup>1</sup>, Franz Baumdicker<sup>2,3</sup>, Paul Schweiger<sup>4</sup>, Sven Kuenzel<sup>5</sup> and Fabian Staubach<sup>1\*</sup> 

## Abstract

**Background:** How microbes affect host fitness and environmental adaptation has become a fundamental research question in evolutionary biology. To better understand the role of microbial genomic variation for host fitness, we tested for associations of bacterial genomic variation and *Drosophila melanogaster* offspring number in a microbial Genome Wide Association Study (GWAS).

**Results:** We performed a microbial 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. While an effect of thiamine on fly development has been described, we show that strain variation in TBP between bacterial isolates from wild-caught *D. melanogaster* contributes to variation in offspring production by the 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).

**Conclusion:** Our study emphasizes the importance of strain variation and highlights that HGT can add to microbiome flexibility and potentially to host adaptation.

**Keywords:** *Drosophila*, Microbiome, GWAS, Horizontal gene transfer, Lateral gene transfer

## Background

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 evolutionary 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

\* Correspondence: [fabian.staubach@biologie.uni-freiburg.de](mailto:fabian.staubach@biologie.uni-freiburg.de)

<sup>1</sup>Department of Evolutionary Biology and Ecology, Institute of Biology I, Albert Ludwigs University Freiburg, Freiburg, Germany

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

important for starvation resistance [25] as well as life span [26].

For targeting 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. Microbe-affected host phenotypes include the life history traits such as development time, fecundity, and life span as well as size of the adults [14, 26–33]. These traits are components of fitness, emphasizing the potential importance of microbes in host evolution and adaptation [14, 32]. 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 metazoan–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 differs 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].

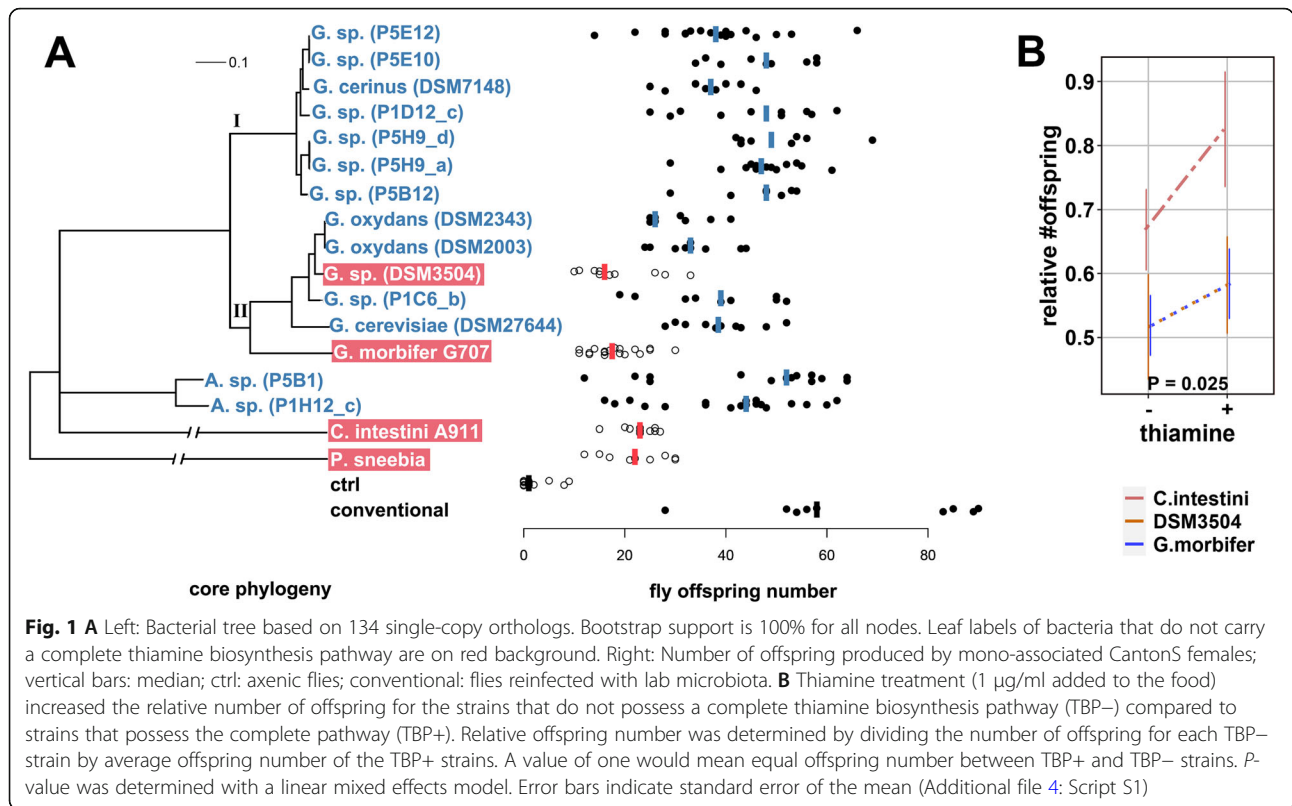
Because variation between closely related bacteria is important for the interaction of the host and its microbiome, it is also important to consider closely related microbes in studies that aim at elucidating the molecular underpinnings of host–microbe interaction. At the same time, limiting a GWAS to the pan-genome of closely related microbes might offer particular power to the approach: limitation to a narrower range of genes that vary in their presence-absence patterns in a similar genomic background, microbial genes that affect the host can be more precisely pinpointed. 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 resolution 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 offspring number per female fly as a fitness component 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.

