**Title:** Geography, taxonomy, and ecological guild: factors impacting freshwater macroinvertebrate gut microbiomes.

Paul A. Ayayee1, Jeff S. Wesner2, Scot P. Ouellette3

1 Department of Biology, University of Nebraska at Omaha, Omaha, NE, USA.

2 Department of Biology, University of South Dakota, Vermillion, SD, USA

3 Department of Pathology and Microbiology, College of Medicine, University of Nebraska Medical Center, Omaha, NE, USA.

Corresponding author: Paul Ayayee. Email: [payayee@unomaha.edu](mailto:payayee@unomaha.edu)

**Abstract**

Despite their diversity, global distribution, and clear effects on host biology the rules of life that govern variation in microbiomes among host species are unclear, particularly in freshwater organisms. In this study, we sought to asses whether geographic location, taxonomy (Order, Family, and Genus), or functional feeding group (FFG) designations would best explain differences in the gut microbiome composition among macroinvertebrates sampled across ten National Ecological Observatory Network's (NEON) freshwater stream sites in the United States. Subsequently, we compared beta diversity of microbiomes among locations, taxonomy (Order, Family, and Genus), and FFGs in a single statistical model, to account for variation within the source microbial community and the types of macroinvertebrates sampled across locations. We determined significant differences in community composition among macroinvertebrate orders, families, genera, and FFGs. Differences in microbiome compositions were underscored by different bacterial ASVs that were differentially abundant among variables (four bacterial ASVs across the 10 NEON sites, 43 ASVs among the macroinvertebrate orders, and 18 bacterial ASVs differing among the five FFGs). Analyses of variations in microbiome composition using the Bray-Curtis distance matric revealed FFGs as the dominant source of variation (mean standard deviation of 0.8), followed by stream site (mean standard deviation of 0.5), and finally family and genus (mean standard deviation of 0.3 each). This study is the most comprehensive assessment of freshwater macroinvertebrate gut microbiomes and it is notable in it’s broad scope; both in terms of number of streams sampled (across 10 distinct geographic locations), and the number of sites within streams sampled ( at least two sites per stream), and the variables assessed to be impacting associated gut microbiomes of freshwater macroinvertebrates

**Keywords**: Freshwater ecosystems, Gut microbiome, Macroinvertebrates, feeding groups

**Introduction**

A sizeable portion of earth’s diversity occurs in the microbiomes, the micro-organisms that live within host species (Hug et al., 2016; Louca et al., 2019). Despite their diversity, global distribution, and clear effects on host biology (Werren et al. 1997), the rules of life that govern variation in microbiomes among host species are unclear, particularly in freshwater organisms (Charlesworth et al. 2019). In studies of freshwater invertebrate microbiomes, there are conflicting results comparing the importance of habitat, taxonomy (Family and Genus), and functional feeding group (FFG) categorization in structuring gut microbial assemblages of freshwater macroinvertebrates. One study determined taxonomy of the assessed macroinvertebrates to be a more relevant determinant of gut microbiome composition in freshwater aquatic than local habitat, stream conditions, or FFG (Kroetsch et al., 2020). In contrast, the vast majority of studies have determined FFGs (regardless of taxonomic affiliation of samples invertebrates) to be the principal variable explaining differences in microbiome composition across freshwater macroinvertebrates (Kaufman et al., 2000; Pechal and Benbow, 2016; Ayayee et al., 2018; Receveur et al., 2020). Furthermore, a study of macroinvertebrate gut microbiome samples collected from two different streams in the same region did not uncover any stream-specific effects on gut microbiome composition (Ayayee et al., 2018). Similarly, another study of macroinvertebrate gut microbiome samples from multiples sites along the reach of one major river did not uncover any significant site-specific effects on associated gut microbiomes (Kroetsch et al., 2020).

In studies of terrestrial insect gut microbiomes, stochastic and deterministic (at opposite ends of a spectrum) are important drivers of gut microbial assembly (Hanson et al., 2012; Jizhong and Daliang, 2017). Stochastic processes (or ecological processes) such as priority effects, dispersal limitation, and ecological drift have been determined to be responsible for variations in microbial community composition in a large number of studies (Jizhong and Daliang, 2017). Similarly, various deterministic factors (or evolutionary processes) have also been uncovered to account for considerable variations in microbial community composition (free-living and host-associated). Deterministic processes mainly involve variation in environmental parameters (e.g., pH, salinity, DO, etc.,), local habitat conditions, nutrient availability, and, ultimately, co-evolution between host and associated gut microbiomes (phylosymbiosis) (Jizhong and Daliang, 2017). These insights have generated various conceptual frameworks with which to assess dynamics governing community assembly (Vellend, 2010) and, by extension, gut microbial community assembly in insect hosts (Brown et al., 2020).

In this study, we sought to asses whether geographic location, taxonomy (Order, Family, and Genus), or FFG designations would best explain differences in gut microbiome composition among sampled macroinvertebrates across ten streams in North America. A major limitation of prior studies characterizing the gut microbiomes of freshwater macroinvertebrates has been the systematic sampling of freshwater ecosystems at a broad scale. Most of these studies only sampled multiple sites within single streams or sampled limited (1-3) streams within one geographic region. The novel aspect of our study is that it is the first to sample multiple streams and multiple sites within the same streams at a large geographical scale. We assessed the gut microbiomes of insects collected from ten National Ecological Observatory Network's (NEON) freshwater stream sites. Subsequently, in a single statistical model, we compared beta diversity of microbiomes among locations, taxonomy (Order, Family, and Genus), and FFGs . This allowed us to account for variation within the source microbial community and the types of macroinvertebrates sampled across locations. These are expected to vary due to the underlying geology, stream conditions, and land usage (Wakelin et al., 2008; Drury et al., 2013; Atashgahi et al., 2015; Medeiros et al., 2016; Fang et al., 2017; Hosen et al., 2017; Sujay. et al., 2018). In addition, functional feeding groups are known to have differences in gut physiology (Cummins, 1979; Martin et al., 1980, 1981b; Anderson, N.H. and Cargill, 1987; Austin and Baker, 1988; Tierno de Figueroa et al., 2011), associated gut microbial composition (Pechal and Benbow, 2016; Ayayee et al., 2018; Receveur et al., 2020), and functions (Stief et al., 2009; Stief, 2013).

We hypothesized that differences in gut microbial community composition (β-diversity) among sampled macroinvertebrates would be a function of deterministic factors (reflecting FFG designations) rather than stochastic factors (geographical location or taxonomical identification). This is based on the assumption that different macroinvertebrate taxa from different sites will have comparable gut physiologies, which will select for taxonomically comparable microbes, thus, resulting in similar gut microbiomes.

**Materials and Methods**

*Sample Collection*

Aquatic insect samples were obtained from collections acquired by the National Ecology Observatory Network (NEON) sites in the United States in 2020 (<https://www.neonscience.org/field-sites/about-field-sites>). We obtained macroinvertebrate samples collected from 10 NEON-managed field sites across North America that ranged across 11 degrees of latitude with mean annual air temperatures from 0.2 (BLDE) to 16°C (BLUE) (Fig. 1). We obtained samples from at least two sites within each of the ten NEON streams. Due to permit restrictions, we limited sample acquisition to five individuals per taxa per site. Macroinvertebrate samples were identified and verified to both species and family taxonomic levels according to NEON protocols. The insect samples were stored in 1.5 ml tubes in 80 % ethanol and kept at -20 ℃ until DNA extraction. Upon receipt, insect samples were categorized into functional feeding groups using a combination of published (Merritt et al., 2008; Cummins, 2021) and online resources ( <https://www.macroinvertebrates.org/> ).

*Sample processing, DNA extraction, and Illumina sequencing*

Before DNA extraction, we surface-sterilized insect samples by washing them in a 1% detergent solution for 1 minute, followed by two one-minute rinses in DI water. If samples were large enough, then we dissected the entire dietary system; if not, then we used the whole insect. Next, we performed DNA extraction using the DNeasy Blood & Tissue Kit (Qiagen, Germantown, MD, USA) with modifications to the manufacturer's directions. We verified the presence of microbial 16S rRNA marker gene in all extracted DNA samples via PCR using the universal 27F and 1492 R bacterial primer pair (Frank et al., 2008). Samples were submitted for high-throughput paired-end MiSeq library preparation and sequencing at the University of Nebraska Medical Center Genomics Core. Briefly, a limited cycle PCR reaction was performed on each sample to create a single amplicon, including the V4 (515-F) and V5 (907-R) variable region (Keskitalo et al., 2017). The resulting libraries were validated using the Agilent BioAnalyzer 2100 DNA 1000 chip (Agilent, Santa Clara, CA, USA), and DNA was quantified using Qubit 3.0 (QubitTM, Thermofisher, Waltham, MA, USA). A pool of the libraries was loaded into the Illumina MiSeq at 10 pM. The pool was spiked with 25 % PhiX (a bacteriophage) at 10 pM for MiSeq run quality as an internal control (Mukherjee et al., 2015) to generate 300 bp paired ends with the 600 cycle kit (version 3). The raw reads were deposited into the Sequence Read Archive database (Accession number: PRJNA825559).

*Microbiome Data Processing and statistical analyses*

Acquired fastq primer-trimmed MiSeq paired-end reads from the sequencing center were processed using DADA2 (Callahan et al., 2016). Across both forward and reverse reads, filtering excluded reads with more than two expected erroneous base calls, any reads identified as part of the PhiX bacteriophage genome for quality control, and reads less than 175 base pairs. Forward reads were truncated to 250 base pairs, and reverse reads to 200 base pairs. Truncation was performed to maintain median quality scores above 30 across samples. Reads were merged, and chimeras subsequently filtered out. We determined amplicon sequence variants (ASVs) and representative sequences against the SILVA 138.1 16S rRNA gene reference database (Quast et al., 2012). We combined the count and taxonomy information for the generated ASVs into a classical OTU/ASV table, and further analyses were performed in QIIME v.1.8 (Caporaso et al., 2010; Kuczynski et al., 2012). Before analyses, we curated the table by removing any reads unclassified at the bacterial or archaeal domain level and any reads assigned as Eukaryotes. Finally, samples with less than 1,000 reads per sample were removed from the table before analyses. We then summarized the filtered and curated ASV table to the family level, and all subsequent analyses were performed on this table.

Briefly, we rarefied the family-level table to 1,110 reads per sample and replicated ten times across all samples. The rationale and justification for rarefying have been discussed in prior studies (Weiss et al., 2017; McKnight et al., 2019; Cameron et al., 2021). For alpha diversity, the chao1 (Huang and Zhang, 2013), Shannon's evenness (Shannon C.E, 1957), and observed\_OTUs indices were calculated in QIIME, and significant differences among categorical groupings determined via non-parametric Wilcoxon tests in JMP Pro 15 (S.A.S., Cary, NC, USA). We generated the Bray-Curtis dissimilarity distance matrix (Bray and Curtis, 1957) using the 1,110-rarefied table. The calculated distance matrix was used to calculate the non-metric multidimensional scales (NMDS) in QIIME. The NMDS scales are used to visualize categorical sample groupings that differ in microbiome composition following a test of differences among these categorical groupings via permutational multivariate analysis of variance (PERMANOVA) (Anderson, 2017) in QIIME using the compare\_categories.py command.

