Author for correspondence:
Christian Schlötterer
e-mail: christian.schloetterer@vetmeduni.ac.at

THE ROYAL SOCIETY
PUBLISHING

Long-term gut microbiome dynamics in *Drosophila melanogaster* reveal environment-specific associations between bacterial taxa at the family level

Rupert Mazzucco and Christian Schlötterer

Institut für Populationsgenetik, Veterinärmedizinische Universität Wien, Veterinärplatz 1, Wien 1210, Austria

RM, 0000-0002-7608-0196

The influence of the microbiome on its host is well-documented, but the interplay of its members is not yet well-understood. Even for simple microbiomes, the interaction among members of the microbiome is difficult to study. Longitudinal studies provide a promising approach to studying such interactions through the temporal covariation of different taxonomic units. By contrast to most longitudinal studies, which span only a single host generation, we here present a *post hoc* analysis of a whole-genome dataset of 81 samples that follows microbiome composition for up to 180 host generations, which cover nearly 10 years. The microbiome diversity remained rather stable in replicated *Drosophila melanogaster* populations exposed to two different temperature regimes. The composition changed, however, systematically across replicates of the two temperature regimes. Significant associations between families, mostly specific to one temperature regime, indicate functional interdependence of different microbiome components. These associations also involve moderately abundant families, which emphasizes their functional importance, and highlights the importance of looking beyond the common constituents of the *Drosophila* microbiome.

1. Introduction

The interaction of the microbiome with its host is highly topical and has been investigated in many species [1–4]. In particular in the context of human health [5], a wide variety of host conditions has been linked to the microbiome [6–13]. The causal relations between microbiome composition and its effect on the host remain poorly understood [14,15].

The composition of the microbiome is highly dynamic and affected by the host genotype as well as by environmental factors [16–21]. Compositional changes of the microbiome, which are triggered by environmental challenges, provide the potential to indirectly modulate the response of the host to an altered environment [22]. Nevertheless, despite a well-documented turnover of microbiome composition and complexity, our understanding of how microbial communities establish and persist in interaction with their host, and the functional interaction among the members of the microbiome community, is still in its infancy [23,24].

Two approaches to study the functional interactions have been proposed: experimental manipulation of the microbiome composition combined with phenotypic analysis of the host [25–27], and covariance in abundance of specific taxonomic units [28]. The underlying idea is that species with functional independence will vary in their abundance randomly with respect to environmental conditions or host genotypes [29]. Covariation, in contrast, suggests that they are either affected in the same way (positive correlation) or they complement each other (negative correlation). The correlation analysis is, nevertheless,

challenged by spurious associations [30,31], in particular when only one or a few host generations are studied.

By contrast to mammalian microbiomes, the gut microbiome of the fruit fly *Drosophila melanogaster* is relatively simple [32,33]. This provides the opportunity to evaluate the impact of the presence of different combinations of the microbiome on the *Drosophila* host [34,35]. Nevertheless, despite the relatively low complexity of the *Drosophila* microbiome, testing the number of possible combinations is challenging; hence, most studies focus on the most abundant species, (e.g. [36,37]), which in laboratory strains typically belong to the genera *Acetobacter* and *Lactobacillus* [38].

For an unbiased exploration of functional associations of different microbiome components, however, the focus on a pre-selected set of taxa is not well-suited and a covariation approach holds more promise: surveying a large number of individuals or following the abundance dynamics in a longitudinal study, can uncover functionally associated taxa. A challenge for the interpretation of longitudinal covariation patterns is that the starting condition (species composition and abundance pattern) affects subsequent time points [39]. Hence, unless the stochastic fluctuations are large, it can be difficult to disentangle random associations from functional ones, in particular when longitudinal studies are only conducted within a single host generation.

Drosophila provides multiple advantages to study functional associations between components of the microbiome. Firstly, the microbiome is not very complex, making the analyses more powerful; secondly, the *Drosophila* microbiome is highly dynamic with large changes in microbiome composition [40], which will break random associations and retain only functionally relevant ones; thirdly, longitudinal studies in mammals often only look at changes within a single host generation, (e.g. [41]). The comparatively short generation time of *Drosophila* enables studies of microbiome time series [39,42] that cover many generations, which provides a substantial increase in power to detect functionally important associations.

Several studies analysed the interplay of thermal adaptation of the *Drosophila* host and its microbiome [43–47], a question of interest in the context of climate change and the possibility of associated range shifts [48,49]. The results remained nevertheless inconclusive, which may be attributed to the small number of host generations covered in those studies. To obtain more robust results, we used an extended longitudinal study of the *D. melanogaster* microbiome during adaptation to novel hot and cold temperature regimes over nearly 10 years, covering 180 generations in hot and 100 generations in cold environments. We characterize the microbiome diversity and changes in composition over time, and analyse associations between taxa and global trends common across replicates, highlighting the significant differences between the two temperature regimes.

2. Material and methods

The Pool-Seq dataset (PRJEB37761) used in this study comes from a long-term evolve-and-resequence experiment of 180 generations of *D. melanogaster* kept in hot conditions (temperature fluctuating between 18°C and 28°C to mimic day and night), and 100 generations kept in cold conditions (fluctuating between 10°C and 20°C), with five replicates each, comprising 81 samples in total (electronic supplementary material, table S1).

(a) Culture conditions

In each temperature regime replicate populations of approximately 1000 flies in five bottles (approx. 200×5) were maintained in parallel using the same conditions and sample handling (including sample freezing). The developmental rates differ between temperature regimes, thus the samples in the two different temperature regimes were handled independently. All flies were maintained on a standard fly food medium (agar–agar, sugar beets syrup, malt syrup, yeast, corn flour, soy flour, in weight proportions of about $1:3:3:3:7:1$), which remained the same throughout the entire experiment. Eclosed flies were transferred to a fresh medium for 4 (hot) or 8 (cold) days and upon transfer to the bottles giving rise to the next generations they were supplemented with additional flies, which emerged during this period. After 2 (hot) and 4 (cold) days of egg laying the flies were transferred to another set of fresh bottles for another round of egg laying. After the last round of egg laying the flies were snap frozen in liquid nitrogen and stored at −80°C until DNA extraction.

(b) DNA extraction and sequencing

DNA was extracted from approximately 500 female flies (except for a single sample that was made from a mixture of males and females; see electronic supplementary material, table S1) using the same high salt extraction procedure for all samples considered here [50]. All reagents used for DNA extraction were prepared in large batches which were used to extract DNA from multiple time points. The exact assignment of batches to specific DNA extractions is not possible. Within a single time point the age of the flies varied between 4 and 8 days for the hot environment and 9–16 days for the cold environment. Pools were sequenced at various time points, using a range of library kits, insert sizes, and read lengths, reflecting the development of Illumina sequencing over a decade ([51]; electronic supplementary material, table S1 and figure S6). Reads were further processed as described below.

(c) Removing barcoded spike-ins

Six lanes (r01F23.93-96, r05F15.58, r10F15.90) contained some additional spike-in reads from *Aviadenovirus A* and/or *Cochliomyia hominivorax* with separate barcodes. The five-base barcodes PEMx1–4 are found at the 5′ end of each read of a pair. We removed all read pairs with the barcode on both reads, allowing up to one mismatch each, with homebrew R code [52] using functions from Bioconductor::ShortRead v. 1.42.0 [53].