## 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* [58] 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 hemolymph [59]. All bacterial genomes analyzed were >99% complete with the exception of *P. sneebia* (>96%, Additional file 1: Table S1). The mean number of offspring varied significantly between flies mono-associated with different isolates ( $P = 4.2 \times 10^{-15}$ , Kruskal-Wallis test, Fig. 1A) up to a 2.8-fold difference between *Gluconobacter morbifer* and *Gluconobacter sp.* P5H9\_d. Differences between bacterial strains were also a significant covariate of offspring number when we accounted for bacterial loads per fly ( $P = 1.4 \times 10^{-4}$ , linear model). Furthermore, bacterial load alone was not significantly associated with fly offspring number ( $P = 0.11$ , linear model), suggesting that not only bacterial biomass affects fly fitness. Presence-absence patterns of 11,269 genes were tested for association with the number of offspring that mono-associated females produced using the PA method [60]. Associations were confirmed by permutation tests and TreeWAS [21] (Table 1, Additional file 2 and 3: Figure S1, Table S2). The six highest PA scores depended strongly on presence-absence patterns between the closely related strains P1C6\_b, DSM2003, DSM2343, and DSM3504 (mean ANI = 95.5%) in the branch that



includes *G. morbifer* (Additional file 2: 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, Fig. 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

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

Gene name	Annotation	PA score	treeWAS terminal <i>p</i> -value	treeWAS simultaneous <i>p</i> -value	treeWAS subsequent <i>p</i> -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

medians,  $n = 17$ , Additional file 2: Figure S2), providing no evidence for larval crowding or reduced adult size. Significance of all  $p$ -values was confirmed in a linear model framework that accounts for bacterial load and also in a phylogenetic ANOVA (Additional file 4: Script S1).

We hypothesized that if bacterial thiamine increased the offspring number of females reared on TBP+ strains, supplementing the diet of flies reared on TBP- strains with thiamine would increase offspring number when compared to TBP+ strains. To test this, we supplemented the diet of females that were mono-associated with TBP- strains (*G. sp.* DSM3504, *G. morbifer* G707, *C. intestini* A911) with thiamine and applied the same assay for offspring number as for the initial GWAS. Supplementing the diet of flies reared on closely related TBP+ strains served as control (*G. oxydans* DSM2343, *G. oxydans* DSM2003, *G. sp.* P1C6\_b, *G. cerevisiae* DSM27644). In order to account for variation between experiments, we calculated the relative offspring number between TBP- and TBP+ strains in the initial unsupplemented experiment (“thiamine treatment -”, Fig. 1B) and the supplementation experiment (“thiamine treatment +”, Fig. 1B). Indeed, the relative offspring number on TBP- strains increased with thiamine supplementation ( $P = 0.025$ , linear mixed effects model, Fig. 1B), supporting a role of thiamine production in the number of offspring that flies produced. We found no evidence that the addition of thiamine increased bacterial loads of TBP- strains ( $P = 0.85$ , generalized linear model), suggesting that the increase in offspring number is not due to an increase in bacterial biomass alone.

#### 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 (Fig. 2A) in *Gluconobacter*, we analyzed the synteny of the underlying loci in a phylogenetic framework. The strains in the upper two panels of Fig. 2B possess all genes required for thiamine biosynthesis. A closer inspection of TBP genes on the *G. morbifer* branch (II in Fig. 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 (Fig. 2C) at the locus that is syntenic with branch I. The same pattern was found for thiC and thiD (Additional file 2: Figure S3). thiOSG (Fig. 2C), thiC, and thiD (Additional file 2: Figure S4) are present in the closely related bacteria *Gluconobacter samuiensis* and *Neokomagataea tanensis* at syntenic loci, suggesting deletion on branch II. The strains with an intact operon on branch II carried a TBP operon at loci not syntenic with the locus shown in Fig.

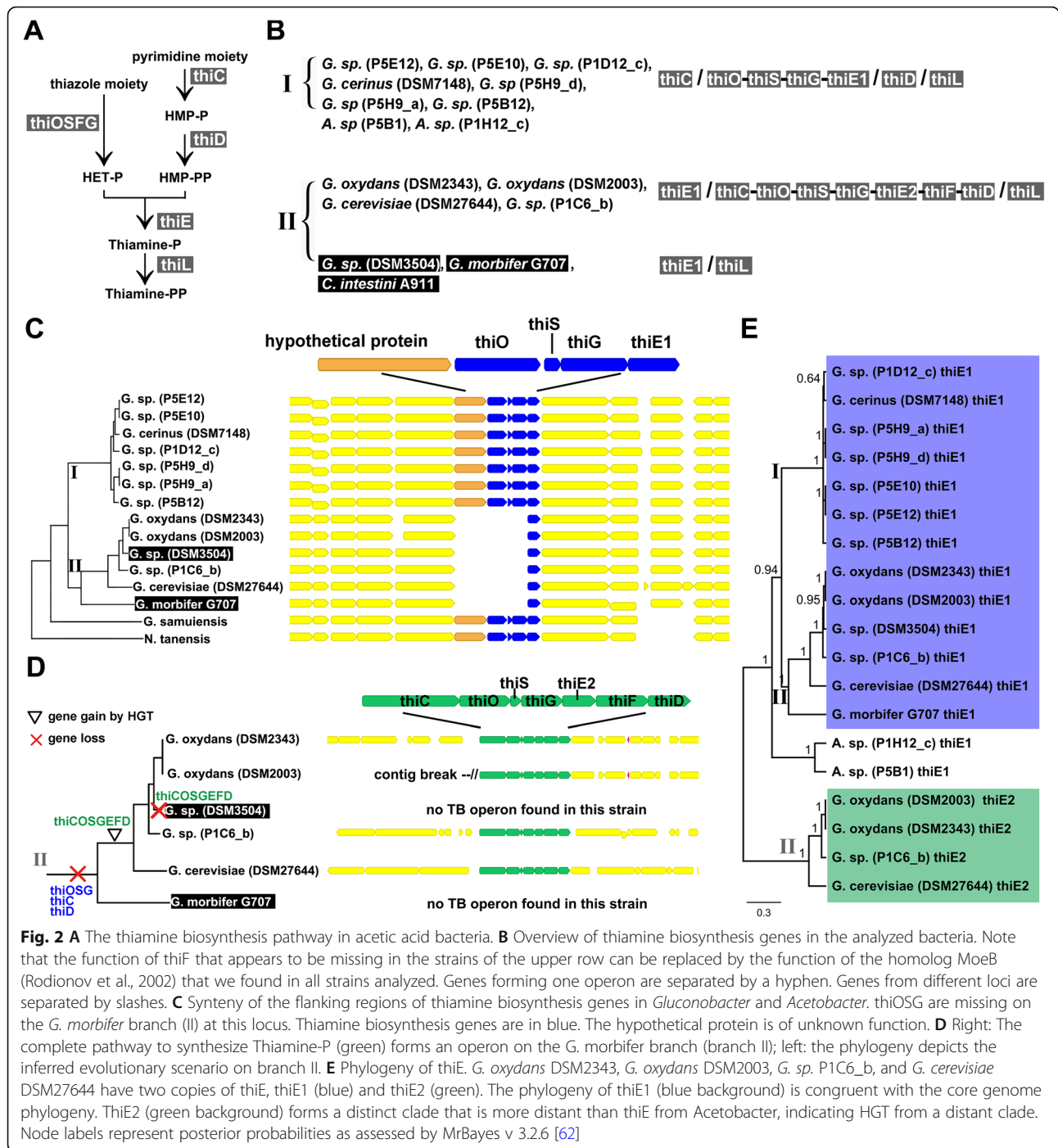
2C as evident from different flanking genes (Fig. 2D, Additional file 2: Figure S5), suggesting insertion.