*Modelling of factors and microbiome composition*

We compared the influence of freshwater insect taxonomy (orders, family, and genus), functional feeding groups, and geographical (i.e., NEON) sites on beta diversity of associated gut microbiomes. To do this, we used intercept-only random effects models with the response variable (Bray-Curtis distance) as a measure of diversity and a Beta likelihood with a logit link. Each model contained a response (grand mean) and three varying intercepts [insect order (family and genus nested within this), functional feeding group, and NEON site], also known as random effects (McElreath Richard, 2020). The structure of the model is as follows:

Where is the Bray-Curtis value of the *i*th observation arising from a Beta distribution with parameters alpha = and beta = . In this parameterization, the mean is the primary target of inference and is represented by , while is a scalar. Priors for each parameter were chosen using prior predictive simulation (Wesner and Pomeranz, 2021) and are justified in the Supplemental Information.

There are several advantages to this modeling approach (Dietze, 2017). First, the varying intercepts use partial pooling to pull each group mean (i.e., mean of individual orders or FFG’s or sites, etc.) toward the global mean. That provides conservative estimates of Bray-Curtis values for each group because the amount of pooling is determined by the amount of data in each group. In other words, it provides a correction for outliers such that groups with few data points are treated skeptically and pulled more strongly towards the overall mean (Efron and Morris, 1977). This is especially important for data like ours with relatively low replication within each group, helping to prevent spurious conclusions. Second, the residual variance of the grand mean of Bray-Curtis dissimilarity is partitioned among the varying intercepts. This allows us to identify the variables that contribute most of the variation in beta diversity, and these in turn are likely to be the best variables to focus on in future studies. Third, from these models, we can predict the diversity of the gut microbiomes at each level (e.g., for each species or each stream or each functional feeding group). Fourth, the varying intercepts allow predictions, with proper uncertainty, of the microbiomes of insects that are not currently in our dataset (McElreath Richard, 2020). Finally, the use of varying intercepts automatically adjusts for unbalanced data so that no single sample dominates the inference.

We fit the model using Bayesian inference in R version 4.2.0 (R Core Team, 2022), with the *brms* package (Bürkner, 2017). Posteriors were explored in *rstan* (Stan Development Team, 2022) with Hamiltonion Monte Carlo (No-U-Turn sampler). We used 4 chains with 2,000 iterations each and the first 1,000 discarded as warmup. Model convergence was checked by ensuring that all R-hats were <1.1 and by visually assessing the chains for mixing.

**Results**

*Macroinvertebrate summary*

We obtained 45 macroinvertebrate samples from 10 NEON sites across the continental USA, with each sample representing 1-5 individuals from a single genus or family. After removing four samples that did not meet sequence quality thresholds, our final dataset consisted of 41 samples (Table 1). Taxonomically, the 41 macroinvertebrate samples came from seven orders (Fig. 2B), 26 families (Fig. 2C), and 36 genera across all 10 NEON sites. The major orders in the dataset were Ephemeroptera (n = 11 samples), Diptera (n = 9 samples), and Plecoptera (n = 8 samples) (Fig. 2B). The major families were Heptageniidae(n = 4 samples) and Chloroperlidae(n = 4 samples), followed by Elmidae, Hydropsychidae, and Tipulidae (all n = 3 samples) (Fig.2 C). We further classified the 41 macroinvertebrate samples into four FFGs: filtering collectors, gathering collectors, predators, and shredder/detritivores (Fig. 2D). The taxonomic and ecological information of the 41 macroinvertebrates samples used in this study are provided (Table 1).

*Freshwater macroinverterbate gut microbiome diversity and composition*

Overall, we obtained ~ 6.5 million reads from the sequencing effort. Filtering, merging, chimera removal, and further curating of the resulting combined count and taxonomy table (removal of unassigned reads at the domain level and removal of samples with fewer than 1,000 reads per sample) resulted in ~ 86% of reads retained (5,684,379 reads) distributed across 41 samples yielding 12,658 ASVs (mean reads per sample = 19,468; Minimum: 1235.000, Maximum: 98,479.000). Rarefaction curves for the species diversity and richness indices at ~1,110 reads per sample indicate that most microbial diversity had been sufficiently covered across samples (Fig. S1). We examined microbial alpha diversity across geographical locations, among taxa (order, families, and genera), and among functional feeding groups. In general, there were no significant differences among locations, taxonomy (order, family, genera), or FFGs for all four diversity indices evaluated (Table 2). Further investigation of microbial community composition (β-diversity) among macroinvertebrate samples was dependent on level of inquiry.

Overall, variances did not differ significantly among samples for any of the variables examined at the level of locations (PERMDISP: F = 999, P = 0.314), order (PERMDISP: F = 999, P = 0.086), family (PERMDISP: F = 999, P = 0.308), genus (PERMDISP: F = 999, P = 0.46), and FFGs (PERMDISP: F = 999, P = 0.126). Across locations, there were no significant differences among the 10 NEON sites (PERMANOVA: test statistic = 1.13, P = 0.18). However, there were significant differences in community composition among macroinvertebrate orders (PERMANOVA: test statistic = 1.74, P < 0.001) (Fig. 3A), families (PERMANOVA: test statistic = 1.48, P < 0.001) (Fig. 3B), and genera (PERMANOVA: test statistic = 2.17, P = 0.004) (Fig. 3C). Finally, there was significant difference in microbiome composition among ecological functional feeding group classifications (PERMANOVA: test statistic = 1.35, P = 0.008) (Fig. 3D).

Examination of ASVs that differed significantly (FDR-adjusted P = 0.05) in abundance among the variables yielded four bacterial ASVs across the 10 NEON sites. For macroinvertebrate orders it was one ASV at FDR-adjusted P = 0.05 and 43 bacterial ASVs at P = 0.05 (Fig. 4A). There were 18 bacterial ASVs differing among the five FFGs (FDR-adjusted P = 0.05)(Fig. 4B). For locations all four significantly abundant ASVs were only detected in one NEON site (LEWI) at the FDR-adjusted P = 0.05. These were ASVs classified to the families Bacteroidetes vadinHA17 (32%), Bacteroidetes BD2-2 (28%), *Defluviicoccaceae* (28 %), and *Geobacteraceae* (12 %). Among macroinvertebrate orders, *Anaplasmataceae* (Genus *Wolbachia*), (abundant across five macroinvertebrate orders except Diptera and Megaloptera) and Unassigned Enteobacterales (abundant across six macroinvertebrate orders, except Odonata) were the most abundant ASVs across macroinvertebrate orders. *Anaplasmataceae* (Genus *Wolbachia*) was the most abundant in Coleoptera (70.28 %), followed by Odonata (20.08 %) and Trichoptera (17.25 %) (Fig. 4A). Clustering of Odonata and Coleoptera is underscored by the abundance of *Anaplasmataceae* (Genus *Wolbachia*) in these orders. Odonata is further separated from Coleoptera due to the abundance of *Nostocaceae* (71.1 %)and Unassigned vadinHA49 (4.83 %) in this order, and the presence of several other ASVs only significantly abundant in Coleoptera (Fig. 4A). The clustering of the Megalopteran samples away from the other macroinvertebrate orders is underscored by the dominance of Unassigned Enterobacterales (94.4 %), whereas the clustering of the Plecopteran samples from the other macrcoinvertebrate orders is underscored by the dominance of both Unassigned Enterobacterales (87.1 %) and *Anaplasmataceae* (8.2 %)(Fig. 4A). The comparatively higher abundances of *Lachnospiraceae* (24.2 %), *Rhodocyclaceae* (34.4 %), *Spirosomaceae* (11.57 %), *and* Unassigned enterobacterales (10.64 %) separates the Ephemeroptera from the Diptera and other orders. Among functional feeding groups, the separate clustering of the shredder/detritivore group from the other four functional feeding groups (Fig. 3D) is underscored by the preponderance of bacterial ASVs mainly in the family *Candidatus Hepatincola* (order Rickettsiales)(63.5 %), as well as more diverse representation of ASVs that differed in abundances in this functional feeding group (Fig. 4 B). The clustering of the filtering collectors and predator (Fig. 3D) is underscored by the abundances of Nitrosomonadaceae (26.19 %), UnassignedB1-j (21.43 %), Unassigned MA-A2-108 (19.04 %), Unassigned MBNT15 (11.9 %), and Unassigned RCP2-54 (7.1 %), as well as Unassiged DS-100, Unassigned Acidobacteriae, and *Polyangiaceae* (all 2.38 %) predominantly in predators. Filtering collectors did not have any bacterial ASVs that were significantly abundant across functional feeding groups.The clustering together of gathering collectors and scrapers separate from the other functional feeding groups (Fig. 3D) can be attributed to the relatively higher proportional abundances of Unassigned PLTA13 (10.87 % in gathering collectors and 50 % in scrapers) in both groups (Fig. 4B); Gathering collectors are further characterized by the abundances of *Nitrosomonadaceae* (47.8 %), Unassigned RCP2-54 (26.08 %), Unassigned MBNT15 and *Xanthobaceraceae (both 4.38 %)*, and Unassigned Thermolephilia (2.17 %), whereas the scrapers are characterized by the abundance of Unassigned MA-A2-108 (50%)(Fig. 4B).

An unexpected, but, intriguing result from our microbiome analyses was the abundant detection of the endosymbiont bacterium, *Wolbachia*, (family *Anaplasmataceae*) in analyzed freshwater macroinvertebrate orders (Fig. 4A). *Wolbachia* was most abundant in aquatic Coleoptera (70.28 %), followed by Odonata (20.08 %), Trichoptera (17.28 %), Plecoptera (8.20 %), and Ephemeroptera (0.15 %). This bacterial ASV was not detected in aquatic Diptera and Megloptera macroinvertebrate orders the current microbiome dataset. A subsequent PCR-based investigation of the presence of *Wolbachia* in the macroinvertebrate samples using *wsp* primers with modifications to reported conditions (Sazama et al., 2017), resulted in six positive detections and 41 negatives (Table 3). The six positive results are distributed across four macroinvertebrate orders, Coleoptera (Family *Elmidae*), Odonata (Family *Gomphidae* and *Cordulegastridae*), Trichoptera (Family *Glossosomatidae*), and Plecoptera (Family *Chloroperlidae*) (Table 3).

*Assessment of variables shaping gut micrbiomes*

Among all samples, beta diversity (in units of Bray-Curtis distance) averaged ~0.7 ± 0.09 (posterior mean ± sd) (Fig. 5A). On average, most of the residual variation in beta diversity was attributable to variation among FFGs, which had a mean standard deviation of 0.8 compared to 0.5 for stream site and ~0.3 for both genus and family (Fig. 5A). After correcting for the logit link, these values indicates that, relative to the average Bray-Curtis value of 0.7, a typical functional feeding group deviatoin is ~0.15 units higher or lower (i.e., 0.55 to 0.85), which is nearly double the deviation for a typical genera or family (± 0.8 units) and slightly higher than a typical deviation for a given site (± 0.8). However, there is considerable uncertainty in these estimates (Fig. 5B), such that the probability that functional feeding group deviations are higher than any of the other categories ranges from only 0.65 to 0.85 with (0.5 indicating equivalence and 1 indicating absolute certainty). This indicates that no single factor dominates as a source of variation in explaining microbiome beta-diversity in aquatic insects. Among families, beta-diversity ranged from a mean of 0.5 (95% Credible Interval: 0.3 to 0.8) in Leptophlebiidae (represented by a single genus, *Paraleptophlebia*) to 0.9 (0.69 to 0.97) in Perlidae (represented by *Acroneuria* and an uknown genus) (Fig. 5A). Among FFGs, beta-diversity ranged from 0.7 (0.2 to 0.9) in collector/filterers to 0.8 in predators (0.6 to 0.9) and among sites it ranged from 0.6 (0.3 to 0.8) at POSE to 0.8 (0.6 to 0.95) at BLDE (Fig. 5A).

