Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Unhappy in Its Own Way: LDA Topic Modeling Reveals the Anna Karenina Effect in Daphnia's Microbiome Exposed to Microplastics

Pooja Haushiram Wakchoure
KU Leuven
Anna Greffe
KU Leuven
Karoline Faust
KU Leuven
Ellen Decaestecker
KU Leuven
Jan Aerts
jan.aerts@kuleuven.be
KU Leuven

Research Article

Keywords:

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11636409/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

"Unhappy in Its Own Way": LDA Topic Modeling Reveals the Anna Karenina Effect in *Daphnia*'s Microbiome Exposed to Microplastics

Pooja Wakchoure[1], Anna Greffe[2], Karoline Faust[2], Ellen Decaestecker[3], Jan Aerts[1,*]

[1] Department of Biosystems, Mechatronics, Biostatistics and Sensors (MeBioS), KU Leuven, Kasteelpark Arenberg 30, 3001 Leuven, Belgium

[2] Department of Microbiology, Immunology and Transplantation, Rega Institute for Medical Research, Laboratory of Molecular Bacteriology, KU Leuven, Herestraat 49, 3000 Leuven, Belgium

[3] Department of Biology, Laboratory of Aquatic Biology, Microbiome Eco-Evo Group, KU Leuven, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium

[*] Corresponding author: jan.aerts@kuleuven.be

Abstract

Background: Host-associated microbiomes play important roles in animal health and responses to environmental stress, yet how microplastic (MP) exposure influences among-individual microbiome variability remains poorly understood. This study investigated whether MP exposure is associated with increased microbiome heterogeneity in *Daphnia magna*, consistent with the Anna Karenina principle (AKP).

Results: We reanalyzed 16S rRNA gene amplicon sequencing data from a controlled MP exposure experiment using Latent Dirichlet Allocation (LDA) topic modelling, community composition analyses, and beta-dispersion tests. LDA identified six microbial components representing distinct bacterial assemblages. Pre-exposure

microbiomes showed relatively homogeneous component profiles, whereas adults sampled after 23 days exhibited more heterogeneous microbial-component mixtures and significantly greater beta dispersion. Within the day-23 adult group, MP-exposed individuals showed significantly greater dispersion than no-plastic controls, despite no significant difference in overall community composition. Long-exposure survivors exhibited lower microbiome variability and greater similarity to pre-exposure samples, although their community composition remained significantly different. Microbial-component profiles were also associated with experimental context, particularly bacterioplankton inoculum source.

Conclusions: MP exposure was associated with increased among-individual variability in the adult *Daphnia magna* microbiome, consistent with an Anna Karenina-type response to environmental disturbance. Combining LDA topic modelling with ASV-level analyses highlights the importance of examining microbiome variability alongside compositional changes to better understand host-associated microbiome responses to environmental stress.

Keywords: *Daphnia magna*; Microplastics; Anna Karenina principle; Microbiome variability; Latent Dirichlet Allocation; Beta dispersion; Microbial community composition

1 Introduction

Microplastic (MP) pollution is a growing threat to aquatic ecosystems worldwide and represents one of the key environmental challenges of the 21st century [1]. MPs, defined as plastic particles smaller than 5 mm, are now widely recognized as ubiquitous contaminants in freshwater ecosystems, accumulating in both the water column and sediments of ponds and lakes, particularly in urbanized areas [2]. Because they can be ingested by a wide range of aquatic organisms such as filter-feeders and provide surfaces for microbial colonization, MP particles may interact with the biota not only as physical and chemical stressors but also as potential microbial habitats for their microbial symbionts [2, 3]. Here we focus on *Daphnia magna*, which is a well-established model organism in ecology and ecotoxicology, especially for microbiome studies [4]. Given its central position in the aquatic food web, understanding how MP exposure affects *Daphnia*, therefore, has implications that extend

beyond individual host responses to broader freshwater ecosystem functioning [5, 2, 4].

Host-associated microbiome studies often focus on average bacterial compositional shifts, such as bacterial taxa that increase or decrease after exposure to a stressor [6]. However, this can overlook variation in how individual microbiomes respond to disturbance. This phenomenon can be understood through the Anna Karenina principle of animal microbiomes [7], inspired by Leo Tolstoy's observation that "all happy families look alike; each unhappy family is unhappy in its own way." Applied to animal microbiomes, the Anna Karenina principle (AKP) proposes that healthy or stable microbiomes tend to resemble one another, whereas stressed or disturbed microbiomes become more variable across individuals. Rather than producing a uniform shift toward a single altered microbiome state, environmental stress may lead to distinct microbial community profiles among hosts [7]. For MP exposure, this suggests that *Daphnia* microbiomes may not only change in bacterial composition but also become more variable across individuals.

Our previous analysis of MP effects on the *Daphnia* microbiome largely focused on changes in average community composition and differential taxonomic abundance across treatment conditions [2]. However, whether MP exposure is associated with increased among-individual microbiome variability remains insufficiently explored. To address this gap, we reanalyzed *Daphnia*-associated microbiome data from the previously published controlled MP exposure experiment [2]. We integrated Latent Dirichlet Allocation (LDA) topic modelling, beta-dispersion analysis, PERMANOVA, and differential abundance analysis to examine changes in microbial-component profiles, community composition, and among-individual variability. In addition, comparisons between MP-exposed individuals and no-plastic controls at the 23-day sampling point were used to evaluate MP-associated differences in microbiome variability. This approach provides a complementary framework for examining microbiome responses to environmental disturbance beyond average taxonomic shifts.

2 Methods

2.1 *In-vitro* MP exposure experiment

The full methodology is described in Krzynowek et al. [2]. In brief, the 16S rRNA gene amplicon sequencing data used in this study were generated from a controlled *in-vitro* MP exposure experiment of *Daphnia* magna (2). In this experiment, *D. magna* clones (F, KNO15.04, and BH) were exposed for 23 days to MPs of three plastic polymers, i.e., polyethylene terephthalate (PET), nylon, and polylactic acid (PLA). No-MP treatment was also included as a control. Furthermore, the experiment was designed to test whether the source of the surrounding bacterioplankton influenced the composition of the *Daphnia*-associated microbiome. Therefore, each MP treatment was combined with one of three bacterioplankton inoculum sources: Blauwe Poort, a high-MP urban pond; De Gavers, a low-MP lake; or sterile water as a bacterium-free control.

Samples were collected before exposure and after 23 days of MP co-culture. The pre-exposure samples represented clone-specific controls. At the 23-day sampling point, both adult and juvenile *Daphnia magna* were collected, together with the surrounding bacterioplankton, to compare host-associated and water microbial communities. Juveniles sampled at this time point were offspring produced during the experiment and therefore had shorter individual exposure histories. An additional group of surviving individuals was collected from the experimental jars beyond the scheduled 23-day exposure period. These samples were categorized as long-term survivors, although their exact sampling time was not documented.

2.2 Processing of amplicon sequencing reads

The quality of demultiplexed paired-end reads was assessed using FastQC. The run generated 10832180 raw sequences (min 7781, max 55923 per sample). Paired-end sequences were filtered, denoised, and dereplicated with DADA2 [8] in RStudio. Taxonomy was assigned using the Silva 138.1 SSU Ref NR 99 database. Reads labeled as “mitochondria” or “chloroplast” were removed. The ASV table, taxonomy, and metadata were imported

into a phyloseq object (v. 1.44.0).

For the LDA topic modeling, we used raw ASV count data rather than rarefied data because LDA is designed to model count-based observations directly. To reduce noise, we applied an ASV prevalence threshold of 5% and a minimal sample sequencing depth of 100 reads. Only host-associated *Daphnia* samples relevant to the experiment were retained, while environmental samples and sterile-water controls were excluded. After filtering, the final dataset comprised 117 *Daphnia* samples: pre-exposure controls (n = 9), 23-day exposure adults (n = 53), 23-day exposure juveniles (n = 41), and long-exposure survivors (n = 14). This filtered count matrix was used for LDA topic modelling and subsequent downstream analysis.

2.3 Analysis

The visual analytics approach was used as the guiding research approach in this study. This approach integrates computational analysis and machine learning with interactive visualization, allowing complex biological data to be explored, compared, and interpreted in a more transparent and hypothesis-driven way.

2.3.1 Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) is a generative probabilistic model originally developed for discrete data, such as collections of text documents [9]. In LDA, each sample is represented as a mixture of latent topics, rather than being assigned to a single group. Each topic is defined by a probability distribution over the observed features [9]. In the context of microbiome data, this means that each sample can contain different proportions of microbial "topics" and each topic is characterized by a distinct combination of ASVs.