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 Fig. 2C, followed the phylogeny based on the core genome, while the potentially newly acquired copy thiE2 that is part of the operon thiCOSGEFD formed a distant clade (Fig. 2E), supporting HGT. Within this clade, the phylogeny of thiE2 is again congruent with the core genome phylogeny, consistent with a single reacquisition event of thiCOSGEFD. The same phylogenetic patterns were found for the other TBP genes that were shared across branches (thiCOSGD, Additional file 2: Figure S6), further supporting a single HGT of thiCOSGEFD to the *G. morbifer* branch. Because TBP genes can occur on plasmids [61], we blast searched the plasmids of the strains for which the plasmids were resolved for TBP genes, finding no evidence for TBP genes (data not shown). In order to identify a potential donor of the operon, we blast searched the sequence of the entire operon against the ncbi non-redundant nucleotide database (nr). The best matching non-*Gluconobacter* sequences were from Rhodobacteraceae, a phylogenetically distant bacterial family (Additional file 5: Table S3). A closer inspection of the non-*Gluconobacter* blast hits with the highest scores (query coverage 79–82%, ~73% identity, Additional file 5: Table S3) revealed that gene order within the operon, but not synteny of flanking genes, was conserved (Additional file 2: Figure S7). Despite a modest difference in GC-content between the potential donors (~66% vs 62% in *Gluconobacter oxydans* DSM2343), the GC-content of the putatively inserted operon did not differ from that of the genomic background (Additional file 2: Figure S8), providing no evidence for a recent acquisition from any of the top 3 blast hits. Furthermore, the best non-*Gluconobacter* blast hits were in marine bacteria. Taken together, this implies that the true donors remain enigmatic. A single reacquisition event of the essential TBP genes in the past, close to the base of clade II, as suggested by the concordance of the inserted operon with the core gene phylogeny, implies that the TBP- strain DSM3504 lost the operon again in an independent event, as depicted in Fig. 2D (left).

## Discussion

### Microbial GWAS for host traits can benefit from strain level variation

We applied a microbial GWAS approach that associates bacterial genes with host phenotype focusing on the genus *Gluconobacter*. Microbial GWAS approaches can be particularly powerful, when pan-genomic variation of closely related bacterial strains can be leveraged, as has



**Fig. 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 one 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* and *Acetobacter*. thiOSG are missing on the *G. morbifer* branch (II) at this locus. Thiamine biosynthesis genes are in blue. The hypothetical protein is of unknown function. **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, and *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. Node labels represent posterior probabilities as assessed by MrBayes v 3.2.6 [62]

been shown for, e.g., virulence genes [63]. We showed that genetic variation below the species level between the strains P1C6\_b, DSM2003, DSM2343, and DSM3504 (mean ANI = 95.5%) empowered us to pinpoint the TBP (Additional file 2: Figure S1). Variation between bacteria that have ANI > 95% is considered strain level variation [64]. The only gene that had a higher association score for offspring 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 artifact of its mobility although we cannot exclude an effect of the transposon on the number of fly offspring. From the other genes with significant associations with offspring number presented in Table 1, the oxidoreductase, LysR family transcriptional regulator, and the methyltransferase domain

protein, a plausible link to fly offspring number is more difficult to test. Nonetheless, these genes might also affect fly offspring number. The ferric iron siderophore receptor is located close to the inserted thiamine operon in P1C6\_b and DSM27644 which also possesses a gene with the same annotation at that locus, as is apparent from Additional file 2: Figure S5. While it seems possible that this gene contributes to fly fitness, it must be considered that DSM3504 that lacks TBP genes and confers relatively low fitness to the host also carries a ferric iron siderophore receptor that is orthologous to that shown in Additional file 2: Figure S5 in DSM27644. Given this and the demonstrated fitness effects of the thiamine supplement (Fig. 1B), we must assume that this gene received a high association score mainly due to linkage to the TBP genes.

#### 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, 65, 66]. 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, 67]. Unawareness of strain level variation in bacterial effects on the host might have led to perceived inconsistency between studies [53]. Furthermore, our study provides an example of the limits of 16S rRNA sequencing in functional inference: The 16S rRNA gene sequences, as assessed by full length Sanger sequencing of strains P1C6\_b (TBP+) and DSM3504 (TBP–), were identical (Additional file 2: Figure S9).

#### The number of offspring as a component of fitness

As a fitness component, we assessed the number of adult offspring produced after 16 days. As such, our assay captures developmental rates and fecundity on a time scale that we consider highly relevant for the reproductive success of an organism that is adapted to an ephemeral resource, rotting fruit [68]. The effect of thiamine on offspring number that we describe is consistent with previously described effects of bacterial thiamine on *D. melanogaster* development and survival to adulthood [37]. Yet, a limitation of our study is that other components of life time reproductive success, and thus fitness, were not directly assessed, in particular egg laying and longevity. However, Sannino et al. [37] also showed that bacterial thiamine neither affects egg laying nor longevity in *D. melanogaster*. This suggests that the effect of thiamine on the number of offspring that we observed is

directly related to lifetime reproductive output, and thus fly fitness.