**Discussions**

The nature and dynamics of insect-gut microbial associations are well established and understood for terrestrial insects (Dillon and Dillon, 2004; Yun et al., 2014; Douglas, 2015). For instance, there is a consensus on the impacts of diet, developmental stage, and environment on gut microbiomes of terrestrial insects. Notably, the importance of diet and internal digestive tract conditions, such as pH and dissolved oxygen content of the gut, in shaping both the composition and ultimately function of associated gut microbial assemblages in terrestrial insects is evident from the variety of such studies in the last two decades. This is evidenced in diverse insects, such as wood-, soil-, and humus-feeding termites (Raychoudhury et al., 2013; Ayayee et al., 2015; Rossmassler et al., 2015), wood-feeding and fungus-feeding termites (Su et al., 2016), cockroaches fed high-cellulose or low-cellulose diets (Bertino-Grimaldi et al., 2013; Carrasco et al., 2015; Ayayee et al., 2016), dung- and detritivore-feeding beetles (Franzini et al., 2016), and wood-feeding beetles (Ayayee et al., 2014; Mason et al., 2016). However, unlike terrestrial insects, where factors like diet, environment, and developmental stage are recognized as essential drivers of gut microbiomes, such consensus is lacking for aquatic insects and their associated gut microbiomes. The gut microbiome of freshwater macroinvertebrates has been noted to differ from the surrounding environment under controlled laboratory studies (Ma et al., 2020) and field-collected samples (Pechal and Benbow, 2016; Ayayee et al., 2018; Receveur et al., 2020). However, it remains to de determined how generalizable this is.

In our study, the 41 macroinvertebrates analyzed were categorized into five functional feeding groups, FFGs (filtering collectors, gathering collectors, shredders/detritivores, scrapers, and predators). Despite there not being any significant differences in alpha diversity among them, FFG was an important predictor of both beta diversity and microbiome community composition, presumably driven by the differential abundances of 18 ASVs (Figs. 4B and 5B). These results suggest a clear rationale for more broadly studying how microbiomes of aquatic insects are affected by FFGs particularly because the mechanistic basis is well-developed. For example, freshwater macroinvertebrate functional feeding group (FFG) categorization is based on behavioral mechanisms of food acquisition and the type of materials consumed instead of taxonomic designation of macroinvertebrates (Gökçe, 2018; Cummins, 2021). This approach allows for classifying hundreds of macroinvertebrate taxa into relevant ecological units based on how they function and acquire food in aquatic ecosystems. There are significant physiological differences among various macroinvertebrate functional feeding groups, further making them critical physiological units that are useful beyond the ease of classification. Differences among FFGs are most obvious in their distinct digestive physiologies; a rationale proposed for assessing gut microbiomes of freshwater macroinvertebrates almost two decages ago (Harris, 1993). For example, filter feeders, such as *Hydropsychidae* (order Trichoptera) are characterized by slightly acidic to neutral gut pH (Martin et al., 1981a; Anderson, N.H. and Cargill, 1987) or slightly alkaline gut pH and hindgut enlargements in the case of Chironomidae (order Diptera) (Cummins, 1979; Martin et al., 1980). Grazers/collectors, such as *Baetidae* and *Leptophlebidae* (order Ephemeroptera), are morphologically adapted to feed on biofilm and have neutral to slightly alkaline gut pHs (Austin and Baker, 1988). Finally, predatory freshwater macroinvertebrates tend to have comparatively more alkaline gut pHs than other functional feeding groups (Anderson, N.H. and Cargill, 1987; Tierno de Figueroa et al., 2011). These different gut conditions may be driving differences in associated gut microbiomes among freshwater macroinvertebrate functional feeding groups in this study.

An additional reason to understand mechanisms that govern the microbiome in freshwater insects is that microbiomes may affect ecosystem-level processes like nutrient cycling. For example, differences in gut microenvironmental conditions among FFGs have been determined to underscore differences in microbial functions in these groups. Well-characterized differences in microbe-mediated nitrous oxide (N2O) emission rates among different freshwater macroinvertebrate functional feeding groups have been documented, with filter-feeders emitting the highest N2O, followed by shredders, grazers, and predators (Stief et al., 2009; Stief, 2013). Similarly, potential gut microbial nitrogen provisioning via nitrate reduction to ammonium has also been documented among freshwater macroinvertebrate functional feeding groups (Ayayee et al., 2019). It is reasonable to assume these different microbe-mediated functions (N2O emission rates and nitrate reduction to ammonium) among functional feeding groups could be underscored by different gut microbiome compositions. However, this was not explicitly investigated in the mentioned studies. The functional feeding group classification is not without its drawbacks. As with any categorical classification, there can be substantial variation among individuals within a category. Most studies with aquatic insects can confidently identify taxa down to the order or family level. However, at these levels, there may be multiple functional feeding groups within taxons, in addition to ontogenetic variation, and this can further complicate assessment of gut microbiota. Overall, it stands to reason that FFGs with their inherent differences in gut physiology and digestive requirements (despite some overlap in food consumed) can serve as a good delineator of freshwater macroinvertebrates gut microbiomes.

While the possible mechanistic basis for microbiome variation among FFGs is well-developed and supported by our data, as well as other studies (Ayayee et al., 2018; Receveur et al., 2020), there is also substantial variation among taxa and sites, but the mechanistic basis for this variation is less understood. Given the broad environmental conditions that our sites represent (e.g., ranging 11 degrees of latitude and 16 °C in mean annual temperature), it seems likely that there is also large variation in the source pool of microbiota among sites. One particularly interesting result is the presence of four bacterial ASVs in high abundances in LEWI and not in the other nine locations. Lewis Run (LEWI) is an aquatic NEON field site located about 60 miles west of Washington, D.C. in Clarke County, Virginia. The site is a small wadeable stream that drains a 11.9 km2 (2,940 acre) watershed. The majority of the stream reach flows past and through land managed by Casey Trees, a nonprofit organization that raises trees for planting in and around the Washington, D.C. area. The surrounding region is characterized by general land use types including successional fields, pastures, woodlands, and small ponds. This site is located within NEON's Mid-Atlantic Domain (D02), a densely populated region bounded by the Atlantic Ocean on the east and stretching down the Eastern Seaboard from southern New Jersey to northern Georgia (<https://www.neonscience.org/field-sites/lewi>). It is unclear how impacted or unimpacted this site is relative to the other nine sites. Several studies have previously reported differences in environmental microbiome diversity and composition among different streams varying in quality (Atashgahi et al., 2015; Medeiros et al., 2016) and among sites along streams (Wakelin et al., 2008; Drury et al., 2013; Kroetsch et al., 2020), as well as reported differences between bacterioplankton and sediment partitions within streams (Fang et al., 2017; Hosen et al., 2017; Ayayee et al., 2018).

The detection of *Wolbachia* in five of the freshwater macroinvertebrate orders in this study (Coleoptera, Odonata, Trichoptera, Plecoptera, and Ephemeroptera) supports previous work that concluded that *Wolbachia* can be possibly considered a common endosymbiont of aquatic insects, with an incidence rate of 52 % (Sazama et al., 2017). The detection of *Wolbachia* in Ephemeroptera, in this study, as in other studies (Sazama et al., 2017), was low compared to the other orders it was found in (low abundance from the microbiome data and no detection from the *wsp*-PCR study), whereas, Plecoptera and Trichoptera were well represented in both approaches. However, in contrast to previous studies, we did not detect *Wolbachia* from Diptera via PCR nor in Diptera from the microbiome dataset.The main Dipteran families in this study were *Chironomidae, Dixidae, Psychodidae, Simulidae,* and *Tipulidae*. We did not have any members of the Dipteran *Culicidae* family (known to harbor several pathogens of public health importance) in our dataset. There is a need to examine further the presence of *Wolbachia* across aquatic Dipteran taxa that are of ecological importance and not merely medical importance.

In conclusion, our study provides data in support of the existence of intrinsic processes that screen microbes from the surrounding source microbiota pool in streams prior to colonization and establishment in freshwater macroinvertebrate guts, predominantly underscored by the ecological classification of macroinvertebrates based on the mode of feeding i.e., FFGs in these streams. These processes may be different among FFGs due to the differences in gut physiologies as already mentioned, resulting in different gut microbiomes. Next, we determined that the sampled streams, themselves, also contributed to the observed variations in gut microbiome composition of the smapled macroinvertebrates. Differences in site characteristics (geopgraphic location of stream, stream condition, pH, salinity, etc.,) may be impacting aquatic macroinvertebrate gut microbiomes, mainly via influencing the source microbiota pool in the water column (bacterioplankton) or the biofilm ( on leaves, twigs, rocks, debris, etc.,) in these streams. Finally, we determined that the taxonomy (order, family, and genus) of the sampled macroinvertebrates in this study, accounted for the lowest variation in observed microbiome community composition. Overall, results suggest that future studies characterizing freshwater macroinvertebrate gut mcirobiomes would be best served by focusing on sampling representatives of multiple functional feeding groups within the same streams, and also from

multiple streams, if possible.

**Acknowledgments**

During the writing of this manuscript, JSW was partially supported by NSF Grant #2106067. PAA was supported by start-up funds from the University of Nebraska at Omaha. SQ……………………………………………………………….. The University of Nebraska DNA Sequencing Core, at the Unibversity of Nebraska Medical Center receives partial support from the National Institute for General Medical Science (NIGMS) INBRE - P20GM103427-19 grant and The Fred & Pamela Buffett Cancer Center Support Grant - P30 CA036727. This publication's contents are the authors' sole responsibility and do not necessarily represent the official views of the NIH or NIGMS.

**References**

Anderson, N.H. and Cargill, A. S. (1987). “Nutritional ecology of aquatic detritivorous insects.,” in *Nutritional Ecology of Insects, Mites, Spiders and Related Invertebrates*, ed. IF. Slansky Jr. & J.G. Rodriguez (eds.) (Jon Wiley & Sons. New York, NY), 902–925.

Anderson, M. J. (2017). Permutational Multivariate Analysis of Variance (PERMANOVA). *Wiley StatsRef Stat. Ref. Online*. doi:doi:10.1002/9781118445112.stat07841.

Atashgahi, S., Aydin, R., Dimitrov, M. R., Sipkema, D., Hamonts, K., Lahti, L., et al. (2015). Impact of a wastewater treatment plant on microbial community composition and function in a hyporheic zone of a eutrophic river. *Sci. Rep.* 5, 17284. Available at: http://dx.doi.org/10.1038/srep17284.

Austin, D. A., and Baker, J. H. (1988). Fate of bacteria ingested by larvae of the freshwater mayfly, <Emphasis Type="Italic">Ephemera danica </Emphasis>. *Microb. Ecol.* 15, 1–10. Available at: papers2://publication/uuid/7D8AF746-F9FF-4CD9-82CD-7973F28D294D.

Ayayee, P. A., Cosgrove, C. R., Beckwith, A., Roberto, A. A., and Leff, L. G. (2018). Gut bacterial assemblages of freshwater macroinvertebrate functional feeding groups. *Hydrobiologia* 822, 157–172. doi:10.1007/s10750-018-3671-3.

Ayayee, P. A., Larsen, T., and Sabree, Z. (2016). Symbiotic essential amino acids provisioning in the American cockroach, Periplaneta americana (Linnaeus) under various dietary conditions. *PeerJ* 4, e2046. doi:10.7717/peerj.2046.

Ayayee, P. A. P. A., Jones, S. C. S. C., and Sabree, Z. L. Z. L. (2015). Can13C stable isotope analysis uncover essential amino acid provisioning by termite-associated gut microbes? *PeerJ* 3, e1218. doi:10.7717/peerj.1218.