Within the LDA framework, samples were represented as documents and ASVs as words, with each sample encoded as a count-based “bag of words” in which ASVs appeared according to their observed read counts. The inferred latent topics were interpreted as microbial components (MCs), defined as recurrent assemblages of co-occurring ASVs. LDA provides a soft clustering representation: each sample is described by a probability distribution over MCs, and each MC is described by a probability distribution over

ASVs. Thus, samples can contain multiple microbial components simultaneously, and individual ASVs can contribute to several components with different probabilities. This representation reflects the overlapping and compositional nature of microbiome data [10]. The resulting per-sample MC probability vectors were used as compositional state vectors for downstream analyses, while the per-MC ASV probability distributions were used to examine the taxonomic identity of each component.

LDA models were fitted using the gensim (v4.4.0) Python library [11]. Models were trained with 10 passes over the corpus, an automatically optimized asymmetric document–topic prior (alpha = "auto"), and a fixed random seed to ensure reproducibility. To determine the optimal number of microbial components, models were fitted for candidate values ranging from $k = 2$ to $k = 12$. For each model, we calculated the topic coherence score [12] and log-perplexity. The final model was selected based on the highest coherence score, resulting in $k = 6$ microbial components, labeled MC1–MC6. Per-sample MC probabilities were extracted from the fitted model, including components with zero probability, and were renormalized to sum to one before downstream analyses.

2.3.2 Statistical analysis of microbial components from topic modeling

The probability matrix of sample-MC derived from LDA modeling was used to evaluate whether experimental variables such as exposure category, inoculum pond, *Daphnia* clone, plastic treatment type, and plastic content associated with microbial component profiles and whether MP exposure increased the among-sample variability. Each sample was represented by a probability vector across six microbial components, MC1-MC6, obtained from the final LDA model, because these vectors represent compositional probability profiles. Pairwise differences between samples were calculated using Jensen-Shannon distance [13].

Principal coordinate analysis (PCoA) was used to visualize the difference in microbial component profiles among exposure categories. To test whether experimental variables explained variation in overall microbial-component composition, permutational multivariate analysis of variance (PERMANOVA) was performed on the Jensen-Shannon distance matrix

and PERMDISP was used to test for differences in among-sample dispersion between exposure categories.

2.3.3 Statistical analysis of ASV-level community composition

Differences in microbiome composition and dispersion were analyzed at the ASV level using Bray–Curtis dissimilarity [14]. To compare microbiome composition and dispersion across adult exposure times, the primary analysis was restricted to pre-exposure adults (n = 9), adults sampled at day 23 (n = 53), and long-exposure survivors (n = 14). Juvenile samples were excluded from this comparison. ASVs were retained when detected in at least three samples and represented by at least ten reads across the dataset. Counts were subsequently converted to relative abundances before calculating Bray–Curtis dissimilarities.

Differences in community composition among exposure times were assessed using permutational multivariate analysis of variance (PERMANOVA; `adonis2`, vegan) with 9,999 permutations, followed by pairwise PERMANOVA comparisons between exposure times. Because PERMANOVA results can be influenced by differences in within-group variability, compositional differences were interpreted alongside tests of multivariate homogeneity of dispersion (PERMDISP; `betadisper`, vegan). For PERMDISP, the distance of each sample from its group centroid was calculated in principal-coordinate space, and differences among exposure times were evaluated using 9,999 permutations, with centroids specified as the group reference. Community structure and the relative positions of group centroids were visualized using principal coordinate analysis (PCoA). Distributions of pairwise Bray–Curtis dissimilarities were also examined to compare within-group variability and between-group separation.

Adult and juvenile microbiomes sampled at day 23 were compared directly using the same Bray–Curtis PERMANOVA and PERMDISP procedures to assess differences in community composition and dispersion between developmental stages. To investigate the association between MP treatment and microbiome variability, MP-exposed animals were compared with no-plastic controls separately within adults and juveniles using Bray–Curtis PERMANOVA and PERMDISP.

Because sample sizes differed among adult exposure times, the robustness of the dispersion results to group-size imbalance was additionally assessed using repeated balanced subsampling. In each of 1,000 iterations, each exposure group was subsampled to the size of the smallest group (n = 9), and Bray–Curtis multivariate dispersion was recalculated using bias-adjusted distances to group centroids. The direction and distribution of differences in mean dispersion were summarized across iterations.

2.3.4 Family-level differential abundance analysis

Differentially abundant families across exposure times were identified using ANCOM-BC[15], applied to the adult dataset with the same filtering as above. Taxa were agglomerated to family level, and abundance was modelled as a function of exposure condition, yielding contrasts of 23 days exposure and long-term survivors against the pre-exposure baseline, together with a global test across all stages. Analyses used a prevalence cut-off of 0.10 and a library-size cut-off of 1,000 reads, with detection of structural zeros enabled and the conservative variance estimator applied. P-values were adjusted for multiple comparisons using the Benjamini–Hochberg false-discovery-rate procedure, and families with $q < 0.05$ were considered differentially abundant. Bias-corrected log abundances (using the ANCOM-BC sampling-fraction estimates) were used for visualization. To focus interpretation on biologically meaningful shifts, significant families were additionally filtered by effect size ($|\text{log-fold change}| \geq 0.5$) when summarizing the overall response.

3 Results

The following section is presented in a stepwise manner to first define the microbial components identified by LDA and then evaluate their biological and experimental relevance. Section 3.1 describes the six LDA-derived microbial components and their distribution across individual samples. Section 3.1.1 then characterizes the bacterial families defining each component. The following sections examine how microbial component profiles change across exposure times and developmental stages and assess whether these patterns are also reflected at the ASV and family levels.

3.1 LDA identifies six microbial components in the *Daphnia* microbiome

LDA was applied to the ASV count matrix from *Daphnia* samples to identify recurrent microbial components (MCs). Among the evaluated models (k = 2–12), the six-component model achieved the highest topic coherence score (0.7713) and was selected for subsequent analyses (Supplementary Fig. S1 in the supplementary file 1). The six components are hereafter referred to as MC1-MC6.

The resulting sample-component profiles are shown in Fig. 1. Additional visualizations of the sample-component probability matrix using different sample-ordering strategies are provided in the supplementary figures S2, S3 and S4 in the supplementary file 1. Most samples were dominated by one or a small number of components, although the relative contribution of individual components varied among samples. MC3 was broadly represented across the dataset, whereas the remaining components were more prominent in particular subsets of samples. Differences in component profiles among exposure times, developmental stages, and other experimental variables are examined statistically in the following sections.

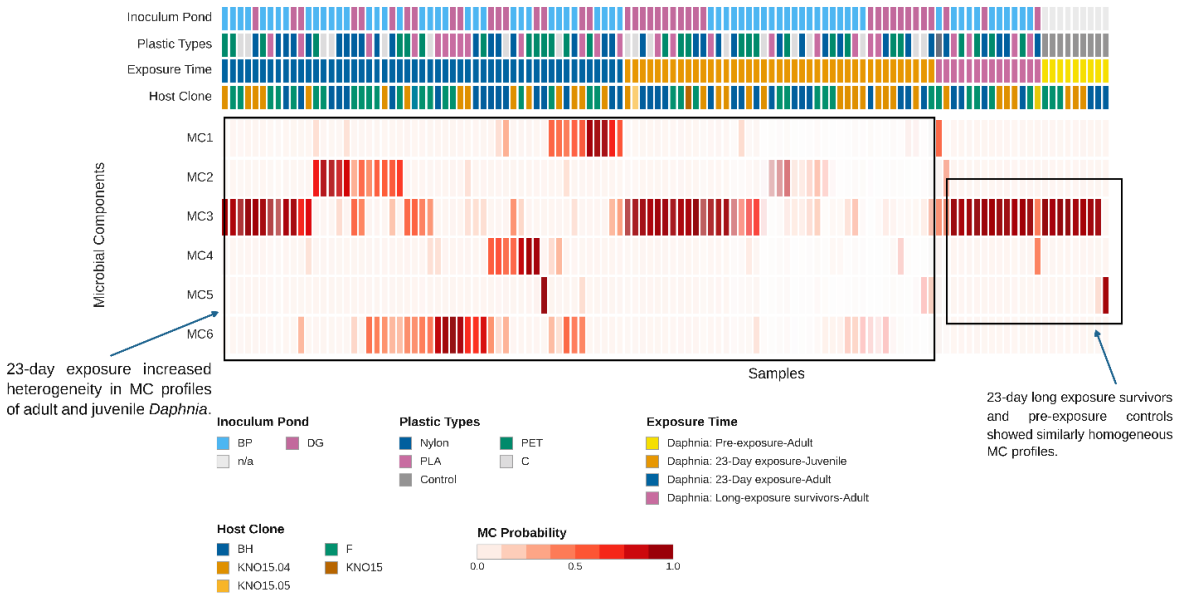


Figure 1: Distribution of LDA-derived microbial components across *Daphnia* samples. Each column represents one of the 117 *Daphnia* samples and each row represents one of the six microbial components (MC1–MC6). Cell values indicate the sample-level LDA component weight, with higher values representing a greater contribution of that component to the corresponding microbiome. Samples are ordered by exposure group and annotated by inoculum source, plastic treatment, exposure group, and host clone.