#### The loss and regain of the TBP by HGT in the context of the evolution of host–microbe interaction

Offspring number was strongly associated with genes from the TBP. The acquisition of B vitamins like thiamine (B1) is a typical benefit that insects receive from microbes [25, 69, 70] 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 pathway to produce Thiamine-P was regained most likely via HGT (Fig. 2D). As such, our study exemplifies that individual events of HGT into a host-associated microbe can alter host fitness outcomes. Other studies that show an effect on host fitness via HGT to a host-associated microbe involve defensive compounds produced by microbial symbionts in plants [71] and animals [17, 72]. In our study, the increase in host fitness with the reacquisition of the TBP is most likely mediated via nutritional benefits. Only a few similar cases have been described so far. The most prominent may be the acquisition of vitamin B7 (biotin) and vitamin B2 (riboflavin) synthesis by planthopper-associated *Wolbachia* [73]. Similarly, bed bug-associated *Wolbachia* [74] and cat flea-associated *Wolbachia* seem to have gained the ability to produce biotin via HGT [75]. In ticks, *pabA* (and possibly *pabB*) required for the synthesis of folic acid was acquired by a *Coxiella*-like symbiont through horizontal gene transfer (HGT) from an Alphaproteobacterium [76] and is thought to affect tick fitness. A *Rickettsia* endosymbiont of deer ticks has acquired the genes necessary for the synthesis of biotin on a plasmid [77]. In this study, as in ours, a complete operon has been transferred.

A difference between our and these previous studies is that although *Gluconobacter* is frequently associated with *D. melanogaster* under natural conditions [54–57], it is neither an obligate symbiont nor is it restricted to the fly gut. While it is interesting that the natural fly isolates of *Gluconobacter* are TBP+, there is currently no evidence that the fly host significantly affects *Gluconobacter* evolution given its occurrence in the environment and the opportunity for horizontal transfer of the bacterium between hosts. Further, taking into account evidence that the abundance of mobile metabolic genes is governed by selection [78, 79], we must assume that loss and gain of the TBP must first of all benefit the bacterium to persist in the bacterial genome. Thiamine is considered essential for bacteria [80], and thus, the TBP can only be lost if enough thiamine is available in the environment. Fruit, the main food substrate of *Drosophila*

under natural conditions [68] and the basis for the food used in our study, is mostly poor in thiamine [81]. However, other bacteria that are associated with *Drosophila*, for example other strains of *Gluconobacter* (this study), *Acetobacter pomorum* or *Lactobacillus plantarum*, 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 [82]. 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, 83, 84]. Waterworth et al. [85] 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 [86–89]; (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, 90] as well as under natural conditions [55, 57].

## Conclusion

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, 91].

## Methods

### Fitness assays

Canton-S stocks were kept at 25 °C on a 12:12 light:dark cycle 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 min, 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:dark 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 h and

normalized to the same optical density (OD<sub>600</sub> = 0.6). One hundred fifty microliters of OD normalized medium was added directly on 10-ml 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). The food was autoclaved without propionic acid and propionic acid added after the food had cooled down. Please note that yeast can in principle serve as a thiamine source, but autoclaving might have reduced the thiamine content of the food, as thiamine is heat labile [92, 93], such that it became a limiting factor for offspring number. Axenic females were transferred to the vial immediately after addition of the bacterial culture. 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. On the day 16, 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 supplementation experiment, food was prepared as described above but we added 1 µg/ml thiamine to the food after autoclaving. That concentration has proven effective for phenotypic rescue in [37]. All statistical analyses were performed in R and can be found in Additional file 4: 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). Please note that glycerol is a standard cryoprotectant that allows to keep bacteria alive at –80 °C for extended time periods [94]. Effective conservation of live bacteria with this method is supported by CFU counts in the range of 10<sup>2</sup>–10<sup>5</sup> CFUs (Additional file 6: Table S4) per fly for the majority of our samples, matching expectations from the literature well [67]. Finally, all samples underwent the freezing procedure in our randomization scheme that should prevent systematic treatment effects. Nonetheless, we cannot fully exclude that strain variation in the response to cryopreservation might have affected colony counts. For CFU enumeration 3–6 replicates per bacterial isolate were picked. To this end, 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 h. CFU counts were done visually or with the OpenCFU software [95] (Additional file 6: 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 that flies were inoculated with on fly offspring using 16S rRNA gene sequencing (Additional file 2: Figure S10). 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, 96]. Sequencing data were analyzed using mothur [97] (see Additional file 7: 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% (Additional file 2: Figure S10A) 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, but there was also some evidence for contaminating 16S gene sequences, which were likely introduced during the PCR or sequencing steps (Additional file 2: Figure S10B).

### 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 (Additional file 1: 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 sequenced using Illumina MiSeq technology and assembled with the A5 MiSeq assembler [98]. Completeness and contamination were assessed with checkM v1.1.2 [99], using standard settings. Assembly statistics were generated with QUAST v5.0.2 [100]. Annotation was performed with prokka v1.1 [101] or imported from GenBank. Average nucleotide identity (ANI) was computed with fastANI (v0.1.2). New isolates were taxonomically classified, using GTDBtk (v0.1.4) [102]. FastANI and GTDBtk were run on the kbase web interface [103].

### Pan-genome clustering and phylogenetic trees

Genomes were analyzed using the panX analysis pipeline [60] with standard parameters (Additional file 8: script S3). Genes were grouped into 11,269 clusters of

homologous sequences, including clusters with a single gene. Thereby, the presence and absence of each gene cluster in the 17 genomes was estimated. Based on the alignments of all 134 inferred single-copy gene clusters that are present in all 17 genomes, panX reconstructs a phylogenetic tree (Fig. 1). For this phylogeny, FastTree 2 [104] and RaxML [105] were applied to all variable positions from these alignments. To create bootstrap values, we used a separate raxml call with the -b option based on the alignments created by panX (see Additional file 8: script S3). For the phylogeny of gene clusters, nucleotide sequences were aligned using MUSCLE v3.8.425 [106] and the tree was built by MrBayes 3.2.6 [62], using a molecular clock with default parameters in the Geneious software suit v1.1 (Biomatters ltd.).

### Microbial pan-genome-wide association study

We calculated the gene presence absence association score (PA score) between each predicted cluster of homologous genes and fly offspring number. That is, if  $D_g$  is the difference between the mean fly offspring for strains with and without gene  $g$ ,  $\sigma$  is the global standard deviation of fly offspring for all strains and  $n_g$  is the number gene gains and losses as inferred from the phylogeny. The 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.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-021-01124-y>.