Ayayee, P., Bhattacharyya, S., Arnold, T., Werne, J., and Leff, L. (2019). Experimental Investigation of Potential Biological Nitrogen Provisioning by Freshwater Insect Gut Microbiomes Using 15N Isotope Analysis. doi:10.20944/PREPRINTS201908.0034.V1.

Ayayee, P., Ferry, J. G., Hoover, K., Saunders, M., Felton, G., and Rosa, C. (2014). Gut Microbes Contribute to Nitrogen Provisioning in a Wood-Feeding Cerambycid. *Environ. Entomol.* 43, 903–912. doi:10.1603/en14045.

Bertino-Grimaldi, D., Medeiros, M. N., Vieira, R. P., Cardoso, A. M., Turque, A. S., Silveira, C. B., et al. (2013). Bacterial community composition shifts in the gut of Periplaneta americana fed on different lignocellulosic materials. *Springerplus* 2, 609. doi:10.1186/2193-1801-2-609.

Bray, J. R., and Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.* 27, 325–349. doi:10.2307/1942268.

Brown, J. J., Mihaljevic, J. R., Des Marteaux, L., and Hrček, J. (2020). Metacommunity theory for transmission of heritable symbionts within insect communities. *Ecol. Evol.* 10, 1703–1721. doi:https://doi.org/10.1002/ece3.5754.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw.* 80, 1–28. doi:10.18637/jss.v080.i01.

Callahan, B. J., Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., et al. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi:10.1038/nmeth.3869.

Cameron, E. S., Schmidt, P. J., Tremblay, B. J.-M., Emelko, M. B., and Müller, K. M. (2021). Enhancing diversity analysis by repeatedly rarefying next generation sequencing data describing microbial communities. *Sci. Rep.* 11, 22302. doi:10.1038/s41598-021-01636-1.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi:10.1038/nmeth.f.303.

Carrasco, P., Moya, A., Latorre, A., Pérez-Cobas, A. E., Angelova, A., Maiques, E., et al. (2015). Diet shapes the gut microbiota of the omnivorous cockroach Blattella germanica. *FEMS Microbiol. Ecol.* 91, fiv022–fiv022. doi:10.1093/femsec/fiv022.

Cummins, K. (2021). The Use of Macroinvertebrate Functional Feeding Group Analysis to Evaluate, Monitor and Restore Stream Ecosystem Condition. *Rep Glob Heal. Res* 4, 1–8. doi:DOI: 10.29011/2690-9480.100129.

Cummins, K. W. K. M. J. (1979). Feeding ecology of stream invertebrates. *Annu. Rev. Ecol. Syst., 10, 147-172,*.

Dietze, M. (2017). *Ecological forecasting*. Princeton University Press.

Dillon, R. J., and Dillon, V. M. (2004). THE Gut Bacteria of Insects : Nonpathogenic Interactions. *Annu. Rev. Entomol.* 49, 71–92. doi:10.1146/annurev.ento.49.061802.123416.

Douglas, A. E. (2015). Multiorganismal Insects: Diversity and Function of Resident Microorganisms. *Annu. Rev. Entomol.* 60, 17–34. doi:10.1146/annurev-ento-010814-020822.

Drury, B., Rosi-Marshall, E., and Kelly, J. J. (2013). Wastewater Treatment Effluent Reduces the Abundance and Diversity of Benthic Bacterial Communities in Urban and Suburban Rivers. *Appl. Environ. Microbiol.* 79, 1897–1905. doi:10.1128/AEM.03527-12.

Efron, B., and Morris, C. (1977). Stein’s Paradox in Statistics. *Sci. Am. - SCI AMER* 236, 119–127. doi:10.1038/scientificamerican0577-119.

Fang, H., Chen, Y., Huang, L., and He, G. (2017). Analysis of biofilm bacterial communities under different shear stresses using size-fractionated sediment. *Sci. Rep.* 7, 1299. doi:10.1038/s41598-017-01446-4.

Frank, J. A., Reich, C. I., Sharma, S., Weisbaum, J. S., Wilson, B. A., and Olsen, G. J. (2008). Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl. Environ. Microbiol.* 74, 2461–2470. doi:10.1128/AEM.02272-07.

Franzini, P. Z. N., Ramond, J.-B., Scholtz, C. H., Sole, C. L., Ronca, S., and Cowan, D. A. (2016). The Gut Microbiomes of Two Pachysoma MacLeay Desert Dung Beetle Species (Coleoptera: Scarabaeidae: Scarabaeinae) Feeding on Different Diets. *PLoS One* 11, e0161118. Available at: https://doi.org/10.1371/journal.pone.0161118.

Gökçe, K. W. C. E.-D. (2018). “Functional Analysis of Stream Macroinvertebrates,” in (Rijeka: IntechOpen), Ch. 4. doi:10.5772/intechopen.79913.

Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C., and Martiny, J. B. H. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.* 10, 497–506. doi:10.1038/nrmicro2795.

Harris, J. M. (1993). The presence, nature, and role of gut microflora in aquatic invertebrates: A synthesis. *Microb. Ecol.* 25, 195–231. doi:10.1007/BF00171889.

Hosen, J. D., Febria, C. M., Crump, B. C., and Palmer, M. A. (2017). Watershed Urbanization Linked to Differences in Stream Bacterial Community Composition. *Front. Microbiol.* 8, 1452. doi:10.3389/fmicb.2017.01452.

Huang, S., and Zhang, H. (2013). The Impact of Environmental Heterogeneity and Life Stage on the Hindgut Microbiota of Holotrichia parallela Larvae (Coleoptera: Scarabaeidae). *PLoS One* 8. doi:10.1371/journal.pone.0057169.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016). A new view of the tree of life. *Nat. Microbiol.* 1, 16048. doi:10.1038/nmicrobiol.2016.48.

Jizhong, Z., and Daliang, N. (2017). Stochastic Community Assembly: Does It Matter in Microbial Ecology? *Microbiol. Mol. Biol. Rev.* 81, e00002-17. doi:10.1128/MMBR.00002-17.

Kaufman, M. G., Walker, E. D., Odelson, D. A., and Klug, M. J. (2000). Microbial Community Ecology &amp; Insect Nutrition. *Am. Entomol.* 46, 173–185. doi:10.1093/ae/46.3.173.

Keskitalo, A., Pietilä, S., Munukka, Ë., Eerola, E., Pursiheimo, J.-P., Laiho, A., et al. (2017). Gut Microbiota Analysis Results Are Highly Dependent on the 16S rRNA Gene Target Region, Whereas the Impact of DNA Extraction Is Minor. *J. Biomol. Tech.* 28. doi:10.7171/jbt.17-2801-003.

Kroetsch, S. A., Kidd, K. A., Monk, W. A., Culp, J. M., Compson, Z. G., and Pavey, S. A. (2020). The effects of taxonomy, diet, and ecology on the microbiota of riverine macroinvertebrates. *Ecol. Evol.* n/a. doi:https://doi.org/10.1002/ece3.6993.

Kuczynski, J., Stombaugh, J., Walters, W. A., González, A., Caporaso, J. G., and Knight, R. (2012). “Using QIIME to Analyze 16S rRNA Gene Sequences from Microbial Communities,” in *Current Protocols in Microbiology* (John Wiley & Sons, Inc.), 1E.5.1-1E.5.20. doi:10.1002/9780471729259.mc01e05s27.

Louca, S., Mazel, F., Doebeli, M., and Parfrey, L. W. (2019). A census-based estimate of Earth’s bacterial and archaeal diversity. *PLOS Biol.* 17, e3000106. Available at: https://doi.org/10.1371/journal.pbio.3000106.

Ma, B., LaPara, T. M., N. Evans, A., and Hozalski, R. M. (2020). Effects of geographic location and water quality on bacterial communities in full-scale biofilters across North America. *FEMS Microbiol. Ecol.* 96, fiz210. doi:10.1093/femsec/fiz210.

Martin, M. M., Kukor, J. J., Martin, J. S., Lawson, D. L., and Merrit, R. W. (1981a). Digestive enzymes of larvae of three species of caddisflies (Trichoptera). *Insect Biochem.* v. 11.

Martin, M. M., Martin, J. S., Kukor, J. J., and Merritt, R. W. (1980). The Digestion of Protein and Carbohydrate by the Stream Detritivore, Tipula abdominalis (Diptera, Tipulidae). *Oecologia* 46, 360–364. Available at: http://www.jstor.org/stable/4216181.

Martin, M. M., Martin, J. S., Kukor, J. J., and Merritt, R. W. (1981b). The digestive enzymes of detritus-feeding stonefly nymphs (Plecoptera; Pteronarcyidae). *Can. J. Zool.* 59, 1947–1951. doi:10.1139/z81-265.

Mason, C. J., Scully, E. D., Geib, S. M., and Hoover, K. (2016). Contrasting diets reveal metabolic plasticity in the tree-killing beetle, Anoplophora glabripennis (Cerambycidae: Lamiinae). *Sci. Rep.* 6, 33813. doi:10.1038/srep33813.

McElreath Richard (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan.* Chapman and Hall/CRC.

McKnight, D. T., Huerlimann, R., Bower, D. S., Schwarzkopf, L., Alford, R. A., and Zenger, K. R. (2019). Methods for normalizing microbiome data: An ecological perspective. *Methods Ecol. Evol.* 10, 389–400. doi:https://doi.org/10.1111/2041-210X.13115.

Medeiros, J. D., Cantão, M. E., Cesar, D. E., Nicolás, M. F., Diniz, C. G., Silva, V. L., et al. (2016). Comparative metagenome of a stream impacted by the urbanization phenomenon. *Brazilian J. Microbiol.* 47, 835–845. doi:http://dx.doi.org/10.1016/j.bjm.2016.06.011.

Merritt, R. W., Cummins, K. W., and Berg, M. B. (2008). *An Introduction to the Aquatic Insects of North America.* 5th ed. Kendall/Hunt Publishing Co. Dubuque, Iowa, USA.

Mukherjee, S., Huntemann, M., Ivanova, N., Kyrpides, N. C., and Pati, A. (2015). Large-scale contamination of microbial isolate genomes by Illumina PhiX control. *Stand. Genomic Sci.* 10, 18. doi:10.1186/1944-3277-10-18.

Pechal, J. L., and Benbow, M. E. (2016). Microbial ecology of the salmon necrobiome: Evidence salmon carrion decomposition influences aquatic and terrestrial insect microbiomes. *Environ. Microbiol.* 18, 1511–1522. doi:10.1111/1462-2920.13187.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi:10.1093/nar/gks1219.

R Core Team (2022). R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna, Austria. URL http//www.R-project.org/*.

Raychoudhury, R., Sen, R., Cai, Y., Sun, Y., Lietze, V. U., Boucias, D. G., et al. (2013). Comparative metatranscriptomic signatures of wood and paper feeding in the gut of the termite Reticulitermes flavipes (Isoptera: Rhinotermitidae). *Insect Mol. Biol.* 22, 155–171. doi:10.1111/imb.12011.

Receveur, J. P., Fenoglio, S., and Benbow, M. E. (2020). Insect-associated bacterial communities in an alpine stream. *Hydrobiologia* 847, 331–344. doi:10.1007/s10750-019-04097-w.

Rossmassler, K., Dietrich, C., Thompson, C., Mikaelyan, A., Nonoh, J. O., Scheffrahn, R. H., et al. (2015). Metagenomic analysis of the microbiota in the highly compartmented hindguts of six wood- or soil-feeding higher termites. *Microbiome* 3, 56. doi:10.1186/s40168-015-0118-1.

