

Manuscript Title

This manuscript ([permalink](#)) was automatically generated from [imasianxd/ProjectFrog_MANUSCRIPT@f085f3b](#) on March 29, 2021.

Authors

- **John Doe**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [johndoe](#) ·  [johndoe](#)

Department of Something, University of Whatever · Funded by Grant XXXXXXXX

- **Jane Roe**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [janeroe](#)

Department of Something, University of Whatever; Department of Whatever, University of Something

Abstract

Municipal wastewater is one of the principal sources of contaminants delivered into urban aquatic environments. These contaminants include endocrine disrupting compounds (EDCs), which can perturb diverse biological processes by disrupting normal endocrine functions even at the levels found in treated wastewater effluents. Current wastewater regulations, testing methodologies and management strategies often do not take EDCs into account, further, even when testing strategies are in place, EDCs are often biologically active below detection thresholds. The effect of EDCs is seen in a North American *Xenopus* (bullfrog) metamorphosis model, where transformation of an aquatic tadpole into a juvenile frog can be disrupted by their presence. Although there is the potential to use amphibian metamorphosis as an assay for EDC presence, current methodologies rely on a 21-day morphological assessment. However, the use of microbiome profiling is quickly emerging as a novel tool for understanding environmental toxicants and their impact with a lower detection threshold than current chemical tests. Here, we used 16S rRNA sequencing to profile the skin microbiome of the bullfrog at different developmental stages as well as upon exposure to thyroid hormones T3 and T4 to investigate if the bullfrog skin microbiome can be used as an indicator of environmental EDC contamination. We found that T3 exposure but not T4 exposure had an effect on tadpole skin microbiomes. We found that the bacterial communities on the skin of older frogs had significantly lower alpha diversity and higher abundance of a *Luteolibacter* sp. than found on younger frogs. The thyroid hormone T4, which strongly affects frog development and metamorphosis, had a significant effect on the skin microbiome. These results suggest that ECD-contaminated waste water exposure can have minimal direct effects on the *Lithobates catesbeianus* skin microbiome, unless the EDC has a strong effect on the host animal, in this case with metamorphic progress. Taken together with the differences observed in the microbiome with metamorphic stage, these results demonstrate that the *Lithobates catesbeianus* skin microbiome is predominantly determined by the host animal. This suggests that the skin microbiome may be a suitable tool to monitor EDCs at levels that are biologically consequential without a tendency for false positives.

Keywords

word1, word2, word3

Introduction

Previous work used RNA-Seq and Kraken filtering to demonstrate that Ranavirus infection resulted in a differential abundance of bacterial transcripts from specific bacterial communities. I have applied the same Kraken filtering techniques to RNA-Seq datasets from tail fin (TF) and olfactory epithelium (OE) tissues from pre-metamorphic bullfrog tadpoles exposed to either E2, T4, or T3. This data will be paired with traditional 16S rRNA sequencing methods to examine the effects these exposures have on the *Rana catesbeiana* bacterial and viral communities.

Microbiome and why study it? Environmental changes don't just impact large animals, but also effect microbes that play pivotal role in an ecosystem. Every niche/surface/environment has a specific microbial community which contribute to the function of the system.

Why using frog to study environmental pollutants?

Bullfrog developmental models are very sensitive to exogenous T3, as well as their sex ratios are affected by endocrine disruption in the environment. Therefore bullfrogs' skin can act as a sensitive model to study effects of hormones and disruptors on the environment. Frog metamorphosis represents the most accessible and striking example of TH action. This process is entirely dependent upon TH and completion of this developmental period requires a rise in TH levels from zero baseline levels found in the premetamorphic tadpole. TH levels naturally increase from premetamorphosis through prometamorphosis to reach maximal levels at metamorphic climax at which time rapid, overt remodeling of the tadpole into a froglet occurs. Virtually all tissues are targets for TH action, yet the end result includes diverse responses e.g. complete tail resorption, liver reprogramming, brain remodeling, and limb growth. These changes are precociously induced in the premetamorphic tadpole by exogenous administration of T3 or T4

Our frog species, Ranidae frogs The current OECD 21 day assay evaluates the whole organismal response with assessment of external morphological criteria at days 7 and 21 and qualitative assessment of thyroid gland morphology at day 21. It relies upon highly inbred laboratory strains of *Xenopus* species that are not relevant to the North American environment.