**Additional file 1: Table S1.** List of bacterial strains used in the experiments including assembly information.

**Additional file 2: Figures S1–S10. Fig. S1.** Distribution of PA scores.

**Fig. S2.** Offspring weight. **Fig. S3.** ThiC and thiD missing at syntenic loci.

**Fig. S4.** Synteny in *Swingsia samuiensis* and *Neokomagataea tanensis* for thiC, thiD, and thiOSG. **Fig. S5.** Detailed view of thiamine operon

insertion locus with flanking genes. **Fig. S6.** Phylogeny of the HGT TBP

genes thiC, thiD, and thiOSG. **Fig. S7.** Thiamine operon loci in potential

donors. **Fig. S8.** GC-content at putative thiamine operon insertion sites.

**Fig. S9.** multiple sequence alignment of 16S rRNA genes. **Fig. S10.** Con-

tamination control using 16S rRNA gene sequencing after the experiment.

**Additional file 3: Table S2.** Full Panx PA scores and Treewas results table.

**Additional file 4: Script S1.** statistical analyses.

**Additional file 5: Table S3.** Blast results for HGT operon.

**Additional file 6: Table S4.** Fitness experiment data (offspring number, CFUs, fly weight).



**Additional file 7: Script S2.** 16S rRNA gene sequence analysis with mothur.

**Additional file 8: Script S3.** Microbial GWAS.

### Acknowledgements

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

### Authors' contributions

YW, PS, and FS designed the experiments. YW and SK performed the experiments. YW, FB, and FS analyzed that data. YW, FB, and FS drafted the manuscript. All authors have read and revised the manuscript and approved the final manuscript.

### Authors' information

Twitter handle: Yun Wang, @FulaibaoWang; Franz Baumdicker, @fbaumdicker; Fabian Staubach, @FabianStaubach; Paul Schweiger, @drpschweig

### Funding

This work was funded by the DFG (STA1154/4-1; Projektnummer 408908608 and BA5529/1-1; Projektnummer 405974812). In addition, FB is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2064/1 – Project number 390727645, and EXC 2124 – Project number 390838134. The authors acknowledge support by the state of Baden-Württemberg through bwHPC. Open Access funding enabled and organized by Projekt DEAL.

### Availability of data and materials

The sequencing data generated and analyzed during the current study are available in the NCBI SRA repository, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA656529> [107]. The bacterial genomes and the assemblies are either available under SRA number SRS7200184 – SRS7200194 [107] or from the sources described in Additional file 1: Table S1 [108–113] with raw data available in [114–119]. The 16S rRNA gene sequences are available under SRA number SRS7426971 – SRS7427068 [107] with the sample titles corresponding to the column "name\_in\_mothur" in Additional file 6\_Table S4. Sequences of the closely related species used for the alignment in Additional file 2: Figure S4 are from [120, 121] with raw data available in [122, 123].

### Declarations

#### Ethics approval and consent to participate

Not applicable

#### Consent for publication

All authors have read and agreed to publish the manuscript.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Department of Evolutionary Biology and Ecology, Institute of Biology I, Albert Ludwigs University Freiburg, Freiburg, Germany. <sup>2</sup>Cluster of Excellence 'Controlling Microbes to Fight Infections', University of Tübingen, Tübingen, Germany. <sup>3</sup>Department of Mathematical Stochastics, University of Freiburg, Freiburg, Germany. <sup>4</sup>Microbiology Department, University of Wisconsin – La Crosse, La Crosse, WI, USA. <sup>5</sup>Max Planck Institute for Evolutionary Biology, Ploen, Germany.

Received: 3 February 2021 Accepted: 13 August 2021

Published online: 27 September 2021

### References

- 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(9):3229–36. <https://doi.org/10.1073/pnas.1218525110>.
- Douglas AE. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera. *Annu Rev Entomol*. 1998;43(1):17–37. <https://doi.org/10.1146/annurev.ento.43.1.17>.
- Moran NA. Symbiosis as an adaptive process and source of phenotypic complexity. *Proc Natl Acad Sci*. 2007;104(Suppl 1):8627–33. <https://doi.org/10.1073/pnas.0611659104>.
- 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? *Zool*. 2018;127:1–19. <https://doi.org/10.1016/j.zool.2018.02.004>.
- Salem H, Kirsch R, Pauchet Y, Berasategui A, Fukumori K, Moriyama M, et al. Symbiont digestive range reflects host plantbreadth in herbivorous beetles. *Current Biology*. 2020;30:2875–2886.e4.
- Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ. Adaptation via symbiosis: recent spread of a *Drosophila* defensive symbiont. *Science*. 2010;329(5988):212–5. <https://doi.org/10.1126/science.1188235>.
- 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(6026):254–6. <https://doi.org/10.1126/science.1199410>.
- Bordenstein SR, Theis KR. Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLoS Biol*. 2015;13(8):e1002226. <https://doi.org/10.1371/journal.pbio.1002226>.
- Moran NA, Yun Y. Experimental replacement of an obligate insect symbiont. *Proc Natl Acad Sci*. 2015;112(7):2093–6. <https://doi.org/10.1073/pnas.1420037112>.
- 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;8:2518. <https://doi.org/10.3389/fmicb.2017.02518>.
- 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(1):13. <https://doi.org/10.1186/s42523-019-0014-2>.
- 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(1870):20172599. <https://doi.org/10.1098/rspb.2017.2599>.
- 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(40):20025–32. <https://doi.org/10.1073/pnas.1907787116>.
- Walters AW, Hughes RC, Call TB, Walker CJ, Wilcox H, Petersen SC, et al. The microbiota influences the *Drosophila melanogaster* life history strategy. *Mol Ecol*. 2020;29(3):639–53. <https://doi.org/10.1111/mec.15344>.
- 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. *N Engl J Med*. 2011;365:1771–80. <https://doi.org/10.1056/NEJMoa1106483>.
- Sachs JL, Skophammer RG, Bansal N, Stajich JE. Evolutionary origins and diversification of proteobacterial mutualists. *Proc Royal Soc London B: Biol Sci*. 2014;281:20132146.
- 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):2478. <https://doi.org/10.1038/s41467-018-04955-6>.
- Melnyk RA, Hossain SS, Haney CH. Convergent gain and loss of genomic islands drive lifestyle changes in plant-associated *Pseudomonas*. *ISME J*. 2019;13(6):1575–88. <https://doi.org/10.1038/s41396-019-0372-5>.
- 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(29):11923–7. <https://doi.org/10.1073/pnas.1305559110>.
- 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(1):238. <https://doi.org/10.1186/s13059-016-1108-8>.
- Collins C, Didelot X. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. *PLoS Comput Biol*. 2018;14(2):e1005958. <https://doi.org/10.1371/journal.pcbi.1005958>.