Sazama, E. J., Bosch, M. J., Shouldis, C. S., Ouellette, S. P., and Wesner, J. S. (2017). Incidence of Wolbachia in aquatic insects. *Ecol. Evol.* 7, 1165–1169. doi:10.1002/ece3.2742.

Shannon C.E (1957). A Mathematical Theory of Communication. *Bell Syst. Tech. J.* 27, 379–423. doi:10.1016/S0006-3495(96)79210-X.

Stan Development Team (2022). R: A language and environment for statistical computing.

Stief, P. (2013). Stimulation of microbial nitrogen cycling in aquatic ecosystems by benthic macrofauna: Mechanisms and environmental implications. *Biogeosciences* 10, 7829–7846. doi:10.5194/bg-10-7829-2013.

Stief, P., Poulsen, M., Nielsen, L. P., Brix, H., Schramm, A., Peter, L., et al. (2009). Nitrous oxide emission by aquatic macrofauna. *Proc. Natl. Acad. Sci. U. S. A.* 106, 4296–300. doi:10.1073/pnas.0808228106.

Su, L. J., Yang, L. Le, Huang, S., Su, X. Q., Li, Y., Wang, F. Q., et al. (2016). Comparative gut microbiomes of four species representing the higher and the lower termites. *J. Insect Sci.* 16, 97. doi:10.1093/jisesa/iew081.

Sujay., K., E., L. G., L., P. M., M., U. R., Shahan, H., Julia, G., et al. (2018). Freshwater salinization syndrome on a continental scale. *Proc. Natl. Acad. Sci.* 115, E574–E583. doi:10.1073/pnas.1711234115.

Tierno de Figueroa, J. M., Trenzado, C. E., López-Rodríguez, M. J., and Sanz, A. (2011). Digestive enzyme activity of two stonefly species (Insecta, Plecoptera) and their feeding habits. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* 160, 426–430. doi:http://doi.org/10.1016/j.cbpa.2011.07.014.

Vellend, M. (2010). Conceptual Synthesis in Community Ecology. *Q. Rev. Biol.* 85, 183–206. doi:10.1086/652373.

Wakelin, S. A., Colloff, M. J., and Kookana, R. S. (2008). Effect of Wastewater Treatment Plant Effluent on Microbial Function and Community Structure in the Sediment of a Freshwater Stream with Variable Seasonal Flow . *Appl. Environ. Microbiol.* 74, 2659–2668. doi:10.1128/AEM.02348-07.

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5, 27. doi:10.1186/s40168-017-0237-y.

Wesner, J. S., and Pomeranz, J. P. F. (2021). Choosing priors in Bayesian ecological models by simulating from the prior predictive distribution. *Ecosphere* 12, e03739. doi:https://doi.org/10.1002/ecs2.3739.

Yun, J. H., Roh, S. W., Whon, T. W., Jung, M. J., Kim, M. S., Park, D. S., et al. (2014). Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Appl. Environ. Microbiol.* 80, 5254–5264. doi:10.1128/AEM.01226-14.

**Table 1A.** Summary of the macroinvertebrate samples obtained and used in this study

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample ID** | **Order** | **Family** | **taxon\_name** | **NEON Locations** | **Genera** | **Functional Feeding Group** |
| LEWI-A1 | Diptera | Ceratopogonidae | Ceratopogoninae sp. | LEWI | *NA* | Predators |
| LEWI-A2 | Diptera | Tabanidae | Tabanidae sp. | LEWI | *NA* | Predators |
| MCRA-A11 | Ephemeroptera | Heptageniidae | Epeorus sp. | MCRA | *Epeorus* | Scrapers |
| MCRA-A12 | Plecoptera | Perlidae | Perlidae sp. | MCRA | *NA* | Predators |
| MCRA-A13 | Trichoptera | Hydropsychidae | Arctopsychinae sp. | MCRA | *NA* | Filtering collectors |
| MCRA-A15 | Ephemeroptera | Ameletidae | Ameletus sp. | MCRA | *Ameletus* | Gathering collectors |
| BLUE-A21 | Diptera | Chironomidae | Chironomini sp. | BLUE | *NA* | Filtering collectors |
| BLUE-A22 | Coleoptera | Elmidae | Stenelmis sp. | BLUE | *Stenelmis* | Gathering collectors |
| BLUE-A23 | Ephemeroptera | Ephemeridae | Ephemera sp. | BLUE | *Ephemera* | Scrapers |
| BLUE-A24 | Ephemeroptera | Heptageniidae | Leucrocuta sp. | BLUE | *Leucrocuta* | Scrapers |
| BLUE-A25 | Megaloptera | Corydalidae | Corydalus sp. | BLUE | *Corydalus* | Predators |
| MCDI-A31 | Diptera | Simuliidae | Simulium sp. | MCDI | *Simulium* | Filtering collectors |
| MCDI-A32 | Diptera | Tipulidae | Tipula sp. | MCDI | *Tipula* | Predators |
| MCDI-A33 | Ephemeroptera | Heptageniidae | Stenonema femoratum | MCDI | *Stenonema* | Scrapers |
| MCDI-A34 | Trichoptera | Hydropsychidae | Cheumatopsyche sp. | MCDI | *Cheumatopsyche* | Filtering collectors |
| MCDI-A35 | Ephemeroptera | Baetidae | Fallceon sp. | MCDI | *Fallceon* | Gathering collectors |
| HOPB-A41 | Trichoptera | Hydropsychidae | Cheumatopsyche sp. | HOPB | *Cheumatopsyche* | Filtering collectors |
| HOPB-A42 | Trichoptera | Glossosomatidae | Glossosoma sp. | HOPB | *Glossosoma* | Scrapers |
| HOPB-A43 | Ephemeroptera | Heptageniidae | Maccaffertium sp. | HOPB | *Maccaffertium* | Scrapers |
| HOPB-A44 | Ephemeroptera | Leptophlebiidae | Paraleptophlebia sp. | HOPB | *Paraleptophlebia* | Gathering collectors |
| BLDE-A51 | Trichoptera | Brachycentridae | Micrasema sp. | BLDE | *Micrasema* | Gathering collectors |
| BLD3-A52 | Ephemeroptera | Baetidae | Acentrella sp. | BLDE | *Acentrella* | Gathering collectors |
| BLDE-A53 | Diptera | Psychodidae | Pericoma/Telmatoscopus sp. | BLDE | *Pericoma* | Gathering collectors |
| BLDE-A54 | Plecoptera | Chloroperlidae | Sweltsa sp. | BLDE | *Sweltsa* | Predators |
| LECO-A61 | Diptera | Dixidae | Dixa sp. | LECO | *Dixa* | Filtering collectors |
| LECO-A62 | Coleoptera | Ptilodactylidae | Anchytarsus bicolor | LECO | *Anchytarsus* | Shredder/detritivore |
| LECO-A63 | Plecoptera | Pteronarcyidae | Pteronarcys sp. | LECO | *Pteronarcys* | Shredder/detritivore |
| LECO-A64 | Plecoptera | Perlidae | Acroneuria sp. | LECO | *Acroneuria* | Predators |
| LECO-A65 | Plecoptera | Chloroperlidae | Alloperla sp. | LECO | *Alloperla* | Predators |
| BIGC-A71 | Diptera | Tipulidae | Tipulidae sp. | BIGC | *NA* | Predators |
| BIGC-A72 | Odonata | Cordulegastridae | Cordulegaster sp. | BIGC | *Cordulegaster* | Predators |
| BIGC-A73 | Odonata | Gomphidae | Gomphidae sp. | BIGC | *NA* | Predators |
| BIGC-A74 | Plecoptera | Chloroperlidae | Sweltsa sp. | BIGC | *Sweltsa* | Predators |
| BIGC-A75 | Plecoptera | Leuctridae | Leuctridae sp. | BIGC | *NA* | Shredder/detritivore |
| MART-B1 | Diptera | Tipulidae | Tipula sp. | MART | *Tipula* | Predators |
| MART-B2 | Plecoptera | Chloroperlidae | Sweltsa sp. | MART | *Sweltsa* | Predators |
| MART-B3 | Ephemeroptera | Ephemerellidae | Drunella doddsii | MART | *Drunella* | Scrapers |
| MART-B4 | Coleoptera | Elmidae | Narpus sp. | MART | *Narpus* | Gathering collectors |
| MART-B5 | Coleoptera | Psephenidae | Ectopria sp. | MART | *Ectopria* | Scrapers |
| POSE-B11 | Coleoptera | Elmidae | Optioservus sp. | POSE | *Optioservus* | Gathering collectors |
| POSE-B12 | Ephemeroptera | Leptophlebiidae | Paraleptophlebia sp. | POSE | *Paraleptophlebia* | Gathering collectors |

**Table 1B.** Summary of the macroinvertebrate samples obtained and used in subset analyses

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SampleID** | **Locations** | **Family** | **Feeding\_Group** | **order** | **genus\_taxized** |
| MCDI-A32 | MCDI | Tipulidae | Predators | Diptera | Tipula |
| HOPB-A44 | HOPB | Leptophlebiidae | Gathering collectors | Ephemeroptera | Paraleptophlebia |
| BLDE-A54 | BLDE | Chloroperlidae | Predators | Plecoptera | Sweltsa |
| BIGC-A74 | BIGC | Chloroperlidae | Predators | Plecoptera | Sweltsa |
| MART-B1 | MART | Tipulidae | Predators | Diptera | Tipula |
| MART-B2 | MART | Chloroperlidae | Predators | Plecoptera | Sweltsa |
| POSE-B12 | POSE | Leptophlebiidae | Gathering collectors | Ephemeroptera | Paraleptophlebia |

**Table 2.** Non-parametric test output of alpha diversity metrices for macroinvertebrate sample variables (location, order, family, genera, and functional feeding group (FFGs).