This work is part of the WEC project in which the response of pre-metamorphic tadpoles is measured to 3 different hormone treatments: Thyroxine (T4), triiodothyronine (T3), and 17-beta-estradiol (E2). T4 is the precursor to the more bioactive thyroid hormone, T3. E2 here, acts a control group for hormone exposure, but is also used to monitor potential crosstalk between thyroid- and estrogen-mediated pathways.

Methods

Sampling

Premetamorphic tadpoles (TK stages I-VI) were exposed at 21 °C for 48 h to concentrations of T3 (Sigma-Aldrich, Oakville, ON; Catalog #T2752, CAS 55-06-1), T4 (Sigma, Catalog #T2501, CAS 6106-07-6), or E2 (Sigma, Catalog #E4389, PubChem Substance ID: 329799056). Five tadpoles per exposure condition were exposed to either 10 nM T3 in 40nM NaOH, 50 nM T4 in 50nM NaOH exposure, 10 nM E2 in dechlorinated water, or their respective vehicle control. The concentrations chosen were based upon observed physiological and environmental relevance [??? et al., 2016] After completion of the experimental exposures, tadpoles were euthanized in 0.1% buffered tricaine methanesulfonate (TMS, Aqua Life, Syndel Laboratories, Nanaimo, BC, Canada) and the TF and rostrum were dissected, preserved in RNAlater solution (Ambion, Foster City, CA, USA), and stored at -20 °C. The rostrum was halved and sub-dissected to isolate the olfactory sac containing the OE. The OE and TF from each sample animal were then combined in 300 µL TRIzol and mechanically disrupted in a Retsch MM301 Mixer Mill (Thermo Fisher Scientific, Ottawa, Canada) at 25 Hz for two 3-min intervals, separated by a

180° rotation of the samples. After pelleting insoluble material with centrifugation at 12,000 ×g for 10 min at 4 °C, the supernatant was transferred to a new RNase-free tube. RNA was extracted and washed using chloroform, isopropanol, and ethanol treatments and subsequently dissolved in 30 µL diethyl pyrocarbonate-treated water (Sigma-Aldrich) and stored at -80 °C.

16S rRNA Amplicon Sequencing

We performed 16S ribosomal RNA (rRNA) sequencing on 38 tadpole and frog samples, 5 control swabs from gloves and swabs, and 6 control swabs from tank biofilm and water. A set of positive controls consisting of predefined communities, and negative controls were also included as part of the sequencing experiment. The 16S rDNA was amplified using the following primer pair. Forward: Klin16SF: 5'-TCGTCCGGCAGCGTCAGATGTGTATAAGAGACAGAYTGGGYDTAAAGNG-3' and reverse Klin16SR: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACNVGGGTATCTAATCC-3' After amplification, the amplicons are prepared for sequencing following the Illumina 16S Metagenomic Sequencing Library Preparation guide (Ref. Illumina. 2017). The success of library preparation was verified by agarose gels. Equimolar amount of all samples with an additional positive and negative control was pooled and sequenced in four independent runs with the Illumina MiSeq platform using the MiSeq Reagent Kit V2 (Illumina), producing 251 bases paired end reads. The sequence data is available in NCBI SRA database under biosample ID 123456789.

16S rRNA Amplicon Analysis

Raw sequencing reads were demultiplexed and processing within QIIME2 (Ver. 2018.4) [1] following the standard ASV identification workflow. Amplicon sequence variants (ASVs) identified using QIIME2 submodule DADA2 [2]. Taxonomic assignment of ASVs was done using QIIME2 VSEARCH submodule against the Silva 16s rRNA database (ver. 132, 99%;) [3]. Alpha diversity of microbiomes was calculated using the Shannon-Wiener Diversity Index. Beta diversity was measured using the weighted-unifrac distance and visualized using PCoA. Taxa correlation with life stage was identified using pearson correlation.