(d) Quality control and trimming

Quality control with FastQC v. 0.11.8 [54] revealed consistent anomalies with base frequencies over the first few bases on the 5′ end, as well as occasional minor adapter contamination. We, therefore, clipped the first five bases on the 5′ end and removed adapter fragments with trimmomatic v. 0.39 [55] (parameters: ILLUMINACLIP:{adapters}:2:30:10:2 HEADCROP:5 MINLEN:50 AVGQUAL:28, with {adapters} chosen as TruSeq2.fa or TruSeq-3.fa as applicable to the respective library), also discarding reads below a minimum length of 50 and an average quality below 28.

(e) Contaminant and duplicate removal

Contaminants were removed by mapping the trimmed read pairs with bowtie2 v. 2.3.5.1 [56] against the *D. melanogaster* host and a set of other genomes that were sequenced together with the target libraries, but may have not been fully removed: we used a combined reference of *D. melanogaster* (GCF_000001215.4), *w*Mel (NC_002978.6), *Homo sapiens* (GCF_000001405.39), *Mus musculus* (GCF_000001635.26), *Gallus gallus* (GCF_000002315.4),

Table 1. The seven most abundant families across all samples (read counts larger than 1% of the total).

family	read pairs	fraction	genera
Acetobacteraceae	12 585 915	0.68	93% *Acetobacter*, 4% *Komagataeibacter*, 2% *Gluconacetobacter*, 1% *Gluconobacter*
Lactobacillaceae	1 556 816	0.084	100% *Lactobacillus*
Streptomycetaceae	603 057	0.033	62% *Streptomyces*
Enterobacteriaceae	397 342	0.021	27% *Salmonella*, 17% *Escherichia*, 4% *Klebsiella*, 4% *Citrobacter*, 3% *Enterobacter*
Burkholderiaceae	275 562	0.015	89% *Ralstonia*, 6% *Cupriavidus*, 2.5% *Paraburkholderia*, 2.5% *Burkholderia*
Leuconostocaceae	263 390	0.014	100% *Leuconostoc*
Bradyrhizobiaceae	195 634	0.011	100% *Bradyrhizobium*

Aviadenovirus A (NC_001720.1), and Illumina PhiX (modified NC_001422.1 with 587:G > A, 833:G > A, 2731:A > G, 2793:C > T, 2811:C > T), and removed all concordantly mapped pairs (parameters: --end-to-end --maxins 1000 --no-mixed --no-discordant). Although *Wolbachia* is part of the microbiome in a wider sense as it lives inside the host cells, we did not include it in our analysis as is not part of the gut microbiome; its dynamics were studied separately [51,57]. As a final preparation step, duplicates were removed with fastuniq v. 1.1 ([58]; default parameters).

(f) Classification with Kraken2

Deduplicated read pairs were classified with kraken2 v. 2.0.8-beta [59] (parameters: --paired --confidence 0.04). Kraken2 has two parameters that influence classification. We did not modify the minimum base quality parameter from its default value of 0 as an exploratory analysis confirmed that our stringent quality filtering makes this parameter uninformative. Increasing the confidence parameter did, however, remove rare dubious classifications (i.e. hits to eukaryotes that could not have contaminated the libraries, e.g. golden eagles, salmon, turtles, wine, cucumbers etc. and which were based only on very few k-mers), while also decreasing the number of classified reads. After some exploration we chose $c = 0.04$ for this parameter, as with higher values the disadvantage of losing read pairs outweighed the gain in classification confidence.

We classified read pairs against a database built with the kraken2-build script (default parameters) from the NCBI NT database [60], with higher animals (Metazoa) and plants (Viridiplantae) removed and the dmel6_iso1MT (GCF_000001215.4) RefSeq [61] genome added.

(g) Postprocessing with bracken

Kraken2 assigns reads on the lowest taxonomic level for which a unique assignment is possible, which implies that different reads can be assigned at different taxonomic levels. We, therefore, used the companion program bracken v. 2.5.0 [62] for re-estimating read counts for a given taxonomic level using a Bayesian model, using the untrimmed read length of each sequencing lane.

We analysed the results on the family level after excluding any remnants of the known contaminants Drosophilidae, Adenoviridae and Anaplasmataceae. Most of the large families are dominated by a single genus, e.g. *Acetobacter*, *Lactobacillus*, *Leuconostoc*, *Ralstonia*, *Bradyrhizobium* (table 1).

(h) Estimating fractions

Because the samples contain only a very small part of the entire microbiome, observed read counts may be subject to substantial sampling error. Although this issue is often ignored [63], accounting for imperfect detection generally improves the analysis [64,65]. Therefore, given the $m = 806$ families classified across

all samples, we estimated their fractions x_i, with $i = 1, \ldots, m$, in each sample from their read counts n_i as $x_i = (n_i + 1)/(N + m)$ with $N = \sum n_i$, which assumes that the combined process of library preparation, sequencing, and read extraction is described by a multinomial sampling model. We note that our method is a more refined version of the commonly applied procedure of setting counts of 0 to 1 in observation tables: for an observed count $n = 0$ and $m \ll N$, the estimated count would indeed be $Nx = 1/(1 + m/N) \approx 1$.

(i) Correlation between families

A known problem of compositional data analysis is that component fractions may exhibit spurious negative correlations [66]. We thus used the normalized and log-transformed component fractions (additive log-ratio transform; [67], choosing as common denominator the fraction of the most abundant family Acetobacteraceae) across all samples with the R function Hmisc::rcorr v. 4.4-0 to obtain Spearman's rank correlation coefficients p as a measure of association between families.

(j) Piece-wise linear models

Figure 2 indicates the presence of abundance peaks at intermediate time points. Since such intermediate peaks cannot be captured by simple linear models, we identified significant trends by fitting segmented linear models using the R segmented package v. 1.1-0 [68,69] with an initial breakpoint located at generation 90 in the hot environment and generation 50 in the cold environment to allow for an intermediate peak. The algorithm moves and possibly drops this breakpoint as it identifies the best-fitting model, thus identifying both linear trends and intermediate peaks or dips in the data.

3. Results

We studied the long-term microbiome dynamics of *D. melanogaster* hosts exposed to novel hot (18/28°C) and cold environments (10/20°C). We first considered the changes in composition over time; then, focusing on the dynamics of the seven most abundant families of microbiota, we identified significant associations between several of them. Finally, we identified significant trends of community restructuring characteristic for the respective environmental conditions.