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Diversity indices** |  |
| Variables | Observed\_species | Chao1 | Shannon's |
| Location | χ2 = 12. 76, P = 0.17 | χ2 = 12. 61, P = 0.18 | χ2 = 11.61, P = 0.24 |
| Order | χ2 = 8.10, P = 0.23 | χ2 = 5.37, P = 0.49 | χ2 = 3.67, P = 0.72 |
| Family | χ2 = 27.71, P = 0.32 | χ2 = 23.03, P = 0.57 | χ2 = 27.98, P = 0.31 |
| Genus | χ2 = 34.90, P = 0.52 | χ2 = 34.18, P = 0.55 | χ2 = 37.02 P = 0.42 |
| Functional feeding group | χ2 = 5.74, P = 0.22 | χ2 = 3.92, P = 0.42 | χ2 = 9.15, P = 0.057 |

**Table 3.** Results of wsp-PCR on freshwater macroinvertebrates obtained from 10 NEON freshwater longterm ecological research sites.Positive samples are highlighted in red.

|  |  |  |
| --- | --- | --- |
| **Sample ID** | **Order** | **Family** |
| LEWI-A1 | Diptera | Ceratopogonidae |
| LEWI-A2 | Diptera | Tabanidae |
| MCRA-A11 | Ephemeroptera | Heptageniidae |
| MCRA-A12 | Plecoptera | Perlidae |
| MCRA-A13 | Trichoptera | Hydropsychidae |
| MCRA-A15 | Ephemeroptera | Ameletidae |
| BLUE-A21 | Diptera | Chironomidae |
| *BLUE-A22* | *Coleoptera* | *Elmidae* |
| BLUE-A23 | Ephemeroptera | Ephemeridae |
| BLUE-A24 | Ephemeroptera | Heptageniidae |
| BLUE-A25 | Megaloptera | Corydalidae |
| MCDI-A31 | Diptera | Simuliidae |
| MCDI-A32 | Diptera | Tipulidae |
| MCDI-A33 | Ephemeroptera | Heptageniidae |
| MCDI-A34 | Trichoptera | Hydropsychidae |
| MCDI-A35 | Ephemeroptera | Baetidae |
| HOPB-A41 | Trichoptera | Hydropsychidae |
| *HOPB-A42* | *Trichoptera* | *Glossosomatidae* |
| HOPB-A43 | Ephemeroptera | Heptageniidae |
| HOPB-A44 | Ephemeroptera | Leptophlebiidae |
| BLDE-A51 | Trichoptera | Brachycentridae |
| BLD3-A52 | Ephemeroptera | Baetidae |
| BLDE-A53 | Diptera | Psychodidae |
| *BLDE-A54* | *Plecoptera* | *Chloroperlidae* |
| LECO-A61 | Diptera | Dixidae |
| LECO-A62 | Coleoptera | Ptilodactylidae |
| LECO-A63 | Plecoptera | Pteronarcyidae |
| LECO-A64 | Plecoptera | Perlidae |
| LECO-A65 | Plecoptera | Chloroperlidae |
| BIGC-A71 | Diptera | Tipulidae |
| *BIGC-A72* | *Odonata* | *Cordulegastridae* |
| *BIGC-A73* | *Odonata* | *Gomphidae* |
| BIGC-A74 | Plecoptera | Chloroperlidae |
| BIGC-A75 | Plecoptera | Leuctridae |
| MART-B1 | Diptera | Tipulidae |
| MART-B2 | Plecoptera | Chloroperlidae |
| MART-B3 | Ephemeroptera | Ephemerellidae |
| MART-B4 | Coleoptera | Elmidae |
| MART-B5 | Coleoptera | Psephenidae |
| *POSE-B11* | *Coleoptera* | *Elmidae* |
| POSE-B12 | Ephemeroptera | Leptophlebiidae |

**Figure 1.** The distribution of the ten NEON longterm ecological research freshwater stream sampling sites across the United States.

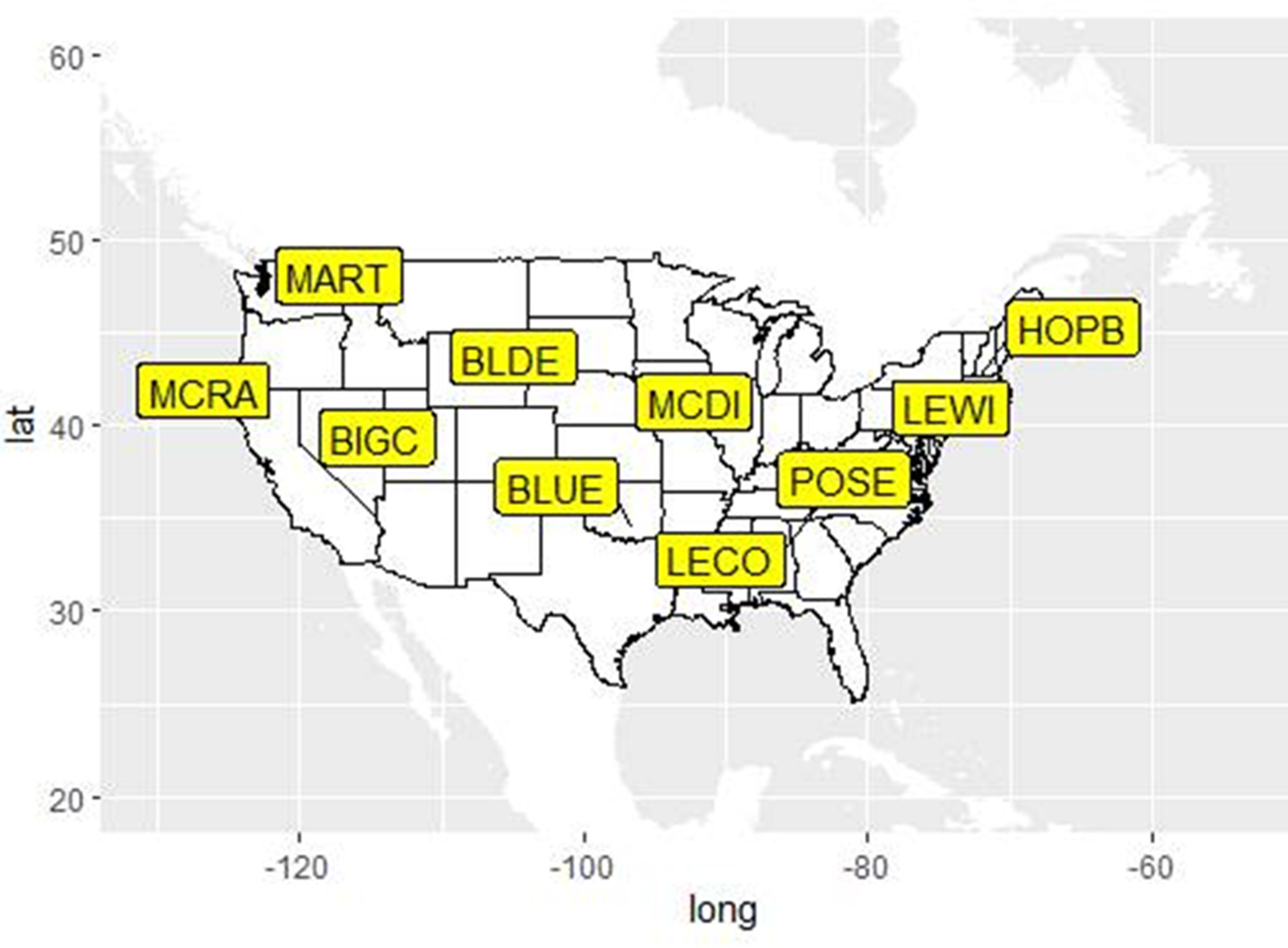
**Figure 2.** Macroinvertebrate summary. The breakdown of freshwater macroinvertebrates obtained from the 10 NEON sites, seven macroinvertebrate orders, 26 families, and five functional feedging group designations used in this study.

**Figure 3.** Gut microbial composition among macroinvertebrate orders (**A**). Gut microbial composition among macroinvertebrate families (**B**). Gut microbial composition among macroinvertebrate genera (**C**). Gut microbial composition among macroinvertebrate functional feeding groups (**D**). Gut microbial composition among macroinvertebrate sample clusters (**E**). stress 0.16.

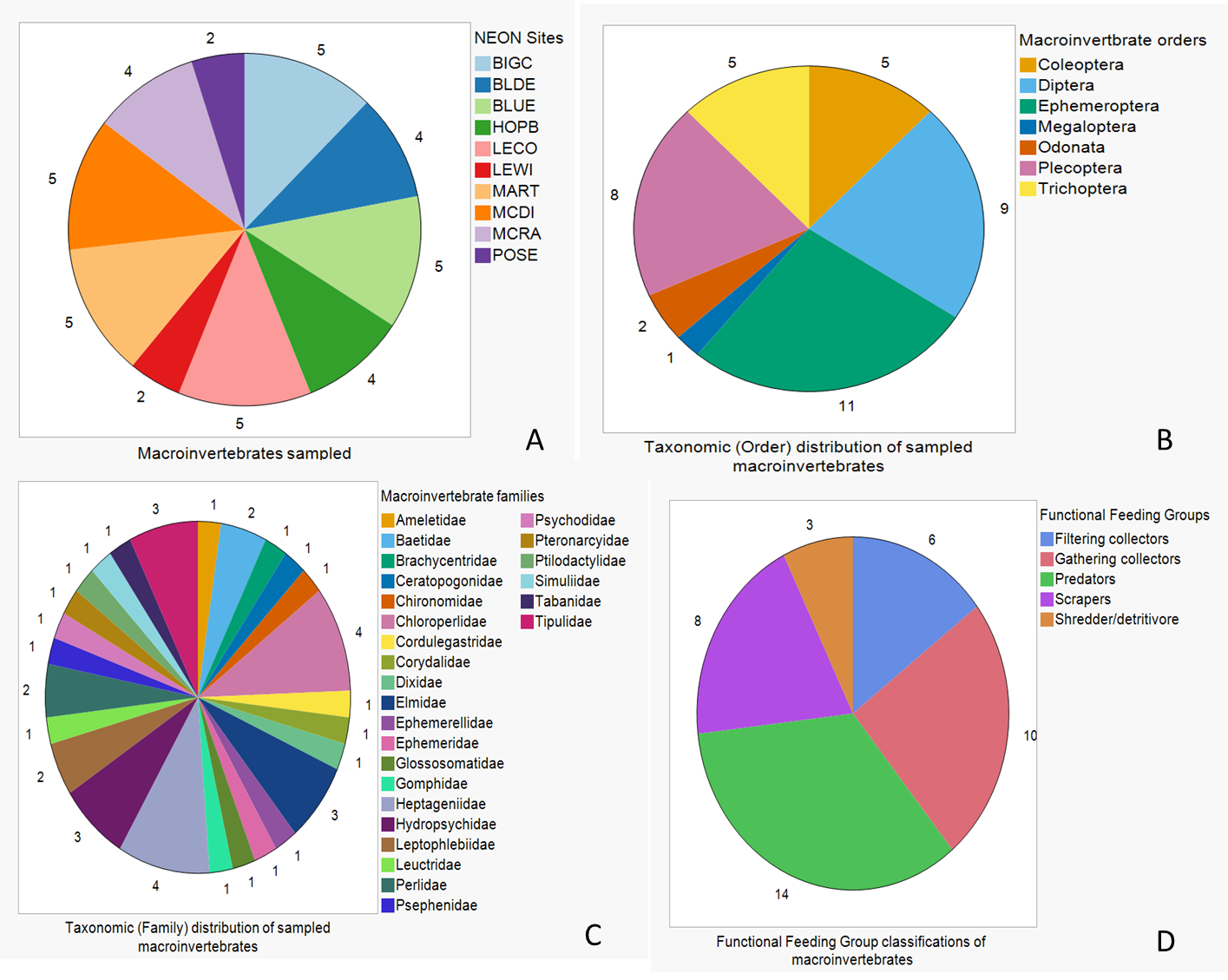
**Figure 4.** Relative abundance of differentially abundant bacterial ASVs at the family level that may be undercoring differences in microbiome community composition among freshwater macroinvertebrate orders (**A**) and functional feeding groups (FFGs)(**B**) used in this study.

**Figure 5.** Model results for beta diversity as measured by Bray-Curtis Distances. A) Model predictions for mean Bray-Curtis distances for stream locations, macroinvertebrate family, genus, and functional feeding group. Different colors represents the families and genera of each macroinvertebrate order; light gray shows the posterior distributions of Bray-Curtis distances for each level of stream, family, genus, and functional feeding group,and black shows the mean distance. In B) we attributed the variation in the entire sample (gray) to each of the four groupings (functional group, stream, insect family, and genus), while black shows the mean distance.