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(11):1951–2. <https://doi.org/10.1093/bioinformatics/bty001>.
23. White KM, Matthews MK, Hughes RC, Sommer AJ, Griffiths 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;8:1119–27.
24. Chaston JM, Newell PD, Douglas AE. Metagenome-wide association of microbial determinants of host phenotype in *Drosophila melanogaster*. *mBio*. 2014;5(5):e01631–14. <https://doi.org/10.1128/mBio.01631-14>.
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:e00662–18.
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;86:200305–20.
27. Brummel T, Ching A, Seroude L, Simon AF, Benzer S. *Drosophila* lifespan enhancement by exogenous bacteria. *Proc Natl Acad Sci U S A*. 2004;101(35):12974–9. <https://doi.org/10.1073/pnas.0405207101>.
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(6056):670–4. <https://doi.org/10.1126/science.1212782>.
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 Metab*. 2011;14(3):403–14. <https://doi.org/10.1016/j.cmet.2011.07.012>.
30. Téfrit MA, Leulier F. *Lactobacillus plantarum* favors the early emergence of fit and fertile adult *Drosophila* upon chronic undernutrition. *J Experimental Biol*. 2017;jeb.151522.
31. Fast D, Duggal A, Foley E. Monoassociation with *Lactobacillus plantarum* disrupts intestinal homeostasis in adult *Drosophila melanogaster*. *mBio*. 2018;9:e01114–8.
32. Gould AL, Zhang V, Lamberti L, Jones EW, Obadia B, Korasidis N, et al. Microbiome interactions shape host fitness. *PNAS*. 2018;115(51):E11951–60. <https://doi.org/10.1073/pnas.1809349115>.
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(1):975. <https://doi.org/10.1038/s41467-018-03070-w>.
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(1):6312. <https://doi.org/10.1038/ncomms7312>.
35. Yamada R, Deshpande SA, Bruce KD, Mak EM, Ja WW. Microbes promote amino acid harvest to rescue undernutrition in *Drosophila*. *Cell Rep*. 2015;10(6):865–72. <https://doi.org/10.1016/j.celrep.2015.01.018>.
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(2):671–9. <https://doi.org/10.1128/AEM.03301-15>.
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):4236. <https://doi.org/10.1038/s41467-020-18049-9>.
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 Biol*. 2020;18(3):e3000681. <https://doi.org/10.1371/journal.pbio.3000681>.
40. Ankrah NYD, Douglas AE. Nutrient factories: metabolic function of beneficial microorganisms associated with insects. *Environ Microbiol*. 2018;20(6):2002–11. <https://doi.org/10.1111/1462-2920.14097>.
41. Moran NA, Ochman H, Hammer TJ. Evolutionary and ecological consequences of gut microbial communities. *Annu Rev Ecol Syst*. 2019;50(1):451–75. <https://doi.org/10.1146/annurev-ecolsys-110617-062453>.
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. *Nat Commun*. 2020;11(1):2471. <https://doi.org/10.1038/s41467-020-16274-w>.
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;4:247–59.
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. *Nat Microbiol*. 2017;2(12):1635–47. <https://doi.org/10.1038/s41564-017-0038-x>.
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(2):788–96. <https://doi.org/10.1128/AEM.02742-13>.
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;5:576.
49. Winans NJ, Walter A, Chouaib B, Chaston JM, Douglas AE, Newell PD. A genomic investigation of ecological differentiation between free-living and *Drosophila*-associated bacteria. *Mol Ecol*. 2017;26(17):4536–50. <https://doi.org/10.1111/mec.14232>.
50. Pais IS, Valente RS, Sporniak M, Teixeira L. *Drosophila melanogaster* establishes a species-specific mutualistic interaction with stable gut-colonizing bacteria. *PLoS Biol*. 2018;16(7):e2005710. <https://doi.org/10.1371/journal.pbio.2005710>.
51. Lee J, Han G, Kim JW, Jeon CO, Hyun S. Taxon-specific effects of *Lactobacillus* on *Drosophila* host development. *Microb Ecol*. 2020;79(1):241–51. <https://doi.org/10.1007/s00248-019-01404-9>.
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(6275):854–7. <https://doi.org/10.1126/science.aad8588>.
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(8):e70749. <https://doi.org/10.1371/journal.pone.0070749>.
55. Adair KL, Wilson M, Bost A, Douglas AE. Microbial community assembly in wild populations of the fruit fly *Drosophila melanogaster*. *ISME J*. 2018;12(4):959–72. <https://doi.org/10.1038/s41396-017-0020-x>.
56. Wang Y, Staubach F. Individual variation of natural *D. melanogaster*-associated bacterial communities. *FEMS Microbiol Lett*. 2018;365:fny017.
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. *Environ Microbiol Rep*. 2020;12(2):220–8. <https://doi.org/10.1111/1758-2229.12826>.
58. 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(5864):777–82. <https://doi.org/10.1126/science.1149357>.
59. Galac MR, Lazzaro BP. Comparative pathology of bacteria in the genus *Providencia* to a natural host, *Drosophila melanogaster*. *Microbes and Infection*. 2011;13:673–83, 7. <https://doi.org/10.1016/j.jmici.2011.02.005>.
60. Ding W, Baumdicker F, Neher RA panX: pan-genome analysis and exploration. *Nucleic Acids Res*. 2018;46:e5.
61. Karunakaran R, Ebert K, Harvey S, Leonard ME, Ramachandran V, Poole PS. Thiamine is synthesized by a salvage pathway in *Rhizobium leguminosarum* bv. viciae strain 3841. *J Bacteriol*. 2006;188(18):6661–8. <https://doi.org/10.1128/JB.00641-06>.
62. 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(3):539–42. <https://doi.org/10.1093/sysbio/sys029>.
63. Gori A, Harrison OB, Mlia E, Nishihara Y, Chan JM, Msefula J, et al. Pan-GWAS of *Streptococcus agalactiae* highlights lineage-specific genes associated with virulence and niche adaptation. *mBio*. 2020;11:e00728-0.
64. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species