3.1.1 Taxonomic signatures of microbial components

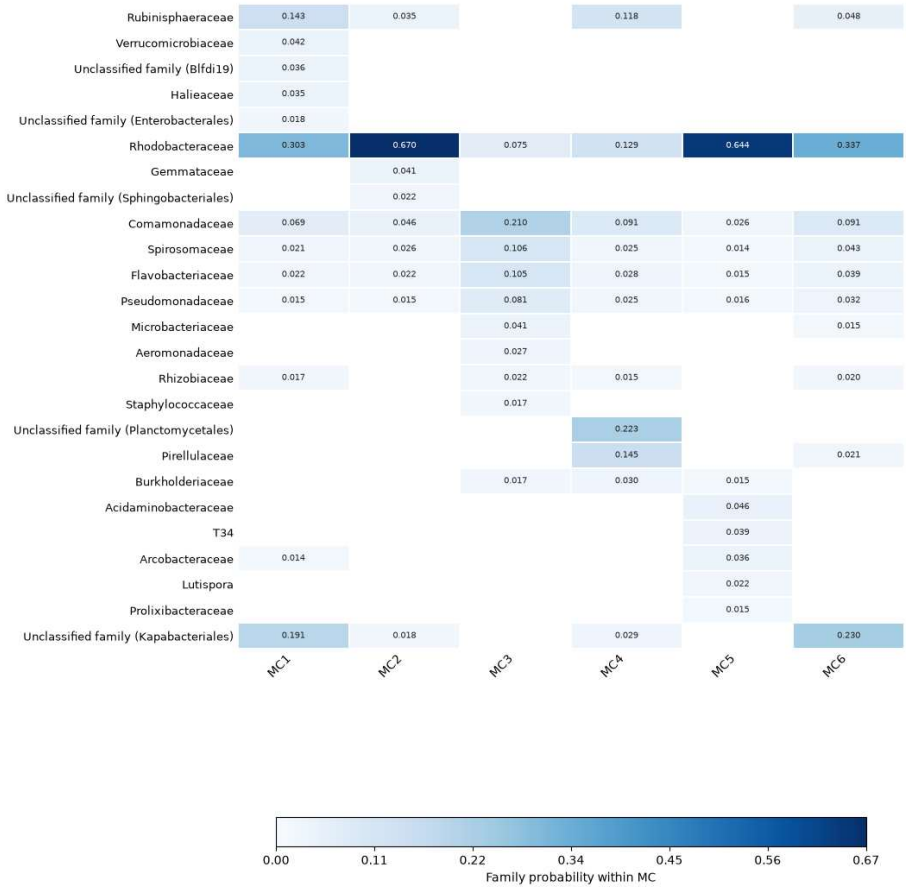


Figure 2: Taxonomic composition of the six LDA-derived microbial components. Columns represent microbial components MC1–MC6 and rows represent bacterial families. Cell values indicate the probability of each family within the corresponding microbial component, with higher probabilities indicating stronger representation of that family within the component.

The six LDA-derived microbial components showed distinct family-level taxonomic profiles (Fig. 2). MC1 was characterized primarily by *Rhodobacteraceae* (probability = 0.303), *Kapabacteriales* (0.191), and *Rubinisphaeraceae* (0.143). MC2 showed the strongest dominance by a single family, with *Rhodobacteraceae* accounting for a probability of 0.670, while the remaining families contributed comparatively little. MC3 displayed a more even family-level composition, with *Comamonadaceae* (0.210), *Spirosomaceae* (0.106), *Flavobacteriaceae* (0.105), and *Pseudomonadaceae* (0.081) among its most prominent families. MC4 was characterized by *Planctomycetales* (0.223), *Pirellulaceae* (0.145), *Rhodobacteraceae* (0.129), and *Rubinisphaeraceae* (0.118). MC5 was again strongly dominated by *Rhodobacteraceae* (0.644), whereas MC6 showed a more mixed profile, characterized mainly by *Rhodobacteraceae* (0.337), *Kapabacteriales* (0.230), and *Comamonadaceae* (0.091).

Overall, the microbial components ranged from strongly family-dominated profiles, such as MC2 and MC5, to more taxonomically mixed components such as MC3, MC4, and MC6, demonstrating that the LDA components represented distinct bacterial assemblages.,

3.2 Exposure time is associated with restructuring and increased microbiome variability

Differences in microbial composition across exposure times can reflect two distinct patterns: a shift in the group centroid, indicating a change in overall community composition, and a change in within-group dispersion, indicating differences in among-sample variability. Both patterns can be examined at the level of microbial components derived from topic modelling and at the level of the underlying ASV composition. Section 3.2.1 evaluates shifts in microbial component profiles across exposure groups, whereas Section 3.2.2 examines differences in their dispersion. Sections 3.2.3 and 3.2.4 evaluate the corresponding patterns at the ASV level, restricted to adult samples, to determine whether the component-level patterns are also evident in the underlying taxonomic composition independently of the topic-modelling representation.Adults and juveniles sampled at day 23 showed distinct microbial-component profiles, with differences between developmental stages examined further in Section 3.3.

3.2.1 Topic-level centroid shifts across exposure times

Microbial-component profiles differed across exposure times (Fig. 3). In the Jensen–Shannon PCoA, samples collected after 23 days of exposure both adults and juvenile occupied a broader region of ordination space than pre-exposure samples, particularly among adults. Long-term survivors were positioned closer to the pre-exposure group. The first two PCoA axes explained 42.0% and 20.3% of the variation, respectively. exposure time was significantly associated with overall microbial-component composition (PERMANOVA, R^2=0.131, p=0.0001). These results indicate significant differences in microbial-component composition across exposure times, with adults sampled at day 23 showing the most heterogeneous component profiles.

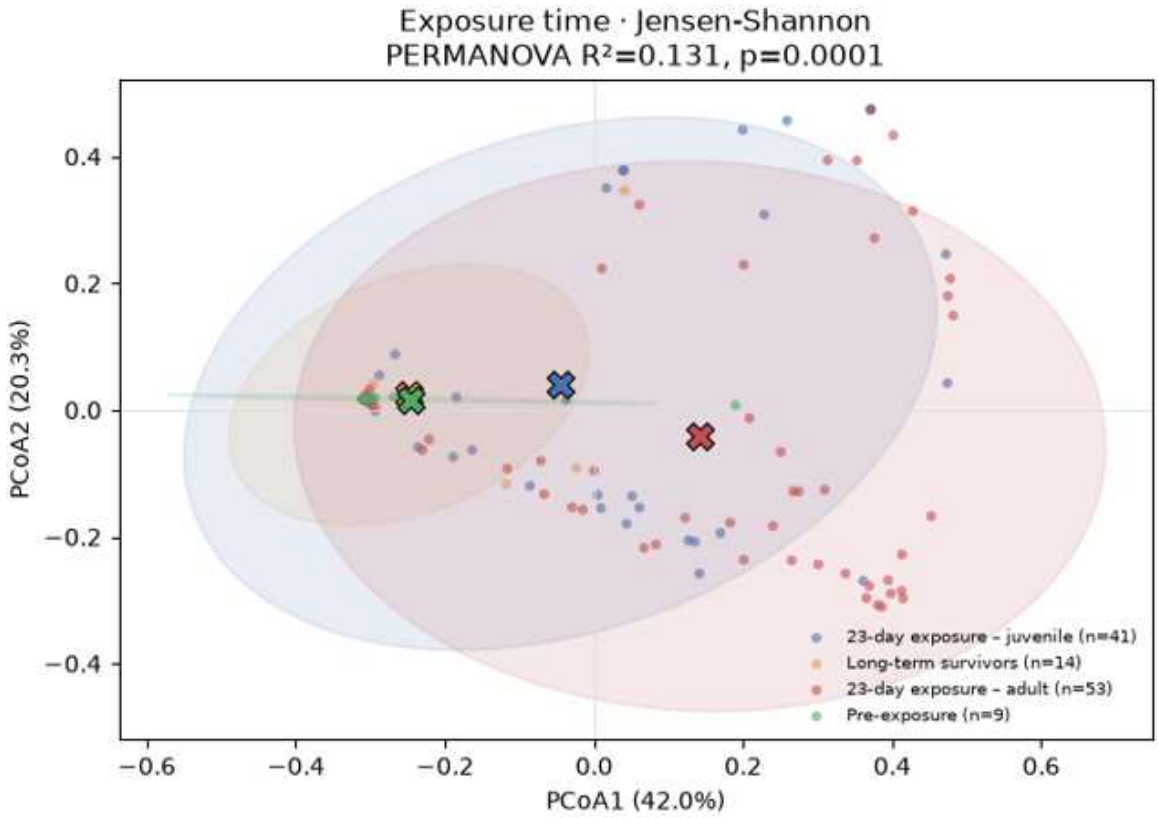


Figure 3: Exposure time explains variation and dispersion in microbial component profiles. Principal coordinate analysis (PCoA) of Jensen–Shannon distances calculated from sample-level microbial-component probability profiles. Points represent individual samples grouped as pre-exposure, 23-day exposure – adult, 23-day exposure – juvenile, and long-term survivors. Crosses indicate group centroids and ellipses summarize the distribution of samples within each exposure time. exposure time was significantly associated with microbial-component composition (PERMANOVA, R^2 =0.131, p=0.0001).