(a) Microbiome composition

Across all replicates and environments, the microbiome is dominated by seven bacterial families (table 1 and figure 1), most of which were also commonly observed in previous studies of the *Drosophila* microbiome. Like for other *Drosophila*
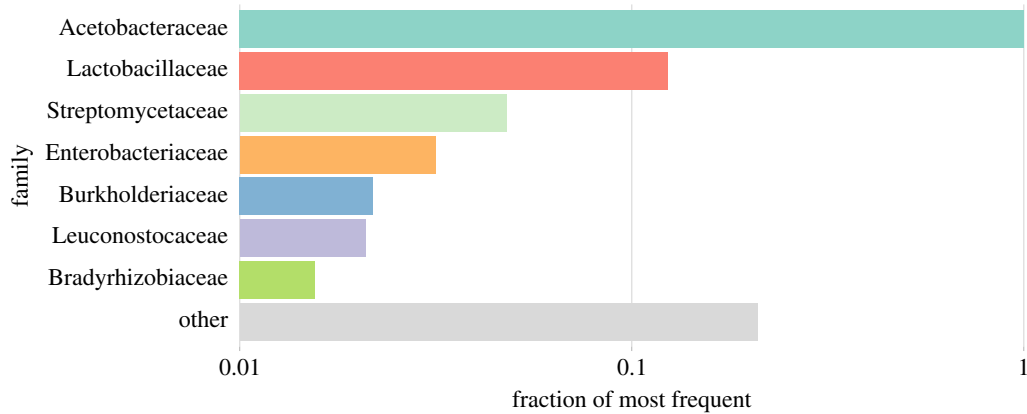
Figure 1. Microbial diversity is dominated by seven bacterial families. Analogous to rank-abundance plots, the relative frequencies of reads classified to the most common families are shown. All frequencies are scaled to the frequency of the Acetobacteraceae, the overall most common family (which thus has a value of 1). The colour coding used here applies to all figures throughout the entire manuscript. The seven most abundant families account for at least 74% of all read pairs in each of the 81 samples. (Online version in colour.)
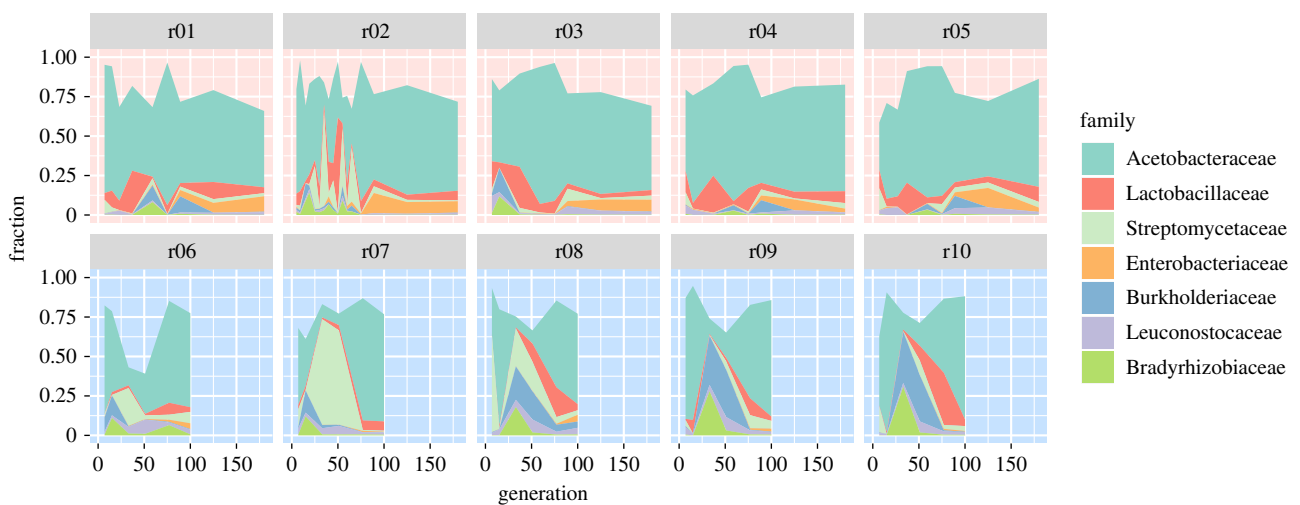

Figure 2. Replicate dynamics reveal characteristic features for each environment. Stacked coloured areas indicate the fraction of the respective family in the given generation in each replicate in the hot (upper row, red background) and cold (lower row, blue background) environment, linearly interpolating between sampled generations (9–17 per replicate in the hot environment, 51 in total; 6 per replicate in the cold environment, 30 in total). Only the seven most abundant families (table 1 and figure 1) are shown. (Online version in colour.)

cultured in the laboratory, Acetobacteraceae are the most abundant family [38,44,46], followed by Lactobacillaceae. The predominant families include further families frequently seen in the *Drosophila* microbiome [33], namely, Enterobacteriaceae and Leuconostocaceae. The high abundance of Enterobacteriaceae, a very common *Drosophila*-associated gut microbe family in the wild [43,70,71], is not surprising because the founder isofemale lines were kept only a small number of generations in the laboratory before the experiment was started. Streptomycetaceae are not typically reported as part of the *Drosophila* microbiome, but have been found in a recent study [72]. Finally, the remaining abundant families Bradyrhizobiaceae and Burkholderiaceae are not typical *Drosophila* gut bacteria, but are considered environmental bacteria [73,74].

(b) Microbiome dynamics

Given that we transferred flies with their associated microbiome to novel environments, we anticipated one of three different scenarios: (i) the microbiome quickly adapts to the new environment, as frequently seen for altered nutrition, (e.g. [75–77]), and then only experiences stochastic changes,

(ii) the microbiome changes continuously because adaptation to the new environment cannot be achieved by a simple change in relative frequency of the community members or (iii) the microbiome remains completely unaffected. None of these predictions fully applied to this experiment; instead, we observed complex dynamics that were also very different between environments (figure 2). Although these changes were not completely synchronized, their occurrence in multiple replicates strongly suggests that these dynamics are not just stochastic effects, but possibly reflect functional turnover of the microbiome during the experiment. In addition to these global patterns, we observed that some bacterial families changed their abundance in a continuous manner; for example, Lactobacillaceae continuously increase in frequency in the cold environment.

(c) Strong associations

While most studies describe microbiome composition in either natural flies or flies exposed to laboratory conditions, our time series data, covering up to 180 host generations, offer the unprecedented opportunity to test for associations
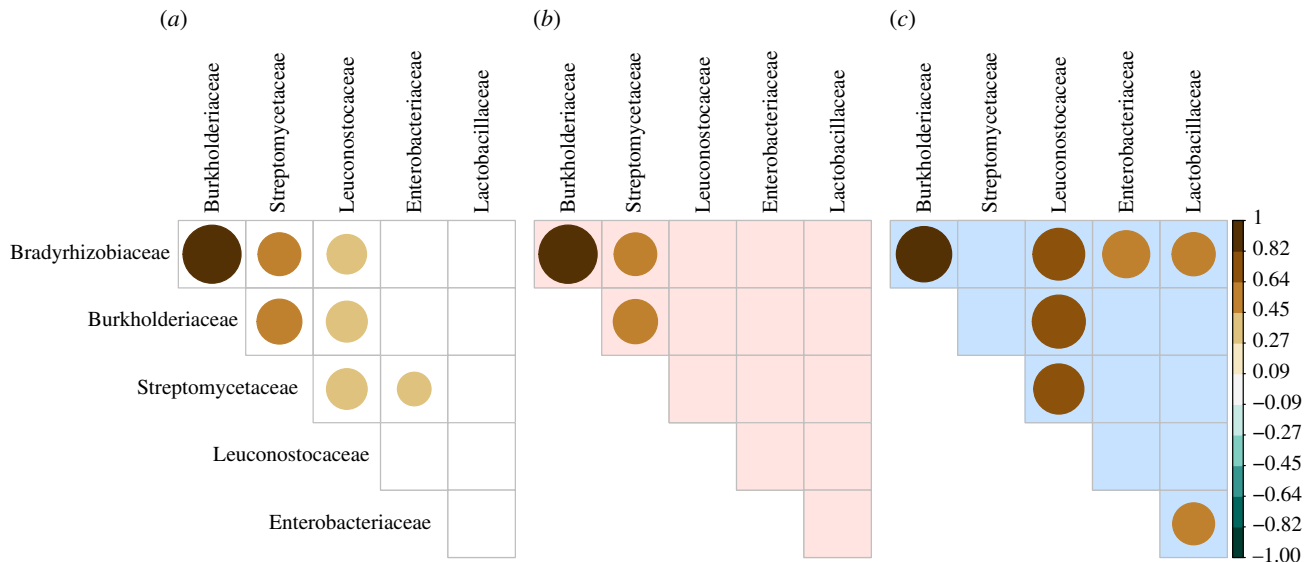
Figure 3. Positive associations among families in hot and cold environments. Filled circles indicate Spearman's correlation coefficient between ALR-transformed family fractions over all replicates in (*a*) all, (*b*) only hot and (*c*) only cold environments by size and fill colour. Only associations with *p*-values < 0.01 are shown. Except for the consistent strong association between Burkholderiaceae and Bradyrhizobiaceae, associations between families are different between environments: Streptomycetaceae are associated with this BB group in the hot environment, while in the cold environment Leuconostocaceae have a substantial association with the BB group and the Streptomycetaceae, and the Enterobacteriaceae and Lactobacillaceae are associated with the Bradyrhizobiaceae and each other. (Online version in colour.)