- boundaries. *Nat Commun.* 2018;9(1):5114. <https://doi.org/10.1038/s41467-018-07641-9>.
65. Brouwer MSM, Roberts AP, Hussain H, Williams RJ, Allan E, Mullany P. Horizontal gene transfer converts non-toxicogenic *Clostridium difficile* strains into toxin producers. *Nat Commun.* 2013;4(1):2601. <https://doi.org/10.1038/ncomms3601>.
  66. Van Rossum T, Ferretti P, Maistrenko OM, Bork P. Diversity within species: interpreting strains in microbiomes. *Nat Rev Microbiol.* 2020;18(9):491–506. <https://doi.org/10.1038/s41579-020-0368-1>.
  67. Obadia B, Güvener ZT, Zhang V, Ceja-Navarro JA, Brodie EL, Ja WW, et al. Probabilistic invasion underlies natural gutmicrobiome stability. *Curr Biol.* 2017;27:1999–2006.e8.
  68. Mansourian S, Enjin A, Jirle EV, Ramesh V, Rehmann G, Becher PG, et al. Wild African *Drosophila melanogaster* areseasonal specialists on marula fruit. *Curr Biol.* 2018;28:3960–3968.e3.
  69. Douglas AE. The B vitamin nutrition of insects: the contributions of diet, microbiome and horizontally acquired genes. *Curr Opin Insect Sci.* 2017;23:65–9. <https://doi.org/10.1016/j.cois.2017.07.012>.
  70. Wang Y, Eum JH, Harrison RE, Valzania L, Yang X, Johnson JA, et al. Riboflavin instability is a key factor underlying the requirement of a gut microbiota for mosquito development. *PNAS.* 2021;118(15):e2101080118. <https://doi.org/10.1073/pnas.2101080118>.
  71. Pinto-Carbó M, Sieber S, Dessein S, Wicker T, Verstraete B, Gademann K, et al. Evidence of horizontal gene transfer between obligate leaf nodule symbionts. *ISME J.* 2016;10(9):2092–105. <https://doi.org/10.1038/ismej.2016.27>.
  72. Lopanik NB. Chemical defensive symbioses in the marine environment. *Funct Ecol.* 2014;28(2):328–40. <https://doi.org/10.1111/1365-2435.12160>.
  73. Ju J-F, Bing X-L, Zhao D-S, Guo Y, Xi Z, Hoffmann AA, et al. *Wolbachia* supplement biotin and riboflavin to enhance reproduction in planthoppers. *ISME J.* 2020;14(3):676–87. <https://doi.org/10.1038/s41396-019-0559-9>.
  74. Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T. Evolutionary origin of insect–*Wolbachia* nutritional mutualism. *PNAS.* 2014;111(28):10257–62. <https://doi.org/10.1073/pnas.1409284111>.
  75. Driscoll TP, Verhoeve VI, Brockway C, Shrewsbury DL, Plumer M, Sevdalis SE, et al. Evolution of *Wolbachia* mutualism and reproductive parasitism: insight from two novel strains that co-infect cat fleas. *PeerJ.* 2020;8:e10646. <https://doi.org/10.7717/peerj.10646>.
  76. 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(3):831–8. <https://doi.org/10.1093/gbe/evv016>.
  77. Gillespie JJ, Joarder V, Williams KP, Driscoll T, Hostetler JB, Nordberg E, et al. A *Rickettsia* genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. *J Bacteriol.* 2012;194(2):376–94. <https://doi.org/10.1128/JB.06244-11>.
  78. 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(7376):241–4. <https://doi.org/10.1038/nature10571>.
  79. 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(7612):435–9. <https://doi.org/10.1038/nature18927>.
  80. Costlow ZA, Degnan PH. Thiamine acquisition strategies impact metabolism and competition in the gut microbe *bacteroidesthetaitaomicron*. *mSystems.* 2017;2:e00116-7.
  81. Ross AC, Caballero BH, Cousins RJ, Tucker KL, Ziegler TR. Modern nutrition in health and disease: Eleventh edition. Wolters Kluwer Health Adis (ESP); 2012.
  82. Morris JJ. Black Queen evolution: the role of leakiness in structuring microbial communities. *Trends Genet.* 2015;31(8):475–82. <https://doi.org/10.1016/j.tig.2015.05.004>.
  83. Bull JJ, Rice WR. Distinguishing mechanisms for the evolution of co-operation. *J Theor Biol.* 1991;149(1):63–74. [https://doi.org/10.1016/S0022-5193\(05\)80072-4](https://doi.org/10.1016/S0022-5193(05)80072-4).
  84. Foster KR, Schluter J, Coyte KZ, Rakoff-Nahoum S. The evolution of the host microbiome as an ecosystem on a leash. *Nature.* 2017;548(7665):43–51. <https://doi.org/10.1038/nature23292>.
  85. 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:e02430–19.
  86. Kiers ET, Rousseau RA, West SA, Denison RF. Host sanctions and the legume–rhizobium mutualism. *Nature.* 2003;425(6953):78–81. <https://doi.org/10.1038/nature01931>.
  87. 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(2):183–94. <https://doi.org/10.1016/j.chom.2013.07.006>.
  88. 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. *Environ Microbiol.* 2019;21(9):3269–83. <https://doi.org/10.1111/1462-2920.14392>.
  89. 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 *Acemispion strigosus*, but host sanctions are uniform. *New Phytol.* 2019;221(1):446–58. <https://doi.org/10.1111/nph.15378>.
  90. 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(1):217–29. <https://doi.org/10.1038/s41396-019-0532-7>.
  91. Voolstra CR, Ziegler M. Adapting with microbial help: microbiome flexibility facilitates rapid responses to environmental change. *BioEssays.* 2020;42(7):2000004. <https://doi.org/10.1002/bies.202000004>.
  92. Bendix GH, Heberlein DG, Ptak LR, Clifcorn LE. Factors influencing the stability of thiamine during heat sterilization. *J Food Sci.* 1951;16(1-6):494–503. <https://doi.org/10.1111/j.1365-2621.1951.tb17411.x>.
  93. Kadakal Ç, Duman T, Ekinci R. Thermal degradation kinetics of ascorbic acid, thiamine and riboflavin in rosehip (*Rosa canina* L) nectar. *Food Sci Technol.* 2017;38(4):667–73. <https://doi.org/10.1590/1678-457x.11417>.
  94. Hubálek Z. Protectants used in the cryopreservation of microorganisms. *Cryobiology.* 2003;46(3):205–29. [https://doi.org/10.1016/S0011-2240\(03\)00046-4](https://doi.org/10.1016/S0011-2240(03)00046-4).
  95. Geissmann Q. OpenCFU, a new free and open-source software to count cell colonies and other circular objects. *PLoS One.* 2013;8(2):e54072. <https://doi.org/10.1371/journal.pone.0054072>.
  96. 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(17):5112–20. <https://doi.org/10.1128/AEM.01043-13>.
  97. 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. *Appl Environ Microbiol.* 2009;75(23):7537–41. <https://doi.org/10.1128/AEM.01541-09>.
  98. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics.* 2015;31(4):587–9. <https://doi.org/10.1093/bioinformatics/btu661>.
  99. 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;25(7):1043–55. <https://doi.org/10.1101/gr.186072.114>.
  100. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29(8):1072–5. <https://doi.org/10.1093/bioinformatics/btt086>.
  101. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30(14):2068–9. <https://doi.org/10.1093/bioinformatics/btu153>.
  102. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics.* 2020;36:1925–7.
  103. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol.* 2018;36(7):566–9. <https://doi.org/10.1038/nbt.4163>.
  104. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One.* 2010;5(3):e9490. <https://doi.org/10.1371/journal.pone.0009490>.
  105. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3. <https://doi.org/10.1093/bioinformatics/btu033>.
  106. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 2004;5(1):113. <https://doi.org/10.1186/1471-2105-5-113>.
  107. Supplementary Datasets. 11-Aug-2020. NCBI BioProject accession: PRJNA656529. [<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA656529>].