3.2.2 Topic-level dispersion changes across exposure times

Multivariate dispersion of microbial-component profiles differed significantly across exposure times (PERMDISP, F=33.408, p=0.0002; Fig. 4). Adult samples from the 23-day exposure group showed the greatest variability, with a mean distance to the group centroid of approximately 0.46. Juvenile samples collected after 23 days of exposure were less dispersed than adults, with a mean distance of approximately 0.33. In contrast, pre-exposure samples and long-term survivors showed comparatively low dispersion, with mean distances to the group centroid of approximately 0.13 and 0.08, respectively. These results demonstrate greater among-sample heterogeneity at day 23, particularly among adults, while long-term survivors exhibited more homogeneous microbial-component profiles.

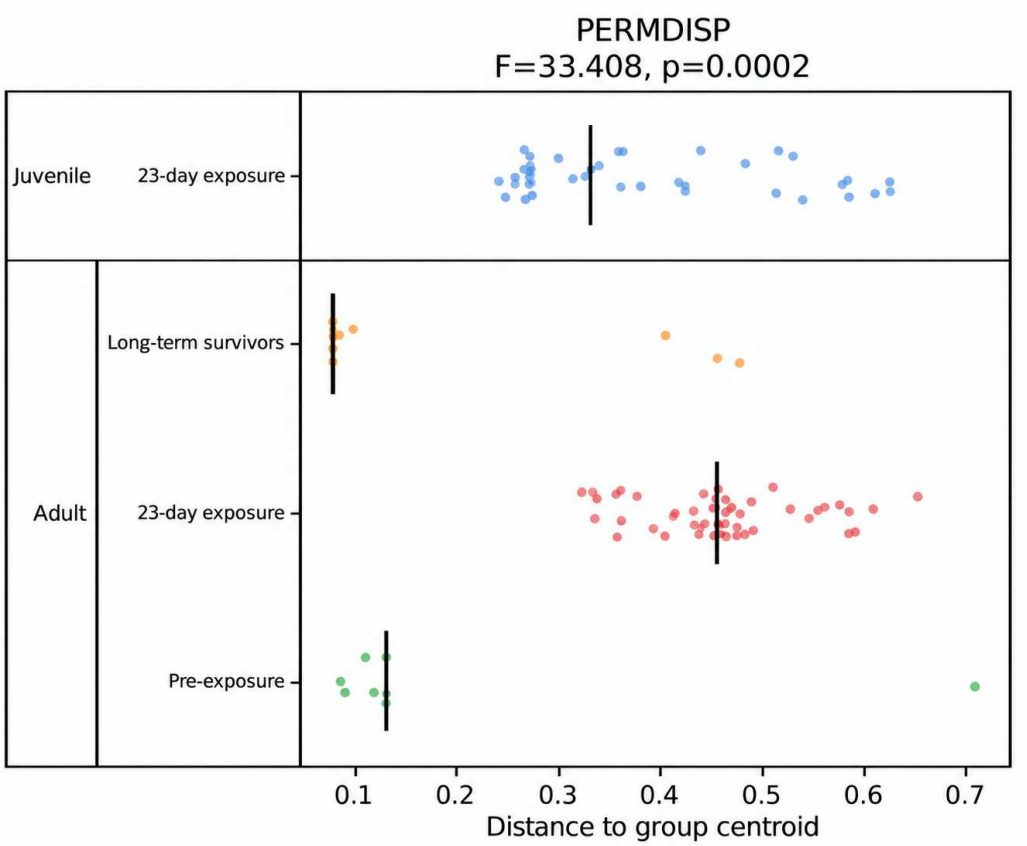


Figure 4: Variability in microbial-component profiles across exposure times. Multivariate dispersion was quantified as the distance of each sample to its group centroid in the Jensen–Shannon-based ordination space. Higher distances indicate greater among-sample variability in microbial-component profiles. Points represent individual samples and black bars indicate group-level summaries. Dispersion differed significantly among exposure times (PERMDISP, F=33.408, p=0.0002).

3.2.3 ASV-level centroid shifts across exposure times in adults

Bray–Curtis PCoA showed differences in adult microbiome composition across exposure times (Fig. 5A). Adults sampled after 23 days of exposure occupied a broader region of the ordination space, whereas pre-exposure samples and long-term survivors were more compact and positioned closer to one another. The first two PCoA axes explained 31.4% and 13.9% of the variation, respectively. Pairwise PERMANOVA confirmed significant differences in community composition among all three adult exposure times (Supplementary file 2[Table_S2A]) . Pre-exposure adults differed from adults after 23 days of exposure (R^2=0.064, p=0.0011), 23-day exposure adults differed from long-term survivors (R^2=0.081, p=0.0001), and pre-exposure adults also differed from long-term survivors (R^2=0.088, p=0.0354).

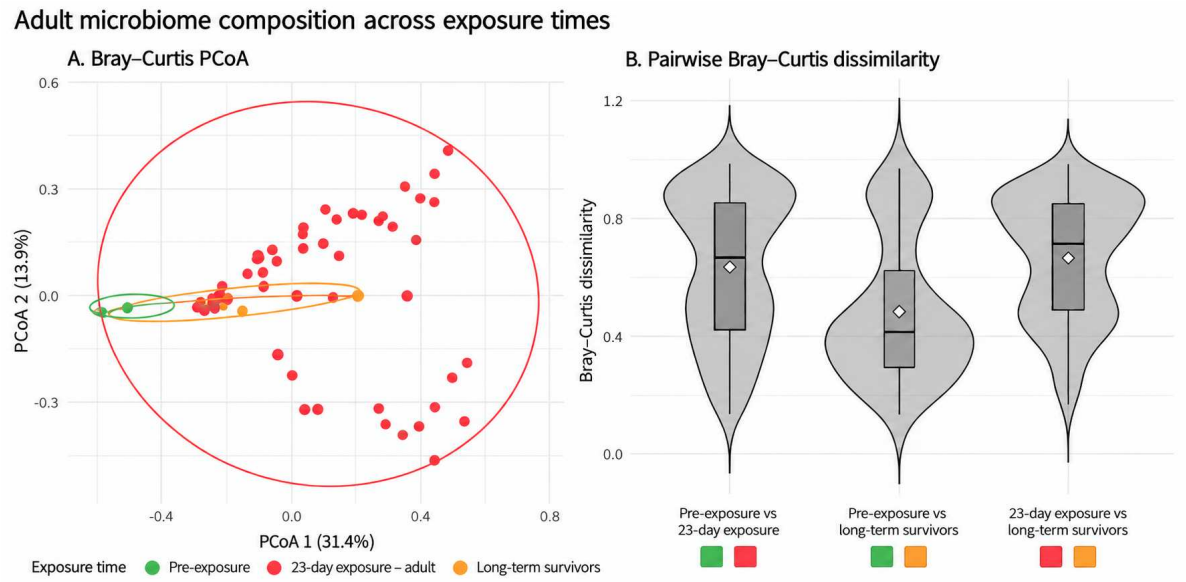


Figure 5: Adult microbiome composition across exposure times. (A) Principal coordinates analysis (PCoA) based on Bray–Curtis dissimilarities for pre-exposure adults, adults after 23 days of exposure, and long-term survivors. PCoA1 and PCoA2 explained 31.4% and 13.9% of the variation, respectively. Ellipses summarize the distribution of samples within each exposure time. (B) Distribution of pairwise Bray–Curtis dissimilarities between adult exposure times. Violin plots show the distribution of pairwise dissimilarities, boxplots indicate the median and interquartile range, and diamonds indicate the mean.

Pairwise Bray–Curtis dissimilarities showed the same overall pattern descriptively (Fig. 5B), with lower dissimilarity between pre-exposure samples and long-term survivors than in comparisons involving the 23-day exposure group. Thus, although long-term survivors were more similar to the pre-exposure state, their microbiome composition remained detectably different.

3.2.4 ASV-level dispersion changes across exposure times in adults

Bray-Curtis beta dispersion differed significantly across adult exposure times (PERMDISP, F=14.64, p=0.0001; Fig. 6). Adults sampled after 23 days of exposure showed the greatest dispersion, with a mean distance to the group centroid of 0.478, compared with 0.276 for pre-exposure adults and 0.335 for long-term survivors.