between different components of the microbiome. Given that the microbiome composition is highly stochastic, with substantial changes between generations [78], large abundance changes are expected. Time-series data provide the opportunity to distinguish between stochastic changes and covariation of abundance patterns between microorganisms. Statistically significant associations imply a functional association between covarying taxonomic groups.

As the interaction of members from different families may be informative about changes in the functional composition of the microbiome, we evaluated the correlated dynamics of the predominant families (figure 3). Bradyrhizobiaceae and Burkholderiaceae have strong positive associations across generations and environments (figure 3*a*–*c*). In the hot environment, they are also significantly associated with Streptomycetaceae, while in the cold environment Leuconostocaceae appear to play a pivotal role, being strongly associated with Bradyrhizobiaceae and Burkholderiaceae on the one hand, and Streptomycetaceae on the other, where also Enterobacteriaceae and Lactobacillaceae are associated with Bradyrhizobiaceae and each other. These highly significant associations provide very strong evidence for the interaction between different members of the microbiome, which only become apparent with the availability of long-term longitudinal data.

(d) Global trends
Because the visual inspection of the abundance plots (figure 2) suggests that the abundance of many families changes in a nonlinear manner, we analysed the family dynamics with segmented linear models to identify significant trends across replicates (figure 4). In cold culture conditions we find clear evidence a strong decrease of Acetobacteraceae and increase of Bradyrhizobiaceae and Burkholderiaceae around generation 30, which is followed by a strong increase of Leuconostocaceae around generation 50. These families, however, then decrease again in relative

abundance as Acetobacteraceae recover and Lactobacillaceae slowly increase (figure 4). The increase of Lactobacillaceae is restricted to the cold environment; in the hot environment, we see the opposite trend—if anything. Interestingly, this pattern of opposing trends in the two environments is seen even more clearly in the Leuconostocaceae, the other member of the group of lactic acid bacteria (electronic supplementary material, §A). Finally, the highly significant increase of Enterobacteriaceae over time occurs in both environments, and thus may be causally unrelated to the other interactions.

4. Discussion
Based on a comprehensive longitudinal dataset of 81 samples from replicated populations covering up to 180 host generations, we found strong evidence for a functional interdependence of various families constituting the *Drosophila* microbiome: (i) parallel changes in abundance patterns across replicates and (ii) significant covariation in abundance of several families.

(a) Parallel dynamics across replicates
The striking compositional similarity of replicate populations in this experiment can be recognized by the clustering of replicates as well as by the similar patterns of frequency changes. The parallel dynamics are most apparent from characteristic, highly pronounced frequency changes of a few families. One example for this is the decrease of Acetobacteraceae at generation 30 across all five replicates in the cold, followed by a recovery at later generations. Interestingly, almost the opposite trend was seen in the hot environment (figure 4).

One explanation for the high parallelism between replicates, which is independent of functional requirements, is related to the culturing method. As the experimental *Drosophila* populations were not maintained under sterile conditions, components of the microbiome may have been transferred accidentally across
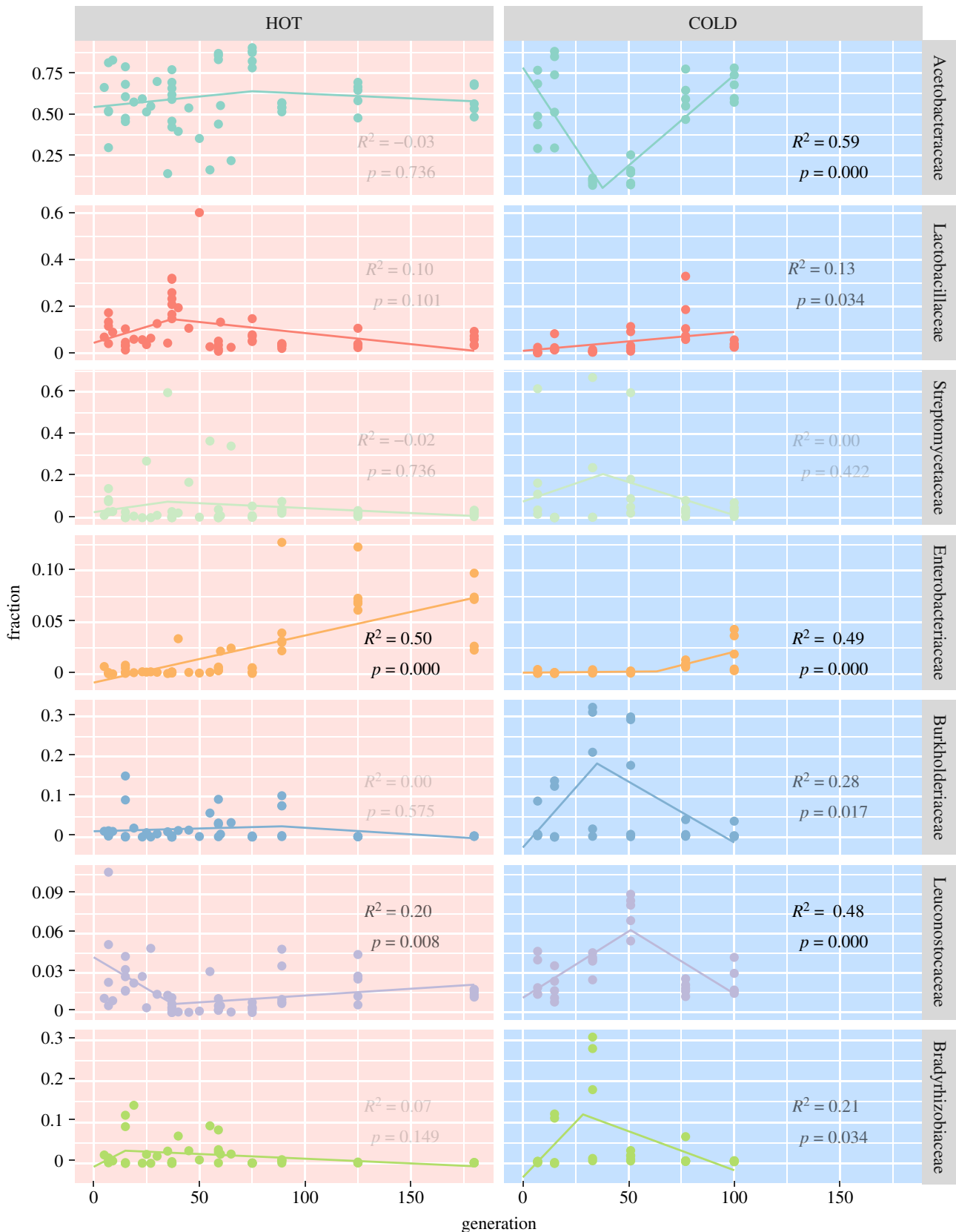
Figure 4. Significant trends differ between environments. Coloured points indicate family frequencies in all replicates in the hot (left column, red background) and cold (right column, blue background) environment (colours as in figure 3, families are also indicated on the margins). Lines indicate the best-fitting segmented linear model, with adjusted R^2- and p-values (Benjamin–Hochberg-corrected for multiple testing and rounded to three digits) given in the respective panel; their opaqueness further indicates significance. The following dynamics were statistically significant: in the hot environment (i) a steady increase of Enterobacteriaceae and (ii) an abundance drop of Leuconostocaceae at generation 37; in the cold environment (iii) a Leuconostocaceae peak around generation 50, which is preceded by (iv) a significant Actetobacteraceae dip at generation 35 and (v) a marginally significant peak in Burkholderiaceae and Bradyrhizobiaceae occurring around the same time; furthermore (vi) a significant increase in Enterobacteriaceae (after generation 60) and (vii) a weakly significant steady increase of Lactobacillaceae. (Online version in colour.)