Meta-transcriptomic Sequencing

The RNA quality and quantity were verified by a Nanodrop spectrophotometer (Thermo Scientific NanoDrop One spectrophotometer, Thermo Scientific). The integrity of isolated RNA was analyzed using a Bioanalyzer 2100 (Agilent Technologies), and samples with RNA integrity number (RIN) of > 7 were used for RNA-Seq analyses (5 biological replicates per treatment). RNA samples were shipped on dry ice to Canada's Michael Smith Genome Sciences Centre (GSC, BC Cancer Research, Canada), where strand-specific mRNA libraries were constructed and sequenced using Illumina HiSeq 2500 (paired-end platform generating 2 x 75 base pair reads for each sample). The sequence data is available in NCBI SRA database under biosample ID 123456789.

Meta-transcriptiomic Analysis

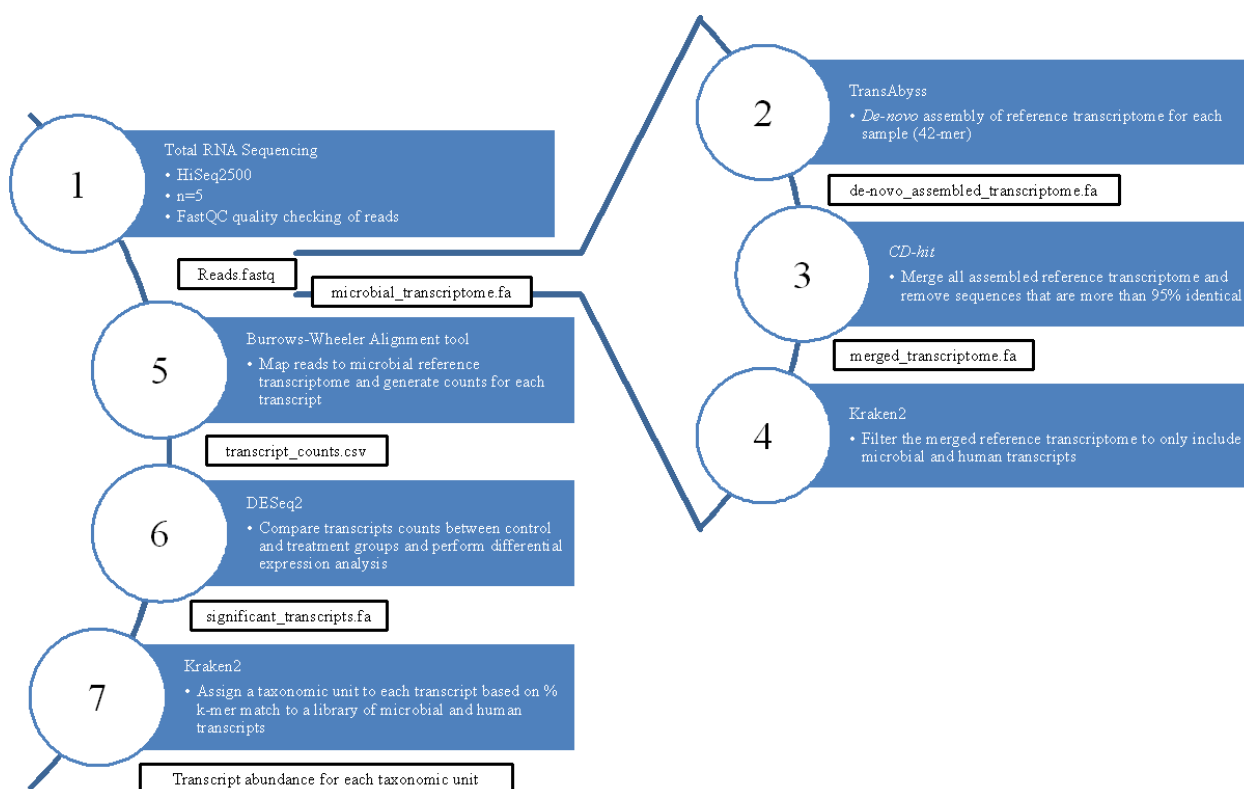
Our RNA-Seq analysis pipeline follows the schematic presented in figure (??). Reads quality was assessed using FastQC to ensure high read quality in all samples. Using TransAbyss, the fastq reads generated from each sample was assembled denovo into a reference transcriptome library representing each transcript that could be assembled given the paired-end reads supplied. We created a complete reference transcriptome assembled using TransAbyss All reference transcriptome libraries were then merged and filtered to remove duplicated sequences, and this merged reference transcriptome was filtered through a large library of microbial RefSeq sequences using Kraken2 to create a library of microbial transcript sequences that were found in one or more of the sequenced samples.