Pairwise comparisons confirmed greater dispersion in the 23-day exposure group than in the pre-exposure group (p=0.0001), and dispersion was significantly lower among long-term survivors than in the 23-day exposure group (p=0.0005). In contrast, dispersion did not differ significantly between pre-exposure adults and long-term survivors (p=0.4218). These results indicate greater adult microbiome heterogeneity at the 23-day sampling point,

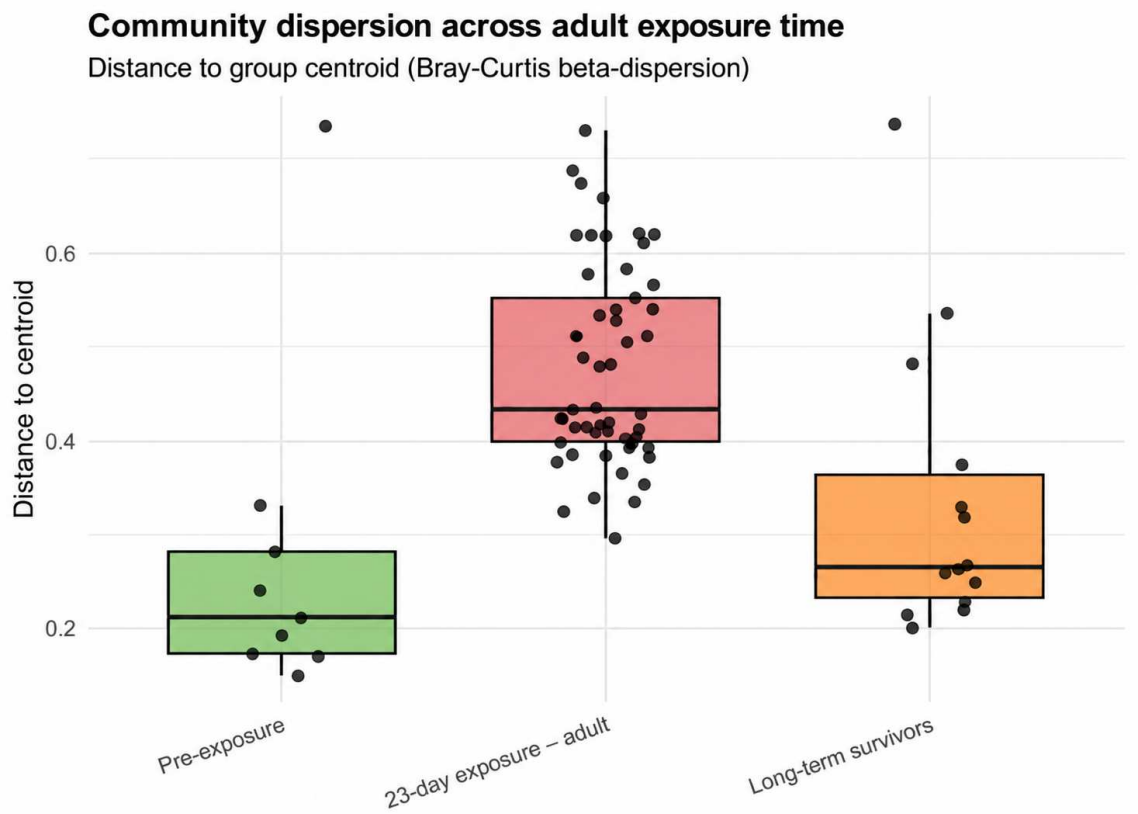


Figure 6: Bray–Curtis beta dispersion across adult exposure times. Beta dispersion was quantified as the distance of each sample to its group centroid in principal-coordinate space derived from Bray–Curtis dissimilarities. Higher distances indicate greater among-individual variability in microbiome composition. Boxplots show pre-exposure adults, adults after 23 days of exposure, and long-term survivors, with points representing individual samples. Dispersion differed significantly among exposure times (PERMDISP, F=14.64, p=0.0001).

whereas long-exposure survivors showed lower variability, approaching the pre-exposure level. This pattern was robust to unequal group sizes. Across 1,000 balanced subsampling iterations (n = 9 per group), 23-day exposure adults were more dispersed than pre-exposure adults in 100% of iterations and more dispersed than long-term survivors in 98.1% of iterations (Supplementary_File_2 [Table_S3]).

3.3 Developmental stage modulates microbiome variability after 23 days of exposure

Adult and juvenile microbiomes differed in their variability after 23 days of exposure (Fig. 7A). Adults showed significantly greater beta dispersion than juveniles (PERMDISP, F=20.03, p=0.0001), with mean distances to the group centroid of 0.479 and 0.369, respectively.

To determine whether this difference was associated specifically with MP exposure,

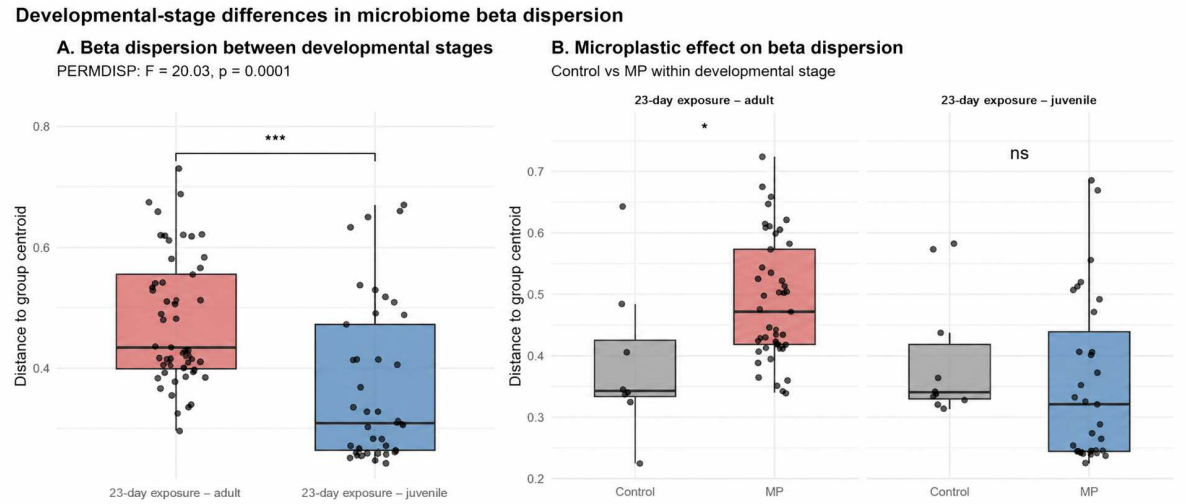


Figure 7: Developmental-stage differences in microbiome beta dispersion after 23 days of exposure. (A) Beta dispersion was significantly greater in adults than in juveniles after 23 days of exposure (PERMDISP, F=20.03, p=0.0001). (B) Beta dispersion of MP-exposed (MP) animals and no-plastic controls compared separately within adults and juveniles. MP-exposed adults showed significantly greater dispersion than controls (p=0.0167), whereas no significant difference was detected in juveniles (p=0.4378). Distances represent sample-to-group-centroid distances in principal-coordinate space derived from Bray-Curtis dissimilarities. Points represent individual samples and boxplots summarize their distributions.

MP-exposed animals were compared with no-plastic controls separately within each developmental stage (Fig. 7B). Among adults, MP-exposed individuals showed significantly greater beta dispersion than controls (mean distance to centroid: 0.489 vs 0.388; p=0.0167). In contrast, no significant difference in dispersion was detected between MP-exposed and control juveniles (0.357 vs 0.394; p=0.4378). Community-composition analyses showed a modest difference between adult and juvenile microbiomes after 23 days of exposure (PERMANOVA, R^2=0.048, p=0.0008). However, within developmental stages, MP-exposed and control groups did not differ significantly in overall community composition in either adults (p=0.5526) or juveniles (p=0.4207). Full PERMANOVA and PERMDISP statistics are provided in the supplementary file 2[Table_S4].

The comparison between MP-exposed individuals and no-plastic controls revealed significantly greater microbiome dispersion among exposed adults, whereas no significant difference was detected among juveniles at the same sampling point.