replicates, thus synchronizing their dynamics. A similar mechanism was recently proposed for the high parallel selection responses of the *Drosophila* host across replicates [79].

Nevertheless, a closer inspection of these dynamics shows that this explanation is unlikely. The diversity peak in the cold culturing regime (figures 3 and 4) is a good example, occurring in a

similar manner across all replicates. Since rare taxa, rather than abundant families, increased in frequency, this is difficult to reconcile with microbiome transfer across replicates: a large increase in frequency by migration requires that many representatives of this family are being transferred. The highly parallel diversity increase makes an extrinsic contamination also unlikely, as all five replicates of the hot temperature regime are affected. Hence, we conclude that accidental contamination during the culturing is unlikely to explain the parallel microbiome dynamics seen in our experiment.

We acknowledge that the use of different sequencing platforms and library preparation protocols [51] could have contributed to the temporal heterogeneity seen in this study (electronic supplementary material, table S2). Importantly, some of the replicates were sequenced with different sequencing platforms (electronic supplementary material, §B and figure S2), and similar pronounced abundance changes can be seen for samples for which the same DNA was sequenced with two different platforms (electronic supplementary material, figure S3a). Furthermore, the temporal dynamics were apparent among time points sequenced using the same technology (electronic supplementary material, figure S3b). While our analyses confirm a certain effect of technology on detected abundance (electronic supplementary material, figures S4 and S5), as previously reported [80–82]; a detailed analysis of the dynamics in each replicate (electronic supplementary material, §C and figures S6 and S7), allows us to conclude that no prominent features of the family dynamics are artefacts of the sequencing technology.

Another contribution to temporal heterogeneity is the degree of gut colonization in the adult flies, which was shown to continue over several days [83]. Since flies were not collected at exactly the same day after eclosure, it is possible that some of the observed changes in abundance may—at least to some extent—reflect heterogeneity in gut colonization. Nevertheless, the functional interdependence of different families inferred from covariation of abundance patterns would be even further strengthened by heterogeneity in colonization status (see below).

(b) Significant covariation between families

Even stronger evidence for the functional non-independence of co-occurring families in the microbiome comes from the significant correlation in abundance between different families (figure 3). Since this signal of covariation comes from the joint analysis of multiple time points, it cannot be explained by contamination across replicates, which would have been limited to single generations. Furthermore, the large fluctuations in abundance patterns across time points highlights that the significant association between various families reflects strong functional links between them and rules out statistical artefacts. It was particularly striking that most of these associations do not involve components of the *Drosophila* microbiome that were previously used for functional testing, (e.g. [36,84–86]).

(c) The microbiome turnover: driver or passenger?

A naive expectation for the impact of rearing temperature on microbiome composition is that the microbiome either changes rapidly as it is exposed to a new environment, and persists in the new, temperature-optimized composition; or, that it changes gradually, reflecting the acquisition of new adaptive

mutations in the microbiome. Interestingly, our data do not fit either of these simple expectations. Rather, we noted a highly dynamic nature of the microbiome with trends that are largely consistent across replicates. The parallel dynamics in multiple replicates of the same temperature regime across multiple generations rules out stochastic changes, but raises the question about the underlying cause. While new functional mutations can occur [87], their appearance should be stochastic and not synchronized across replicates. Contingency on a certain genetic background for certain mutations to be successful has been described for citrate use in *Escherichia coli* [88] and a similar principle may apply to the microbiome composition as well. Nevertheless, this would only explain why some changes do not occur early in the experiment, but cannot synchronize the dynamics across replicates.

Here, we propose that much of the non-random patterns in the microbiome are driven by the *Drosophila* host. Several evolve-and-resequence studies exhibit a highly parallel selection response across replicates [89,90]. Given that the host genotype affects the microbiome composition as demonstrated by GWAS [17,18], genetic changes in the host genome, which are parallel across replicates, may also shape the microbiome composition. A direct link between temperature adaptation of the host and microbiome composition has been recently found [91] where microbiome composition, specifically the ratio of acetic acid bacteria to lactic acid bacteria, changes with latitude of *D. melanogaster* populations from eastern North America. While we do not find strong evidence for this antagonism of acetic and lactic acid bacteria between hot- and cold-evolved populations (electronic supplementary material, figure S1), we consider the influence of the host genotype the best explanation for the high parallelism and most likely for much of the temporal dynamics of the microbiome discovered in this study. Our previous analysis of the *Wolbachia* dynamics in this population [51] showed a turnover of different *Wolbachia* genotypes, but these were monotonic and thus unlikely to contribute to the abundance peaks observed in this study.

If the microbiome dynamics are driven by the *Drosophila* host, the continuous turnover of the microbiome in combination with the linear frequency change of Lactobacillaceae in the cold and the Enterobacteriaceae in the hot environment, may imply that the host has not yet fully adapted to the new environment (i.e. has not yet reached the new trait optimum). Nevertheless, it is important to consider that even after reaching the trait optimum the allele frequencies of contributing loci still experience considerable allele frequency changes [92,93], which may drive microbiome changes even when the phenotype of the host does not change anymore after the trait optimum has been reached. Alternatively, some of the dynamics, in particular those not parallel across replicates, may reflect the acquisition of new mutations by the host; (e.g. [94,95]). Single-strain characterization of evolved microbiomes will be helpful to characterize new mutations that occurred during the experiment. In the current set-up, the distinction between ancestral low frequency genotypes and new mutations may be challenging.