108. Sheng B, Ni J, Gao C, Ma C, Xu P. Draft genome sequence of the *Gluconobacter oxydans* strain DSM 2003, an important biocatalyst for industrial use. *Genome Announc.* 2014;2:e00417–4.
109. Prust C, Hoffmeister M, Liesegang H, Wiewer A, Fricke WF, Ehrenreich A, et al. Complete genome sequence of the acetic acid bacterium *Gluconobacter oxydans*. *Nat Biotechnol.* 2005;23(2):195–200. <https://doi.org/10.1038/nbt1062>.
110. Kostner D, Luchterhand B, Junker A, Volland S, Daniel R, Büchs J, et al. The consequence of an additional NADH dehydrogenase paralog on the growth of *Gluconobacter oxydans* DSM3504. *Appl Microbiol Biotechnol.* 2015;99(1):375–86. <https://doi.org/10.1007/s00253-014-6069-9>.
111. Juneja P, Lazzaro BP. *Providencia sneebia* sp. nov. and *Providencia burhodogranaria* sp. nov., isolated from wild *Drosophila melanogaster*. *Int J Syst Evol Microbiol.* 2009;59 Pt 5:1108–11.
112. Kim E-K, Kim S-H, Nam H-J, Choi MK, Lee K-A, Choi S-H, et al. Draft genome sequence of *Gluconobacter morbifer* G707T, a pathogenic gut bacterium isolated from *Drosophila melanogaster* intestine. *J Bacteriol.* 2012;194(5): 1245. <https://doi.org/10.1128/JB.06670-11>.
113. Kim E-K, Kim S-H, Nam H-J, Choi MK, Lee K-A, Choi S-H, et al. Draft genome sequence of *Commensalibacter intestini* A911T, a symbiotic bacterium isolated from *Drosophila melanogaster* intestine. *J Bacteriol.* 2012;194(5): 1246. <https://doi.org/10.1128/JB.06669-11>.
114. Supplementary Datasets. 13-Dec-2013. NCBI BioProject accession: PRJNA228961. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA228961>
115. Supplementary Datasets. 24-Jan-2005. NCBI BioProject accession: PRJNA13325. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA13325>
116. Supplementary Datasets. 3-Mar-2014. NCBI BioProject accession: PRJNA188081. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA188081>
117. Supplementary Datasets. 13-Nov-2012. NCBI BioProject accession: PRJNA82569. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA82569>
118. Supplementary Datasets. 4-Nov-2011. NCBI BioProject accession: PRJNA73361. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA73361>.
119. Supplementary Datasets. 21-Oct-2011. NCBI BioProject accession: PRJNA73359. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA73359>.
120. Malimas T, Chaipitakchonlatarn W, Thi Lan Vu H, Yukphan P, Muramatsu Y, Tanasupawat S, et al. *Swingsia samuiensis* gen. nov., sp. nov., an osmotolerant acetic acid bacterium in the  $\alpha$ -Proteobacteria. *J Gen Appl Microbiol.* 2013;59(5):375–84. <https://doi.org/10.2323/jgam.59.375>.
121. Yukphan P, Malimas T, Muramatsu Y, Potacharoen W, Tanasupawat S, Nakagawa Y, et al. *Neokomagataea* gen. nov., with descriptions of *Neokomagataea thailandica* sp. nov. and *Neokomagataea tanensis* sp. nov., osmotolerant acetic acid bacteria of the  $\alpha$ -Proteobacteria. *Biosci Biotechnol Biochem.* 2011;75(3):419–26. <https://doi.org/10.1271/bbb.100344>.
122. Supplementary Datasets. 2-Jul-2019. NCBI BioProject accession: PRJNA528164. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA528164>.
123. Supplementary Datasets. 2-Jul-2019. NCBI BioProject accession: PRJNA492196. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA492196>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