3.4 Experimental context shapes microbial-component profiles

Microbial-component composition was structured by several experimental factors, although their contributions differed in magnitude (Fig. 8A). Exposure time explained the largest

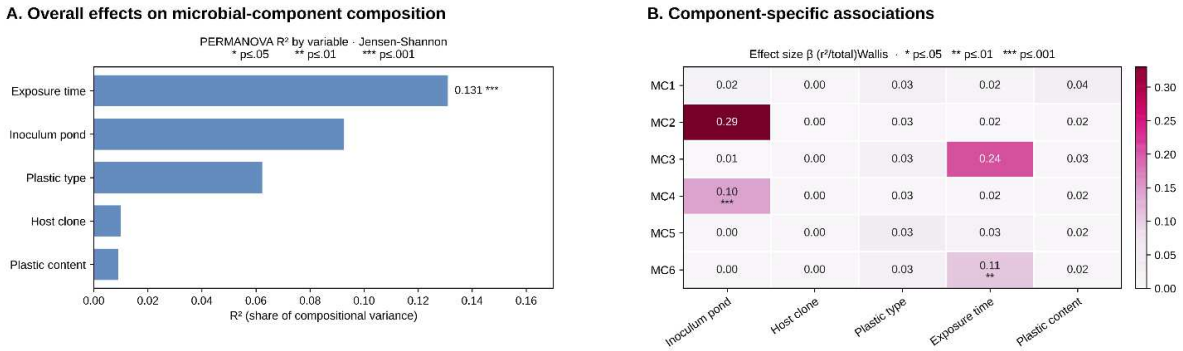


Figure 8: Experimental context shapes microbial-component profiles. (A) PERMANOVA R^2 values showing the proportion of variation in Jensen-Shannon microbial-component composition explained by each experimental variable. Asterisks indicate significant PERMANOVA effects. (B) Component-specific Kruskal–Wallis effect sizes ($\eta 2$) for microbial components MC1-MC6 across experimental variables. Higher values indicate stronger associations, and asterisks denote significance after false-discovery-rate correction ($*q \leq 0.05$, $**q \leq 0.01$, $***q \leq 0.001$).

proportion of variation in overall component profiles (PERMANOVA, R2=0.131, p=0.0001), followed by inoculum pond (R2=0.093, p<0.01) and plastic type (R2=0.062, p<0.05). In contrast, host clone (R2=0.010) and plastic content (R2=0.009) explained little variation and were not significant.

Component-wise analyses showed that these overall effects were concentrated in specific microbial components (Fig. 8B). Inoculum pond was most strongly associated with MC2 ($\eta 2$=0.29, q<0.001) and MC4 ($\eta 2$=0.10, q<0.001). exposure time was primarily associated with MC3 ($\eta 2$=0.24, q<0.001) and, to a lesser extent, MC6 ($\eta 2$=0.11, q<0.01). Although plastic type was significantly associated with overall microbial-component composition in the PERMANOVA, no individual component showed a significant plastic-type association after multiple-testing correction. Likewise, host clone and plastic content showed no significant component-specific associations.

3.5 Family-level differential abundance across exposure conditions

Family-level differential abundance analysis (ANCOM-BC) identified 40 bacterial families that differed significantly between adults after 23 days of exposure and pre-exposure adults (q<0.05), compared with only 13 families between long-term survivors and pre-exposure adults. This reduction indicates that many of the taxonomic shifts detected after 23 days of exposure were no longer detectable among long-term survivors, consistent with the reduced microbiome variability observed at this stage.

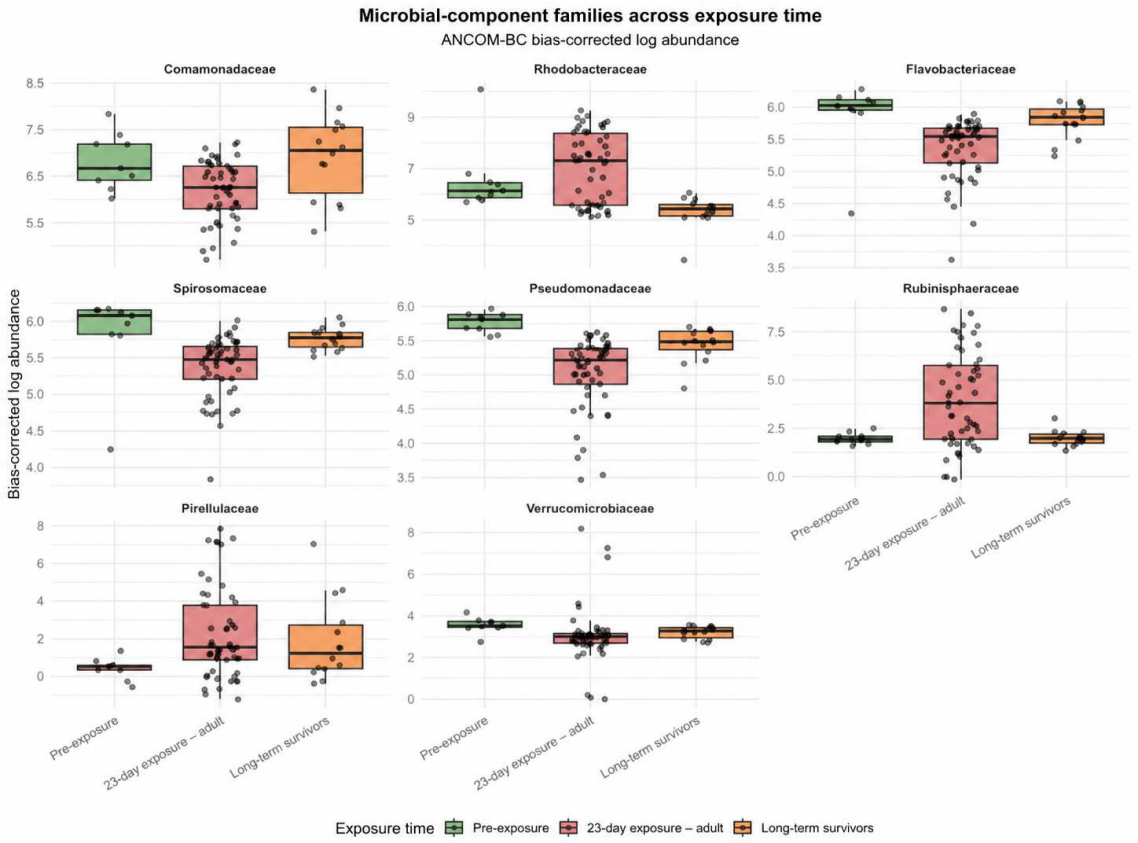


Figure 9: Family-level abundance patterns across adult exposure times. ANCOM-BC bias-corrected log abundances are shown for bacterial families represented in the LDA-derived microbial components across pre-exposure adults, adults after 23 days of exposure, and long-term survivors. Boxplots show the distribution within each exposure time and points represent individual samples.

Among families contributing strongly to the LDA-derived microbial components, the differential-abundance patterns were consistent with the component-level restructuring. *Pirellulaceae*, a prominent family in MC4 (family probability = 0.145), increased strongly after 23 days of exposure (LFC =+1.86, q<0.001). *Rubinisphaeraceae*, which contributed strongly to both MC1 (0.143) and MC4 (0.118), also increased markedly (LFC =+1.97, q<0.001). MC4 was additionally characterized by an unclassified *Planctomycetales* group, indicating a strong *Planctomycetes*-related signature within this component.

In contrast, several families characteristic of MC3 declined after 23 days of exposure. MC3 was strongly represented by *Comamonadaceae* (0.210), *Flavobacteriaceae* (0.105), and *Pseudomonadaceae* (0.081). *Pseudomonadaceae* decreased significantly after exposure (LFC =−0.74, q<0.001), as did *Comamonadaceae* (LFC =−0.63, q=0.04). *Flavobacteriaceae* showed a similar negative trend, although this difference was not significant after multiple-testing correction (LFC =−0.52, q=0.07). Thus, the family-level results

support a reduction in taxa characteristic of MC3 alongside increased representation of *Planctomycetes*-associated families contributing to MC4.

Beyond the families strongly represented within the microbial components, the broader differential-abundance analysis revealed additional exposure-associated changes (Supplementary file 2 Table [S5]). Several freshwater *Alphaproteobacteria* declined after 23 days of exposure, including *Sphingomonadaceae* (LFC =−1.58), *Beijerinckiaceae* (LFC =−1.88), and *Xanthobacteraceae* (LFC =−1.07). In contrast, *Bacteriovoracaceae* increased (LFC =+0.83), indicating that the response extended beyond the component-defining families and involved broader reorganization of the adult microbiome.

Among long-term survivors, several of the strongest changes observed after 23 days of exposure were no longer significant relative to pre-exposure adults. *Rubinisphaeraceae, Pseudomonadaceae,* and *Comamonadaceae* were not significantly different from the pre-exposure group (q>0.05). *Pirellulaceae* retained a positive effect size (LFC =+1.46), but this difference was also not significant after multiple-testing correction (q=0.19).

4 Discussion

4.1 LDA provides an interpretable topic-level representation of microbiome disturbance

LDA provided an interpretable representation of each microbiome as a mixture of microbial components rather than assigning samples to discrete community types. This representation is particularly useful for microbiome data, where samples can contain overlapping bacterial assemblages and differ in the relative contribution of several microbial components. In this study, LDA revealed a comparatively consistent pre-exposure state, more heterogeneous component profiles after 23 days of exposure, differences between adult and juvenile microbiomes, and partial restructuring among long-term survivors.