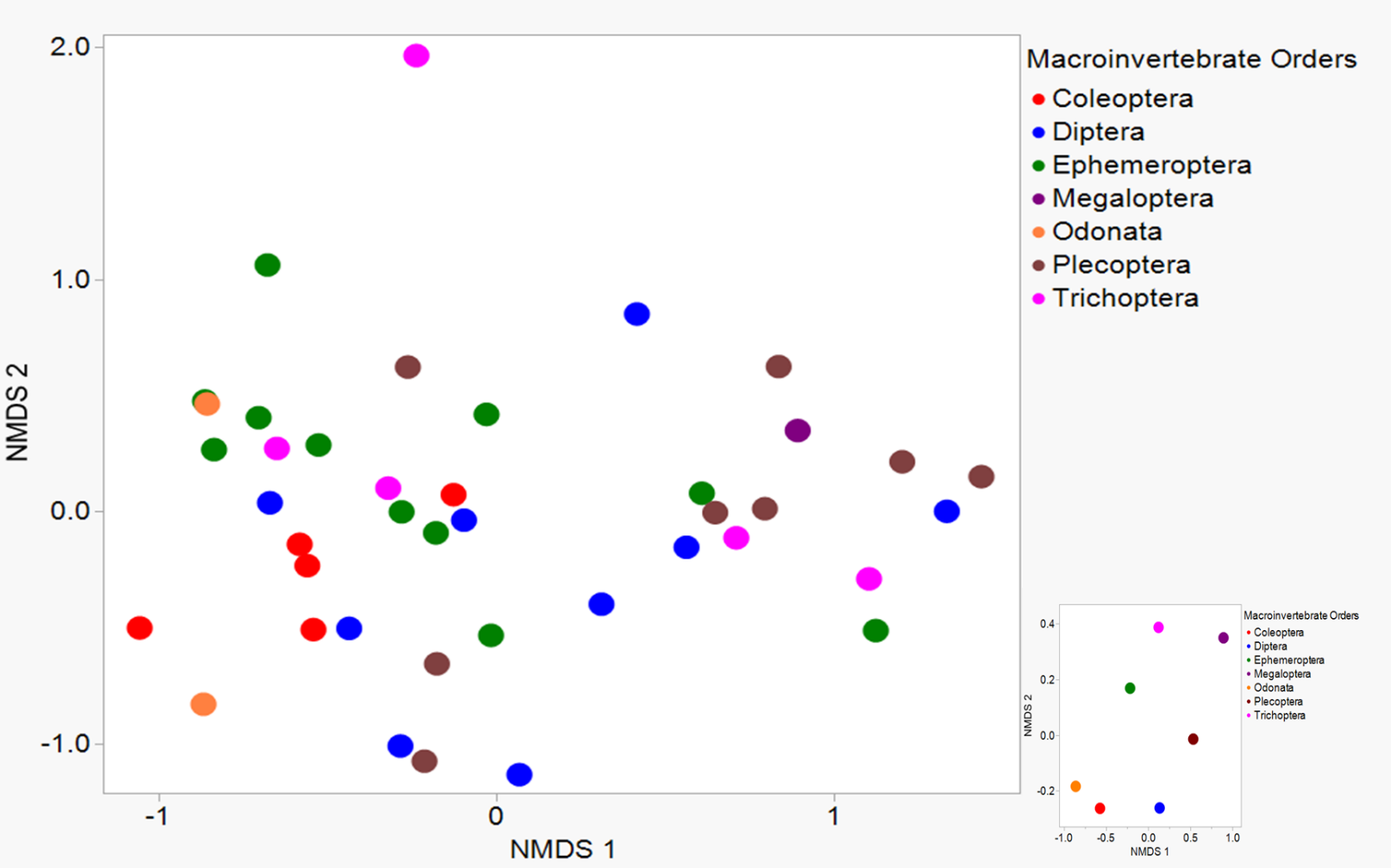
**Figure 1**.**NEON Freshwater stream site locations**



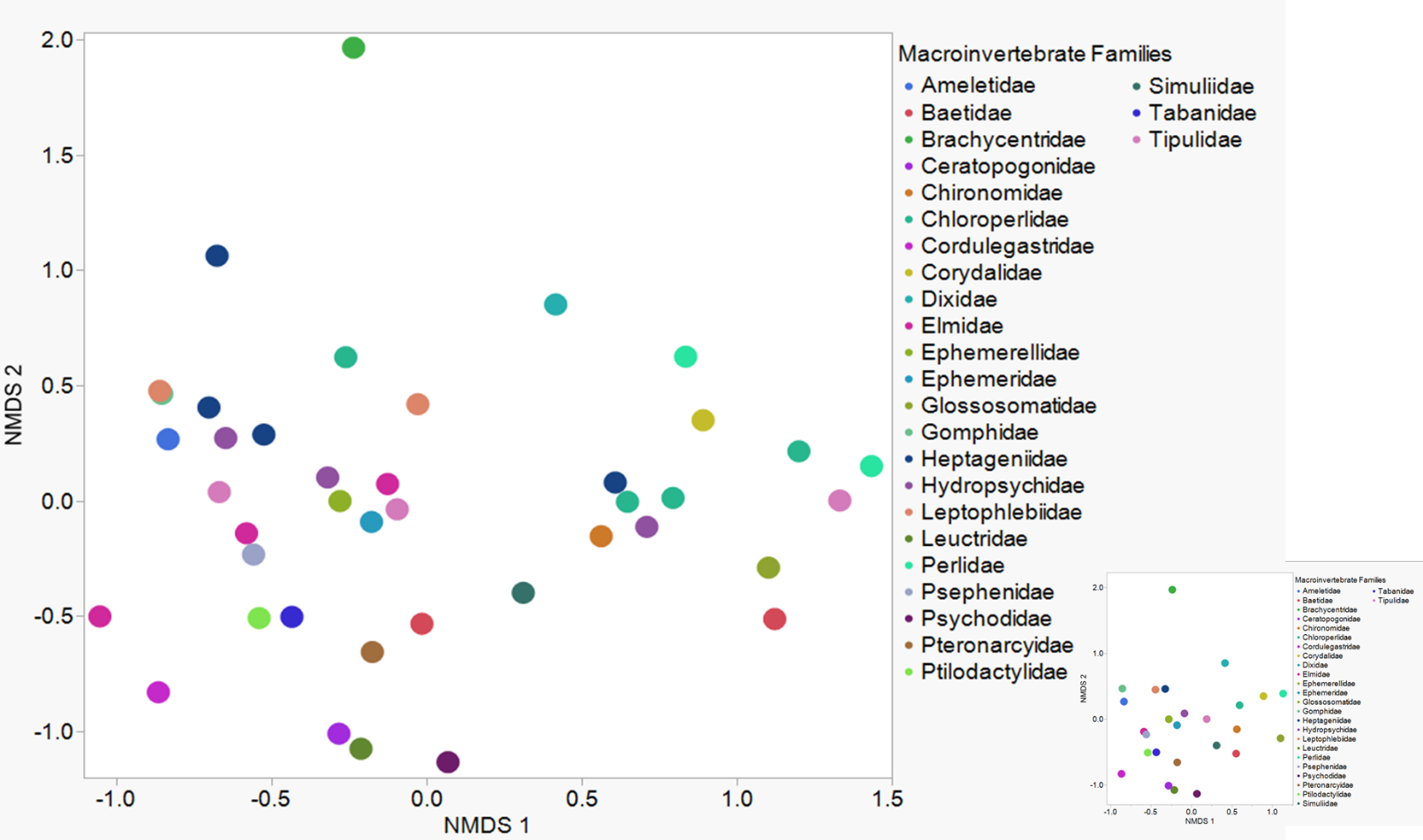
**Figure 2. Macroinvertebrate summary**



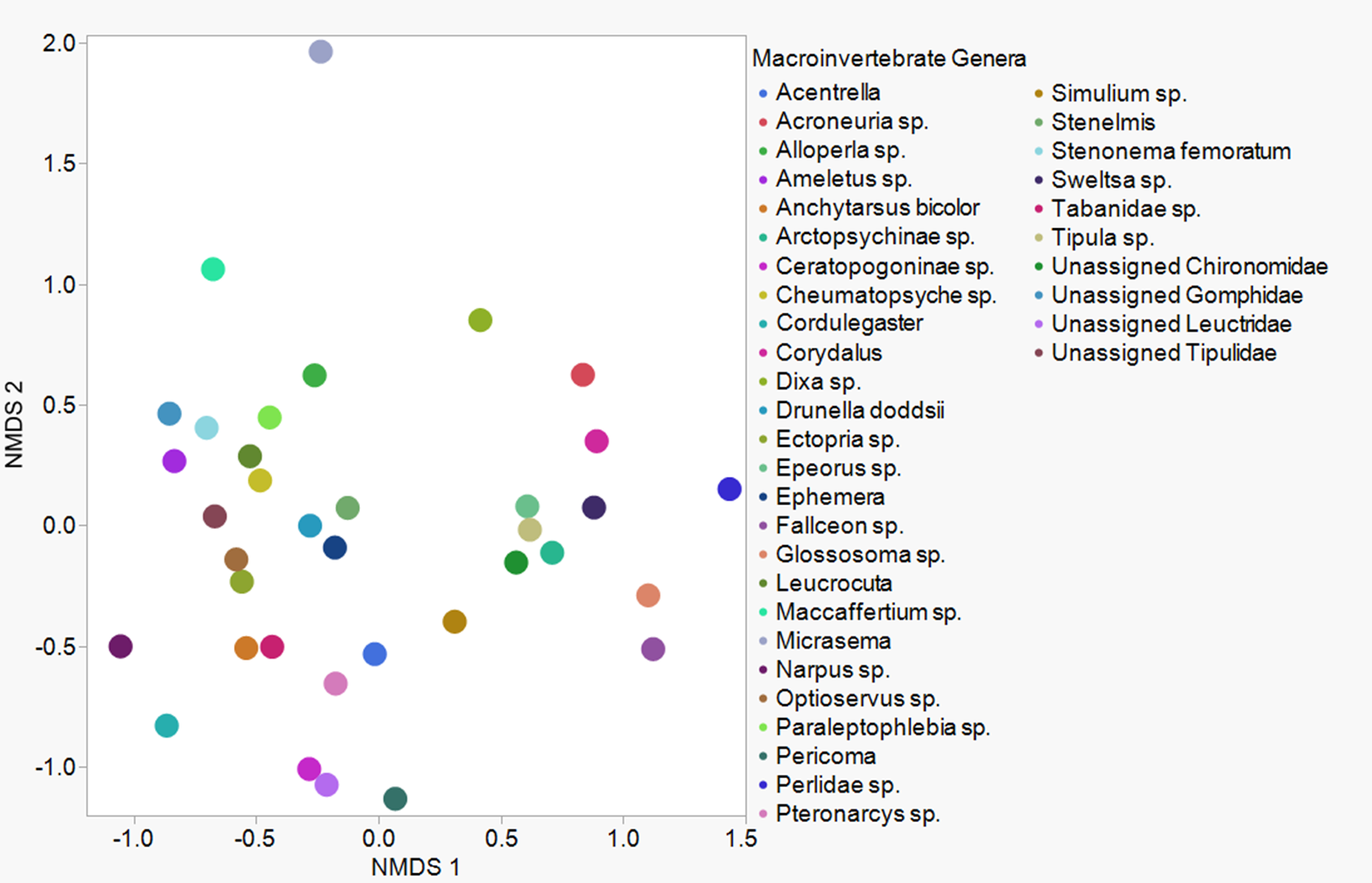
**Figure 3A**. Gut microbial composition among macroinvertebrate orders