5. Conclusion

Strong functional associations have been shown before (reviewed in [96]), but these studies typically manipulated

only a small number of taxa to study the influence of the host. In our experiment, we followed the natural dynamics and inferred the patterns of covariation without any artificial manipulation, thus providing a more natural setting than experimentally combining different components of the microbiome. Our analyses indicate that the strongest associations between different components of the microbiome do not involve the most abundant families, but members of intermediate abundance, which are often ignored in functional studies of the *Drosophila* microbiome. Our results highlight the importance of expanding the functional characterization of the *Drosophila* microbiome beyond highly abundant families like Acetobacteraceae and Lactobacillaceae.

References

1. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007 The human microbiome project. *Nature* **449**, 804–810. (doi:10.1038/nature06244)

2. Isaacson R, Kim HB. 2012 The intestinal microbiome of the pig. *Anim. Health Res. Rev.* **13**, 100–109. (doi:10.1017/S1466252312000084)

3. Liu C et al. 2020 The Mouse Gut Microbial Biobank expands the coverage of cultured bacteria. *Nat. Commun.* **11**, 79. (doi:10.1038/s41467-019-13836-5)

4. O'Hara E, Neves ALA, Song Y, Guan LL. 2020 The role of the gut microbiome in cattle production and health: driver or passenger? *Annu. Rev. Anim. Biosci.* **8**, 199–220. (doi:10.1146/annurev-animal-021419-083952)

5. Cani PD. 2018 Human gut microbiome: hopes, threats and promises. *Gut* **67**, 1716–1725. (doi:10.1136/gutjnl-2018-316723)

6. Lloyd-Price J, Abu-Ali G, Huttenhower C. 2016 The healthy human microbiome. *Genome Med.* **8**, 51. (doi:10.1186/s13073-016-0307-y)

7. Dickerson F, Severance E, Yolken R. 2017 The microbiome, immunity, and schizophrenia and bipolar disorder. *Brain. Behav. Immun.* **62**, 46–52. (doi:10.1016/j.bbi.2016.12.010)

8. Maruvada P, Leone V, Kaplan LM, Chang EB. 2017 The human microbiome and obesity: moving beyond associations. *Cell Host Microbe* **22**, 589–599. (doi:10.1016/j.chom.2017.10.005)

9. Rajagopala SV, Vashee S, Oldfield LM, Suzuki Y, Venter JC, Telenti A, Nelson KE. 2017 The human microbiome and cancer. *Cancer Prev. Res.* **10**, 226–234. (doi:10.1158/1940-6207.CAPR-16-0249)

10. Barko PC, McMichael MA, Swanson KS, Williams DA. 2018 The gastrointestinal microbiome: a review. *J. Vet. Intern. Med.* **32**, 9–25. (doi:10.1111/jvim.14875)

11. Gopalakrishnan V, Helmink BA, Spencer CN, Reuben A, Wargo JA. 2018 The influence of the gut microbiome on cancer, immunity, and cancer immunotherapy. *Cancer Cell* **33**, 570–580. (doi:10.1016/j.ccell.2018.03.015)

12. Zheng P, Li Z, Zhou Z. 2018 Gut microbiome in type 1 diabetes: a comprehensive review. *Diabetes Metab. Res. Rev.* **34**, e3043. (doi:10.1002/dmrr.3043)

13. De Luca F, Shoenfeld Y. 2019 The microbiome in autoimmune diseases. *Clin. Exp. Immunol.* **195**, 74–85. (doi:10.1111/cei.13158)

14. Fischbach MA. 2018 Microbiome: focus on causation and mechanism. *Cell* **174**, 785–790. (doi:10.1016/j.cell.2018.07.038)

15. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. 2018 Current understanding of the human microbiome. *Nat. Med.* **24**, 392–400. (doi:10.1038/nm.4517)

16. Spor A, Koren O, Ley R. 2011 Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat. Rev. Microbiol.* **9**, 279–290. (doi:10.1038/nrmicro2540)

17. Chaston JM, Dobson AJ, Newell PD, Douglas AE. 2016 Host genetic control of the microbiota mediates the *Drosophila* nutritional phenotype. *Appl. Environ. Microbiol.* **82**, 671–679. (doi:10.1128/AEM.03301-15)

18. Early AM, Shanmugarajah N, Buchon N, Clark AG. 2017 *Drosophila* genotype influences commensal bacterial levels. *PLoS ONE* **12**, e0170332. (doi:10.1371/journal.pone.0170332)

19. Chen S, Wang J, Peng D, Li G, Chen J, Gu X. 2018 Exposure to heat-stress environment affects the physiology, circulation levels of cytokines, and microbiome in dairy cows. *Sci. Rep.* **8**, 14606. (doi:10.1038/s41598-018-32886-1)

20. Pathirana E, Fuhrmann M, Whittington R, Hick P. 2019 Influence of environment on the pathogenesis of Ostreid herpesvirus-1 (OsHV-1) infections in Pacific oysters (*Crassostrea gigas*) through differential microbiome responses. *Heliyon* **5**, e02101. (doi:10.1016/j.heliyon.2019.e02101)

21. Wang Y, Kapun M, Waidele L, Kuenzel S, Bergland AO, Staubach F. 2020 Common structuring principles of the *Drosophila melanogaster* microbiome on a continental scale and between host and substrate. *Environ. Microbiol. Rep.* **12**, 220–228. (doi:10.1111/1758-2229.12826)

22. Orland C, Emilson EJS, Basiliko N, Mykytczuk NCS, Gunn JM, Tanentzap AJ. 2019 Microbiome functioning depends on individual and interactive effects of the environment and community structure. *ISME J.* **13**, 1–11. (doi:10.1038/s41396-018-0230-x)

23. Cresci GA, Bawden E. 2015 Gut microbiome: what we do and don't know. *Nutr. Clin. Pract.* **30**, 734–746. (doi:10.1177/0884533615609899)

24. Pais IS, Valente RS, Sporniak M, Teixeira L. 2018 *Drosophila melanogaster* establishes a species-specific mutualistic interaction with stable gut-colonizing bacteria. *PLoS Biol.* **16**, e2005710. (doi:10.1371/journal.pbio.2005710)

25. Tremaroli V, Bäckhed F. 2012 Functional interactions between the gut microbiota and host metabolism. *Nature* **489**, 242–249. (doi:10.1038/nature11552)

26. Gilbert JA, Quinn RA, Debelius J, Xu ZZ, Morton J, Garg N, Jansson JK, Dorrestein PC, Knight R. 2016 Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature* **535**, 94–103. (doi:10.1038/nature18850)

27. Luca F, Kupfer SS, Knights D, Khoruts A, Blekhman R. 2018 Functional genomics of host–microbiome interactions in humans. *Trends Genet.* **34**, 30–40. (doi:10.1016/j.tig.2017.10.001)

28. McHardy IH et al. 2013 Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. *Microbiome* **1**, 17. (doi:10.1186/2049-2618-1-17)