Combining the LDA representation with complementary statistical analyses allowed these patterns to be evaluated quantitatively. PERMANOVA and beta-dispersion analyses assessed changes in the location and variability of microbial profiles, while component-

specific Kruskal–Wallis tests examined associations between individual microbial components and experimental variables. Together with ASV-level analyses and family-level differential abundance analysis, this provided a multilevel view of the microbiome response, linking topic-level restructuring with changes in community composition, among-individual variability, and specific bacterial families.

4.2 Restructuring and increased dispersion happen at both the topic and ASV level

Microbiome restructuring following exposure was evident at both the topic and ASV levels. At the topic level, pre-exposure microbiomes showed relatively similar microbial-component profiles, characterized by a strong contribution of the broadly distributed MC3 background component. After 23 days of exposure, these profiles shifted and became more heterogeneous, with individuals showing varying contributions from MC1, MC2, MC4, and MC6. Similarly, Bray–Curtis analyses at the ASV level revealed changes in community composition accompanied by increased beta dispersion, particularly among adults after 23 days of exposure. These findings indicate that exposure was associated with both compositional restructuring and increased among-individual microbiome variability.

The within-stage comparison further supported an association between MP exposure and increased microbiome variability. Among adults sampled after 23 days, MP-exposed individuals showed significantly greater beta dispersion than no-plastic controls, despite no significant difference in overall community composition between these groups.

4.3 Increased dispersion supports an Anna Karenina-type response

The AKP predicts that stable microbiomes tend to resemble one another, whereas stressed or disturbed microbiomes become more variable and shift in different directions [7]. In the current study, microbiome variability observed after 23 days of exposure, particularly among adult *Daphnia*, aligns with the pattern described by this principle.

Similar patterns have been reported in other host-associated microbiome systems. In the coral model Aiptasia, long-term heat stress increased bacterial microbiome variability,

supporting an Anna Karenina-like trend under thermal stress [16]. Similarly, in tadpoles, Rana-virus infection was associated with gut bacteriome dysbiosis and increased dispersion, again consistent with the idea that disease or stress can increase variability in host-associated microbiomes [17]. Habitat disturbance has also been shown to alter amphibians' skin microbiome structure, suggesting that environmental stressors can reshape host-associated microbial communities in natural systems [18]. More recently, the AKP has also been extended beyond animal hosts, with pathogen stress in plants leading to microbiome divergence and increased stochasticity in community assembly [19].

One possible explanation for the increased inter-individual microbiome variability is that MP exposure alters host-mediated regulation of microbial community assembly. Under relatively stable conditions, host-associated processes may contribute to maintaining more similar microbial communities among individuals, while both host genotype and surrounding bacterioplankton influence microbial colonization and community composition in *Daphnia magna* [20]. Previous studies have also demonstrated that microplastic exposure can affect gut physiology, oxidative stress responses, and microbial community composition in aquatic organisms, including *Daphnia* [21]. Such physiological disturbances could potentially weaken host-mediated microbial filtering, allowing microbial communities to assemble more variably among exposed individuals. As immune activity, gut physiology, and microbial colonization dynamics were not directly measured in the present study, reduced host filtering should be considered a plausible mechanism rather than a demonstrated cause of the observed increase in microbiome variability.

Importantly, AKP should not be interpreted as a universal response to all stressors; for example, work on wild bank vole gut microbiota showed that environmental disturbance can generate more complex patterns where microbiome variation may depend on season, host context, and disturbance history [22]. Context-dependent microbiome variability has previously been observed in *Daphnia magna*. A previous study showed that host genotypes receiving sympatric microbiomes, originating from the same region as the host, developed more convergent gut microbial communities, whereas allopatric host-microbiome combinations showed greater variation among genotypes [23]. This pattern

was particularly evident when donor microbiomes had previously been exposed to the toxic cyanobacterium *Microcystis aeruginosa* and was interpreted in the context of local host-microbiome adaptation [23]. Although the experimental conditions differ, these findings provide a relevant parallel to the association between inoculum source and microbial-component profiles observed in our study. This suggests that the magnitude and direction of microbiome restructuring may depend on the surrounding microbial context.

4.4 Developmental stage modulates the magnitude of destabilization

Adult and juvenile *Daphnia* microbiomes differed after 23 days of exposure. Although overall community composition differed significantly between the two developmental stages, the clearest contrast was in beta dispersion: adult microbiomes were substantially more variable among individuals than juvenile microbiomes. This difference was also reflected at the topic level. Juveniles retained a relatively strong and consistent contribution of the broadly distributed MC3 background component, whereas adults showed less consistent representation of MC3 and more heterogeneous mixtures of microbial components.

The association between MP exposure and microbiome dispersion also differed between developmental stages. Among adults, MP-exposed individuals showed significantly greater beta dispersion than no-plastic controls, whereas no comparable difference was detected between MP-exposed and control juveniles. Thus, the increased microbiome variability associated with MP exposure was evident in adults but not in juveniles under the conditions examined here.

These findings suggest that developmental stage may influence microbiome variability following MP exposure. However, this interpretation should be made cautiously. Juveniles were sampled at the end of the 23-day experiment and, as offspring produced during the experiment, had shorter individual exposure histories than the founding animals. Therefore, the observed differences between adult and juvenile microbiomes may reflect both developmental stage and variation in exposure duration. Furthermore, the absence of pre-exposure juvenile samples prevents determining whether juvenile microbiomes underwent smaller changes from their initial state or were inherently less variable than

adult microbiomes.

4.5 Restructuring is shaped by experimental context

Although exposure category strongly influenced microbial component composition, it did not explain all of the observed restructuring. MC3 and MC6 were most strongly associated with exposure category, consistent with their contrasting patterns across exposure times: MC3 represented the shared background assemblage that declined after exposure, whereas MC6 became more prominent in exposed individuals. In contrast, MC2 and MC4 were primarily associated with the inoculum pond, suggesting that these components retained a signal of the source bacterial community independently of exposure time. No significant component-level association was detected for host clone, plastic type, or plastic content. Within this experiment, microbial component restructuring was therefore shaped mainly by exposure time and inoculum source, rather than by the specific MP treatment.

The association between inoculum source and microbial component profile highlights the importance of environmental microbial availability in shaping host-associated microbiomes. Different inoculum sources may provide distinct bacterial communities for host colonization, potentially contributing to differences in microbiome composition among individuals. Thus the heterogeneity observed across experimental groups may reflect the combined influence of exposure conditions and surrounding microbial community rather than a uniform response to MP exposure.

4.6 Taxonomic changes suggest restructuring of specific bacterial groups

The taxonomic composition of the LDA-derived microbial components provided further insights into the biological patterns underlying microbiome restructuring. The broadly distributed background component, MC3, included bacterial families such as *Comamonadaceae*, *Flavobacteriaceae*, and *Pseudomonadaceae*, whereas MC4 was characterized by *Planctomycetes*-associated families, including *Pirellulaceae* and *Rubinisphaeraceae*. Family-level differential abundance analysis showed a decline in MC3-associated families, particularly *Pseudomonadaceae* and *Comamonadaceae*, following 23 days of exposure, while *Flavobacteriaceae* exhibited a similar but nonsignificant trend. In contrast, *Rubinis-*

phaeraceae and *Pirellulaceae* increased markedly after exposure. These differences were generally no longer statistically significant in long-term survivors, consistent with partial reconvergence toward the pre-exposure taxonomic composition, although this does not necessarily indicate complete recovery.

Several of the affected families have previously been associated with *Daphnia* microbiomes, host performance, or particle-associated aquatic communities. *Comamonadaceae* is particularly relevant because *Limnohabitans*, a member of this family, are frequently reported as an important *Daphnia*-associated bacterial taxon. Recent characterization of the *Daphnia* core microbiome showed that, despite substantial variability among hosts and experimental conditions, associations with *Limnohabitans* occur consistently, particularly under laboratory conditions [4]. The decline in *Comamonadaceae* after exposure may therefore represent a reduction in an established host-associated group. Experimental evidence that some *Limnohabitans* strains can enhance fecundity in *Daphnia magna* further suggests that changes in this family could have implications for host performance, although functional consequences were not directly examined here [4].

Flavobacteriaceae may also be relevant to host responses under environmental stress. In a study of cyanobacterial stress in *Daphnia magna,* gut microbiome composition differed between tolerant and susceptible host genotypes, and *flavobacteria* were proposed as potential contributors to host tolerance through their possible involvement in the digestion or processing of cyanobacterial cells [24]. Although the decline in *Flavobacteriaceae* observed in our study was not statistically significant after multiple testing correction, its direction may indicate changes in a bacterial group previously implicated in host-microbiome interactions under environmental stress.

In contrast, the enrichment of *Rubinisphaeraceae* and *Pirellulaceae* is also biologically plausible given the ecology of *Planctomycetes*. Members of this phylum commonly occur on submerged surfaces, organic particles, aquatic biofilms, and plastic-associated microbial communities.[25] Microplastics can also provide surfaces for microbial colonization, supporting the formation of plastic-associated biofilms commonly referred to as the plastisphere[26]. The increased representation of these families may therefore reflect

changes in environmentally derived, particle-associated, or biofilm-associated bacteria within the *Daphnia* microbiome. However, because the present analysis did not directly determine whether these bacteria colonized microplastic particles or were acquired from the surrounding bacterioplankton, their enrichment should be interpreted as evidence of taxonomic restructuring rather than direct proof of plastisphere-mediated recruitment.