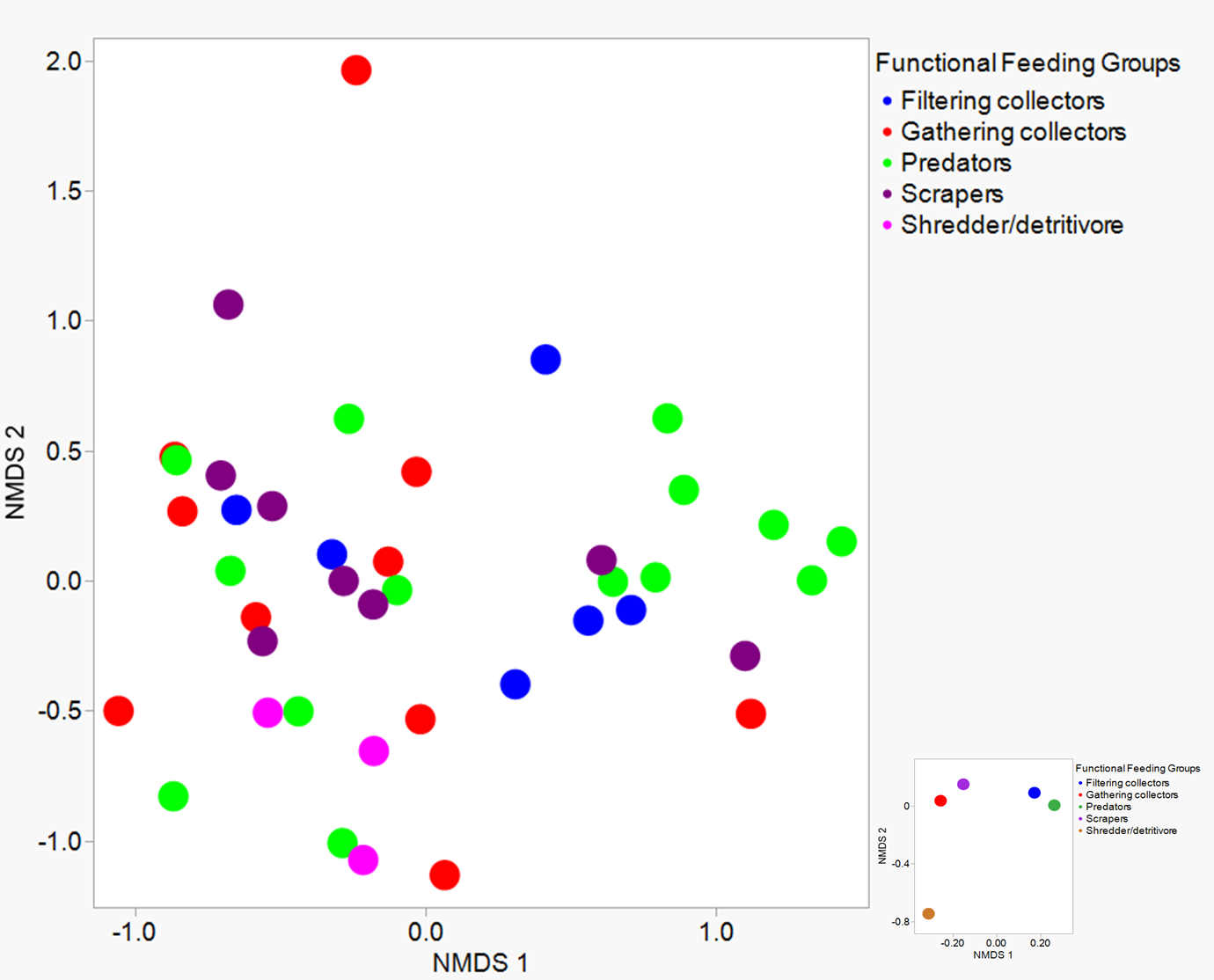
**Figure 3B**. Gut microbial composition among macroinvertebrate families. Stress 0.16



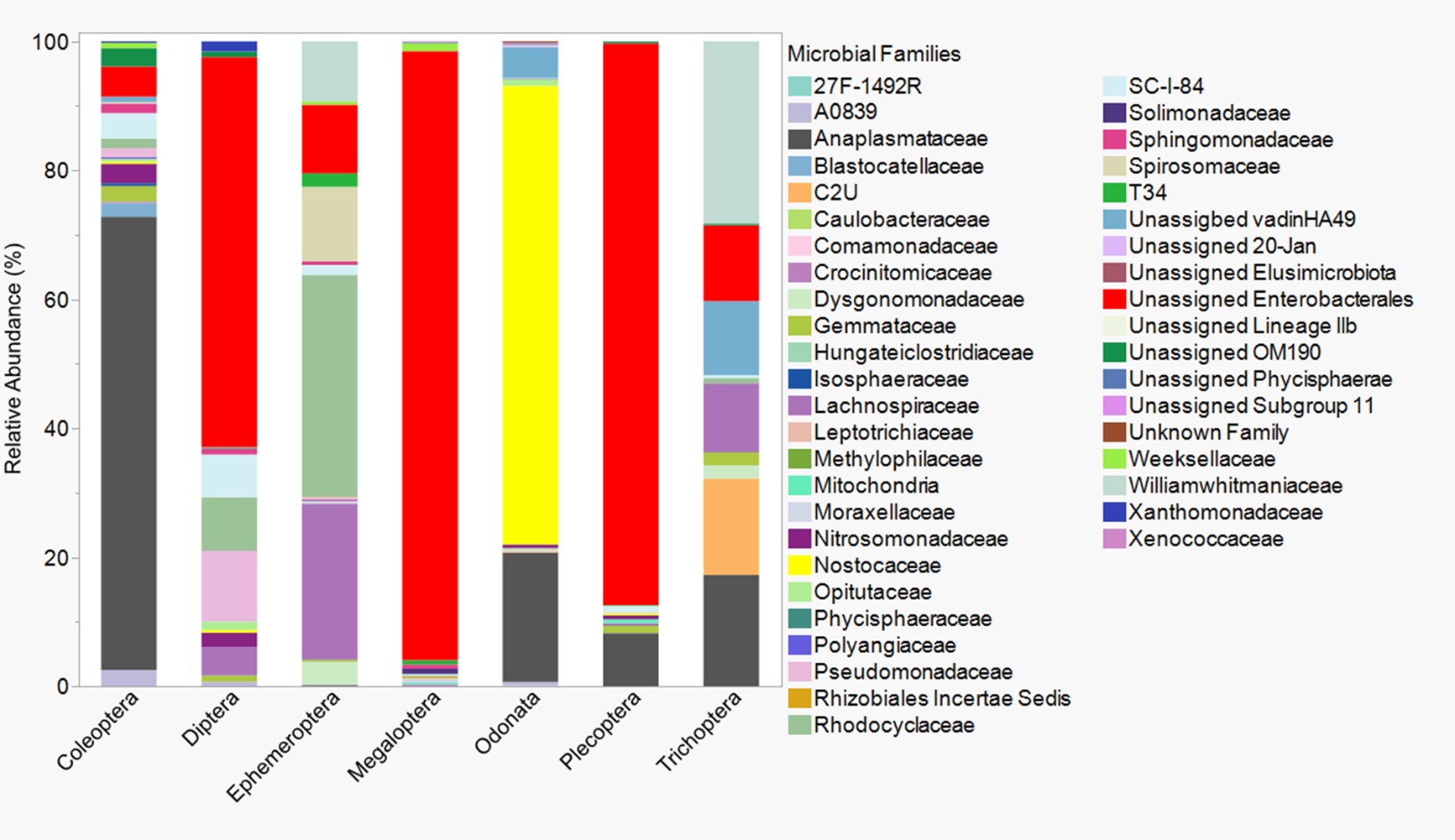
**Figure 3C.** Gut microbial composition among macroinvertebrate genera



**Figure 3D**. Gut microbial composition among macroinvertebrate functional feeding groups



**Figure 4A.** Differentially abundant bacterial ASVs across macroinvertebrate orders

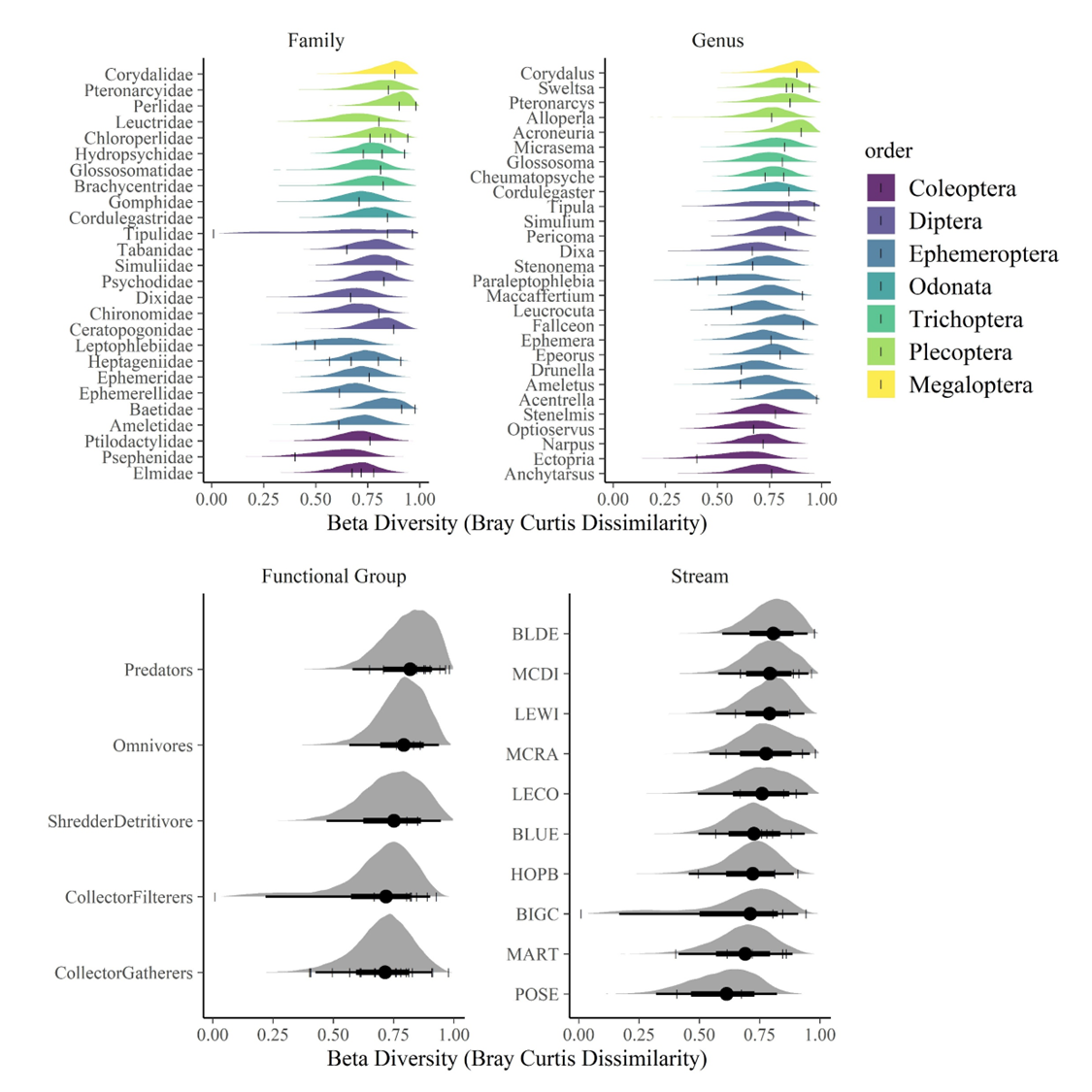


**Figure 4B.** Differentially abundant bacterial ASVs across macroinvertebrate functional feeding groups

Chart, bar chart

Description automatically generated

**Figure 5A.**



**Figure 5B.**