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

30. Jackson DA. 1997 Ccompositional data in community ecology: the paradigm or peril of proportions? *Ecology* **78**, 929–940. (doi:10.1890/0012-9658(1997)078[0929:CDICET]2.0.CO;2)

31. Carr A, Diener C, Baliga NS, Gibbons SM. 2019 Use and abuse of correlation analyses in microbial ecology. *ISME J.* **13**, 2647–2655. (doi:10.1038/s41396-019-0459-z)

32. Broderick NA, Lemaitre B. 2012 Gut-associated microbes of *Drosophila melanogaster*. *Gut Microbes* **3**, 307–321. (doi:10.4161/gmic.19896)

33. Staubach F, Baines JF, Künzel S, Bik EM, Petrov DA. 2013 Host species and environmental effects on

bacterial communities associated with *Drosophila* in the laboratory and in the natural environment. *PLoS ONE* **8**, e70749. (doi:10.1371/journal.pone.0070749)

34. Douglas AE. 2018 The *Drosophila* model for microbiome research. *Lab. Anim.* **47**, 157–164. (doi:10.1038/s41684-018-0065-0)

35. Ludington WB, Ja WW. 2020 *Drosophila* as a model for the gut microbiome. *PLoS Pathog.* **16**, e1008398. (doi:10.1371/journal.ppat.1008398)

36. Rudman SM *et al.* 2019 Microbiome composition shapes rapid genomic adaptation of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **116**, 20 025–20 032. (doi:10.1073/pnas.1907787116)

37. Henriques SF *et al.* 2020 Metabolic cross-feeding in imbalanced diets allows gut microbes to improve reproduction and alter host behaviour. *Nat. Commun.* **11**, 4236. (doi:10.1038/s41467-020-18049-9)

38. Wong CNA, Ng P, Douglas AE. 2011 Low-diversity bacterial community in the gut of the fruitfly *Drosophila melanogaster*: bacterial community in *Drosophila melanogaster*. *Environ. Microbiol.* **13**, 1889–1900. (doi:10.1111/j.1462-2920.2011.02511.x)

39. Coenen AR, Hu SK, Luo E, Muratore D, Weitz JS. 2020 A primer for microbiome time-series analysis. *Front. Genet.* **11**, 310. (doi:10.3389/fgene.2020.00310)

40. Wong AC-N, Chaston JM, Douglas AE. 2013 The inconstant gut microbiota of *Drosophila* species revealed by 16S rRNA gene analysis. *ISME J.* **7**, 1922–1932. (doi:10.1038/ismej.2013.86)

41. Ramiro RS, Durão P, Bank C, Gordo I. 2020 Low mutational load and high mutation rate variation in gut commensal bacteria. *PLoS Biol.* **18**, e3000617. (doi:10.1371/journal.pbio.3000617)

42. Baksi KD, Kuntal BK, Mande SS. 2018 'TIME': a web application for obtaining insights into microbial ecology using longitudinal microbiome data. *Front. Microbiol.* **9**, 36. (doi:10.3389/fmicb.2018.00036)

43. Corby-Harris V, Pontaroli AC, Shimkets LJ, Bennetzen JL, Habel KE, Promislow DEL. 2007 Geographical distribution and diversity of bacteria associated with natural populations of *Drosophila melanogaster*. *Appl. Environ. Microbiol.* **73**, 3470–3479. (doi:10.1128/AEM.02120-06)

44. Moghadam NN, Thorshauge PM, Kristensen TN, de Jonge N, Bahrndorff S, Kjeldal H, Nielsen JL. 2017 Strong responses of *Drosophila melanogaster* microbiota to developmental temperature. *Fly* **12**, 1–12. (doi:10.1080/19336934.2017.1394558)

45. Henry Y, Colinet H. 2018 Microbiota disruption leads to reduced cold tolerance in *Drosophila* flies. *Sci. Nat.* **105**, 59. (doi:10.1007/s00114-018-1584-7)

46. Zare A, Johansson A, Karlsson E, Delhomme N, Stenberg P. 2018 The gut microbiome participates in transgenerational inheritance of low-temperature responses in *Drosophila melanogaster*. *FEBS Lett.* **592**, 4078–4086. (doi:10.1002/1873-3468.13278)

47. Karunakar P, Bhalla A, Sharma A. 2019 Transgenerational inheritance of cold temperature response in *Drosophila*. *FEBS Lett.* **593**, 594–600. (doi:10.1002/1873-3468.13343)

48. Chen I-C, Hill JK, Ohlemuller R, Roy DB, Thomas CD. 2011 Rapid range shifts of species associated with high levels of climate warming. *Science* **333**, 1024–1026. (doi:10.1126/science.1206432)

49. Semenza JC, Suk JE. 2018 Vector-borne diseases and climate change: a European perspective. *FEMS Microbiol. Lett.* **365**, fnx244. (doi:10.1093/femsle/fnx244)

50. Miller SA, Dykes DD, Polesky HF. 1988 A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215. (doi:10.1093/nar/16.3.1215)

51. Mazzucco R, Nolte V, Vijayan T, Schlötterer C. 2020 Long-term dynamics among *Wolbachia* strains during thermal adaptation of their *Drosophila melanogaster* hosts. *Front. Genet.* **11**, 482. (doi:10.3389/fgene.2020.00482)

52. R Core Team. 2019 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. See https://www.R-project.org/.

53. Morgan M, Anders S, Lawrence M, Aboyoun P, Pages H, Gentleman R. 2009 ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* **25**, 2607–2608. (doi:10.1093/bioinformatics/btp450)

54. Andrews S. 2010 *FastQC: a quality control tool for high throughput sequence data*. Babraham Bioinformatics. See https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

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

56. Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. (doi:10.1038/nmeth.1923)

57. Versace E, Nolte V, Pandey RV, Tobler R, Schlötterer C. 2014 Experimental evolution reveals habitat-specific fitness dynamics among *Wolbachia* clades in *Drosophila melanogaster*. *Mol. Ecol.* **23**, 802–814. (doi:10.1111/mec.12643)

58. Xu H, Luo X, Qian J, Pang X, Song J, Qian G, Chen J, Chen S. 2012 FastUniq: a fast de novo duplicates removal tool for paired short reads. *PLoS ONE* **7**, e52249. (doi:10.1371/journal.pone.0052249)

59. Wood DE, Lu J, Langmead B. 2019 Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257. (doi:10.1186/s13059-019-1891-0)

60. Sayers EW *et al.* 2019 Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **47**, D23–D28. (doi:10.1093/nar/gky1069)

61. O'Leary NA *et al.* 2016 Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745. (doi:10.1093/nar/gkv1189)

62. Lu J, Breitwieser FP, Thielen P, Salzberg SL. 2017 Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104. (doi:10.7717/peerj-cs.104)

63. Kellner KF, Swihart RK. 2014 Accounting for imperfect detection in ecology: a quantitative review. *PLoS ONE* **9**, e111436. (doi:10.1371/journal.pone.0111436)

64. Si X *et al.* 2018 The importance of accounting for imperfect detection when estimating functional and phylogenetic community structure. *Ecology* **99**, 2103–2112. (doi:10.1002/ecy.2438)