5 Conclusion and Limitations

This study demonstrates that MP exposure is associated with increased among-individual variability in the adult *Daphnia magna* microbiome, consistent with an Anna Karenina-type response to environmental disturbance. Long-exposure survivors exhibited lower microbiome variability and greater similarity to pre-exposure samples, although their community composition remained significantly different. Microbiome restructuring was also associated with inoculum source, suggesting that the surrounding microbial community may influence host-associated microbiome responses. By integrating LDA topic modelling with ASV-level analyses, this study highlights the value of examining microbiome variability alongside compositional changes to better understand responses to environmental stress.

Several limitations should be considered when interpreting these findings. Microbiomes were sampled from different individuals at each experimental stage, preventing direct assessment of changes within individual hosts. Additionally, juveniles collected at the 23-day sampling point were offspring produced during the experiment and therefore had shorter individual exposure histories, making it difficult to distinguish developmental-stage differences from variation in exposure duration. The exact exposure duration of long-term survivors was not documented, limiting interpretation of their microbiome profiles in relation to exposure time. Their reduced microbiome variability may also reflect selective survival rather than recovery toward the initial microbiome state. Furthermore, although increased beta dispersion is consistent with an Anna Karenina-type response, the underlying mechanisms of microbiome destabilization were not directly examined. Future studies incorporating longitudinal sampling, controlled exposure durations, and host physiological measurements could help clarify these mechanisms.

6 Declarations

Ethics approval and consent to participate

This study involved *Daphnia magna*, an invertebrate crustacean not covered by European Directive 2010/63/EU on the protection of animals used for scientific purposes. Therefore, ethical approval under this directive was not required. Consent to participate was not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by a KU Leuven C1 Project research grant (C16/23/003).

Availability of data and materials

The raw 16S rRNA gene amplicon sequencing reads and associated metadata from the microplastic exposure experiment are publicly available in the NCBI Sequence Read Archive (SRA) under BioProject accession PRJNA1122644 (BioSample accessions SAMN41876284–SAMN41876514). The processed datasets and analysis scripts supporting the findings of this study, including Latent Dirichlet Allocation (LDA) topic modelling, ASV-level community composition analysis, PERMANOVA, PERMDISP, and ANCOM-BC differential abundance analysis, are available in the GitLab repository:`https://gitlab.kuleuven.be/aida-lab/Microplastic-AKP-Paper`. Additional results and supporting figures and tables are provided in the Supplementary Information.

Authors Contributions

PW performed the data analysis, developed the visualizations, interpreted the results, and prepared the original manuscript draft. AG contributed the experimental microbiome data, provided methodological information, and reviewed the manuscript. KF contributed to data interpretation and critically reviewed the manuscript. ED and JA supervised the research, contributed to the interpretation of results, and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

[1] M. Raju, R. Gandhimathi, and P. V Nidheesh. "The cause, fate and effect of microplastics in freshwater ecosystem: Ways to overcome the challenge". *Journal of Water Process Engineering* (2023). DOI: 10.1016/j.jwpe.2023.104199.

[2] A. Krzynowek et al. "Effects of microplastics on Daphnia-associated microbiomes in situ and in vitro". *ISME Journal* 19.1 (2025), wrae234. DOI: 10.1093/ismejo/wrae234.

[3] Vanderville Villegas et al. "Nanoplastic-induced microbiome shifts reduce Daphnia fitness and increase parasite reproduction". *ISME Communications* 6.1 (2026), ycag109. DOI: 10.1093/ismeco/ycag109.

[4] A. Gurung et al. "Investigating the core microbiome concept: Daphnia as a case study". *Environmental Microbiome* 21 (2026), p. 13. DOI: 10.1186/s40793-025-00834-y.

[5] R. Aljaibachi and A Callaghan. "Impact of polystyrene microplastics on Daphnia magna mortality and reproduction in relation to food availability". *PeerJ* (2018). DOI: 10.7717/peerj.4601.

[6] W. Stock et al. “Human impact on symbioses between aquatic organisms and microbes”. *Aquatic Microbial Ecology* (2021). DOI: 10.3354/AME01973.

[7] J. R. Zaneveld, R. McMinds, and R. Vega Thurber. “Stress and stability: Applying the Anna Karenina principle to animal microbiomes”. *Nature Microbiology* 2 (2017), p. 17121. DOI: 10.1038/nmicrobiol.2017.121.

[8] B. J. Callahan et al. “DADA2: High-resolution sample inference from Illumina amplicon data”. *Nature Methods* (2016). DOI: 10.1038/nmeth.3869.

[9] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. “Latent Dirichlet Allocation”. *Journal of Machine Learning Research* 3 (2003), pp. 993–1022.

[10] K. Sankaran and S. P Holmes. “Latent variable modeling for the microbiome”. *Biostatistics* 20.4 (2019), pp. 599–614. DOI: 10.1093/biostatistics/kxy018.

[11] Radim Řehůřek and Petr Sojka. “Software Framework for Topic Modelling with Large Corpora”. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. 2010, pp. 45–50.

[12] M. Röder, A. Both, and A. Hinneburg. “Exploring the space of topic coherence measures”. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. Association for Computing Machinery, 2015, pp. 399–408. DOI: 10.1145/2684822.2685324.

[13] D. M. Endres and J. E Schindelin. “A new metric for probability distributions”. *IEEE Transactions on Information Theory* (2003). DOI: 10.1109/TIT.2003.813506.

[14] J. Roger Bray and John T. Curtis. “An Ordination of the Upland Forest Communities of Southern Wisconsin”. *Ecological Monographs* 27.4 (1957), pp. 325–349. DOI: 10.2307/1942268.

[15] Huijuan Lin and Shyamal D. Peddada. “Analysis of compositions of microbiomes with bias correction”. *Nature Communications* 11.1 (2020), p. 3514. DOI: 10.1038/s41467-020-17041-7.

[16] H. I. Ahmed et al. “Long-Term Temperature Stress in the Coral Model Aiptasia Supports the Anna Karenina Principle for Bacterial Microbiomes”. *Frontiers in Microbiology* 10 (2019), p. 975. DOI: 10.3389/fmicb.2019.00975.

[17] L. S. Billet, O. Hernández-Gómez, and D. K. Skelly. “Ranavirus Epizootics and Gut Bacteriome Dysbiosis in Tadpoles: Evidence for the Anna Karenina Principle?” *Molecular Ecology* 34.16 (2025), e70022. DOI: 10.1111/mec.70022.

[18] R. R. Jiménez et al. “Habitat disturbance influences the skin microbiome of a rediscovered neotropical-montane frog”. *BMC Microbiology* (2020). DOI: 10.1186/s12866-020-01979-1.

[19] D. Li et al. “The Anna Karenina principle in the assembly of plant microbiome under pathogen stress”. *Npj Biofilms and Microbiomes* (2026). DOI: 10.1038/s41522-026-00964-2.

[20] M. Callens et al. “The bacterioplankton community composition and a host genotype dependent occurrence of taxa shape the Daphnia magna gut bacterial community”. *FEMS Microbiology Ecology* (2020). DOI: 10.1093/femsec/fiaa128.

[21] Ve Van Le et al. “Shape-driven toxicity of polystyrene microplastics: Impacts on physiology and gut microbiota in Daphnia magna”. *Marine Pollution Bulletin* 225 (2026), p. 119218. DOI: 10.1016/j.marpolbul.2026.119218.

[22] A. Lavrinienko et al. “Applying the Anna Karenina principle for wild animal gut microbiota: Temporal stability of the bank vole gut microbiota in a disturbed environment”. *Journal of Animal Ecology* (2020). DOI: 10.1111/1365-2656.13342.

[23] S. Houwenhuyse et al. “Locally adapted gut microbiomes mediate host stress tolerance”. *ISME Journal* (2021). DOI: 10.1038/s41396-021-00940-y.

[24] E. Macke et al. “Host-genotype dependent gut microbiota drives zooplankton tolerance to toxic cyanobacteria”. *Nature Communications* (2017). DOI: 10.1038/s41467-017-01714-x.

[25] M. M. Bengtsson and L Øvreås. "Planctomycetes dominate biofilms on surfaces of the kelp Laminaria hyperborea". *BMC Microbiology* (2010). DOI: 10.1186/1471-2180-10-261.

[26] L. A. Amaral-Zettler, E. R. Zettler, and T. J Mincer. "Ecology of the plastisphere". *Nature Reviews Microbiology* (2020). DOI: 10.1038/s41579-019-0308-0.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile2.xlsx
- SupplementaryFile1.docx