65. Benoit D, Jackson DA, Ridgway MS. 2018 Assessing the impacts of imperfect detection on estimates of diversity and community structure through multispecies occupancy modelling. *Ecol. Evol.* **8**, 4676–4684. (doi:10.1002/ece3.4023)

66. Aitchison J. 1982 The statistical analysis of compositional data. *J. R. Stat. Soc. Ser. B Methodol.* **44**, 139–160. (doi:10.1111/j.2517-6161.1982.tb01195.x)

67. Filzmoser P, Hron K, Reimann C. 2009 Univariate statistical analysis of environmental (compositional) data: problems and possibilities. *Sci. Total Environ.* **407**, 6100–6108. (doi:10.1016/j.scitotenv.2009.08.008)

68. Muggeo VMR. 2017 Interval estimation for the breakpoint in segmented regression: a smoothed score-based approach. *Aust. N. Z. J. Stat.* **59**, 311–322. (doi:10.1111/anzs.12200)

69. Muggeo VMR. 2003 Estimating regression models with unknown break-points. *Stat. Med.* **22**, 3055–3071. (doi:10.1002/sim.1545)

70. Cox CR, Gilmore MS. 2007 Native microbial colonization of *Drosophila melanogaster* and its use as a model of *Enterococcus faecalis* pathogenesis. *Infect. Immun.* **75**, 1565–1576. (doi:10.1128/IAI.01496-06)

71. Chandler JA, Morgan Lang J, Bhatnagar S, Eisen JA, Kopp A. 2011 Bacterial communities of diverse *Drosophila* species: ecological context of a host–microbe model system. *PLoS Genet.* **7**, e1002272. (doi:10.1371/journal.pgen.1002272)

72. Davies LR, Loeschcke V, Schou MF, Schramm A, Kristensen TN. 2021 The importance of environmental microbes for *Drosophila melanogaster* during seasonal macronutrient variability. *Sci. Rep.* **11**, 18850. (doi:10.1038/s41598-021-98119-0)

73. Coenye T. 2014 The family Burkholderiaceae. In *The prokaryotes* (eds E Rosenberg, EF DeLong, S Lory, E Stackebrandt, F Thompson), pp. 759–776. Berlin, Germany: Springer.

74. Marcondes de Souza JA, Carareto Alves LM, de Mello Varani A, de Macedo Lemos EG. 2014 The family Bradyrhizobiaceae. In *The prokaryotes* (eds E Rosenberg, EF DeLong, S Lory, E Stackebrandt, F Thompson), pp. 135–154. Berlin, Germany: Springer.

75. Vacchini V *et al.* 2017 Bacterial diversity shift determined by different diets in the gut of the spotted wing fly *Drosophila suzukii* is primarily reflected on acetic acid bacteria: acetic acid bacteria of *Drosophila suzukii*. *Environ. Microbiol. Rep.* **9**, 91–103. (doi:10.1111/1758-2229.12505)

76. Obadia B, Keebaugh ES, Yamada R, Ludington WB, Ja WW. 2018 Diet influences host–microbiota

associations in *Drosophila*. *Proc. Natl Acad. Sci. USA* **115**, E4547–E4548. (doi:10.1073/pnas.1804948115)

77. Jiménez-Padilla Y, Esan EO, Floate KD, Sinclair BJ. 2020 Persistence of diet effects on the microbiota of *Drosophila suzukii* (Diptera: Drosophilidae). *Can. Entomol.* **152**, 516–531. (doi:10.4039/tce.2020.37)

78. Adair KL, Douglas AE. 2017 Making a microbiome: the many determinants of host-associated microbial community composition. *Curr. Opin. Microbiol.* **35**, 23–29. (doi:10.1016/j.mib.2016.11.002)

79. Phillips MA *et al.* 2016 Genome-wide analysis of long-term evolutionary domestication in *Drosophila melanogaster*. *Sci. Rep.* **6**, 39281. (doi:10.1038/srep39281)

80. Goodrich JK, Di Rienzi SC, Poole AC, Koren O, Walters WA, Caporaso JG, Knight R, Ley RE. 2014 Conducting a microbiome study. *Cell* **158**, 250–262. (doi:10.1016/j.cell.2014.06.037)

81. The Microbiome Quality Control Project Consortium. 2017 Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nat. Biotechnol.* **35**, 1077–1086. (doi:10.1038/nbt.3981)

82. Sze MA, Schloss PD. 2019 The impact of DNA polymerase and number of rounds of amplification in PCR on 16S rRNA gene sequence data. *mSphere* **4**, e00163-19. (doi:10.1128/mSphere.00163-19)

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

84. Overend G, Luo Y, Henderson L, Douglas AE, Davies SA, Dow JAT. 2016 Molecular mechanism and functional significance of acid generation in the *Drosophila* midgut. *Sci. Rep.* **6**, 27242. (doi:10.1038/srep27242)

85. Fischer CN, Trautman EP, Crawford JM, Stabb EV, Handelsman J, Broderick NA. 2017 Metabolite exchange between microbiome members produces compounds that influence *Drosophila* behavior. *Elife* **6**, e18855. (doi:10.7554/eLife.18855)

86. Schretter CE, Vielmetter J, Bartos I, Marka Z, Marka S, Argade S, Mazmanian SK. 2018 A gut microbial factor modulates locomotor behaviour in *Drosophila*. *Nature* **563**, 402–406. (doi:10.1038/s41586-018-0634-9)

87. Ferreiro A, Crook N, Gasparrini AJ, Dantas G. 2018 Multiscale evolutionary dynamics of host-associated microbiomes. *Cell* **172**, 1216–1227. (doi:10.1016/j.cell.2018.02.015)

88. Blount ZD, Borland CZ, Lenski RE. 2008 Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **105**, 7899–7906. (doi:10.1073/pnas.0803151105)

89. Graves JL *et al.* 2017 Genomics of parallel experimental evolution in *Drosophila*. *Mol. Biol. Evol.* **34**, 831–842. (doi:10.1093/molbev/msw282)

90. Mallard F, Nolte V, Tobler R, Kapun M, Schlötterer C. 2018 A simple genetic basis of adaptation to a novel thermal environment results in complex metabolic rewiring in *Drosophila*. *Genome Biol.* **19**, 119. (doi:10.1186/s13059-018-1503-4)

91. Walters AW *et al.* 2020 The microbiota influences the *Drosophila melanogaster* life history strategy. *Mol. Ecol.* **29**, 639–653. (doi:10.1111/mec.15344)

92. Franssen SU, Kofler R, Schlötterer C. 2017 Uncovering the genetic signature of quantitative trait evolution with replicated time series data. *Heredity* **118**, 42–51. (doi:10.1038/hdy.2016.98)

93. Hayward LK, Sella G. 2021 Polygenic adaptation after a sudden change in environment. bioRxiv, 792952. (doi:10.1101/792952)

94. Petkau K, Fast D, Duggal A, Foley E. 2016 Comparative evaluation of the genomes of three common *Drosophila*-associated bacteria. *Biol. Open* **5**, 1305–1316. (doi:10.1242/bio.017673)

95. Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. 2017 The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (doi:10.1038/nature24287)

96. Douglas AE. 2016 How multi-partner endosymbioses function. *Nat. Rev. Microbiol.* **14**, 731–743. (doi:10.1038/nrmicro.2016.151)