Kraken2 is a taxonomic classification system that relies on exact k-mer matches instead of percent identity and coverage. Kraken2 matches the query sequences to the lowest common ancestor that is matched within the selected database. Kraken2 will assign a taxonomy if the number of identical k-mer matches passes a threshold percentage of all k-mers of a given sequence, as set by the user. The optimal threshold for taxonomic assignment in this dataset was determined to be a 40% confidence threshold of k-mer matches (Fig. ??).

The Kraken2 database used to filter for microbial transcripts includes the following RefSeq sequences: -Archaea: RefSeq complete archaeal genomes/proteins -Bacteria: RefSeq complete bacterial genomes/proteins -Viral: RefSeq complete viral genomes/proteins -Human: GRCh38 human genome/proteins -Fungi: RefSeq complete fungal genomes/proteins -Plant: RefSeq complete plant genomes/proteins -Protozoa: RefSeq complete protozoan genomes/proteins

Human RefSeq sequences were included in the Kraken2 filter in order identify human contamination and separate these transcripts later in the microbial RNA-Seq pipeline.

We used this new filtered list of microbial transcripts as our microbial reference transcriptome and re-mapped the raw reads from each sequenced sample to this new reference using Burrow-Wheeler Alignment tool. The mapped reads were the assembled into transcripts and counted for differential expression analysis using DESeq2. We then extracted two lists of sequences for each WEC exposure condition and tissue type: one list containing all transcripts that were significantly represented in the majority of samples within a given exposure and control set ($\text{cpm} > 0.1$), and one list containing all transcripts that were significantly differentially expressed with a specific exposure condition. These two lists were then fed through the kraken2 databases consisting of each of the RefSeq libraries above, which will assign a taxonomy to a transcript at a match threshold.



K-mer threshold for taxonomy identification. Excluding counts generated using a 10 to 30 % confidence intervals for kraken2 mapping creates a linear correlation ($R^2 > 0.95$) between percent confidence and number of bacterial.

Results

Characterization of *Lithobates catesbeianus* skin bacterial microbiome across life stages.

Overall, seven phyla were detected at a high abundance and represented >95% of all bullfrog microbiomes (Fig. [???]). Proteobacteria were the most dominant in tadpoles (median 27%) followed by Fusobacteria (median 12%), but both of their relative abundance decreased with life stage. At the froglet stage, Verrucomicrobia (median 89%) became the most dominant phyla.

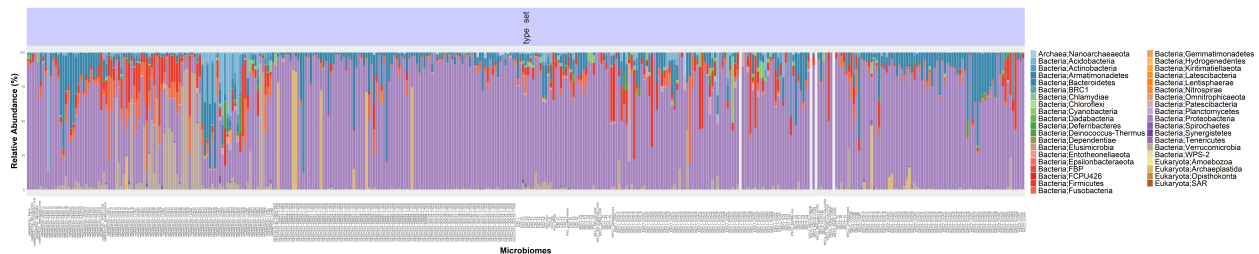
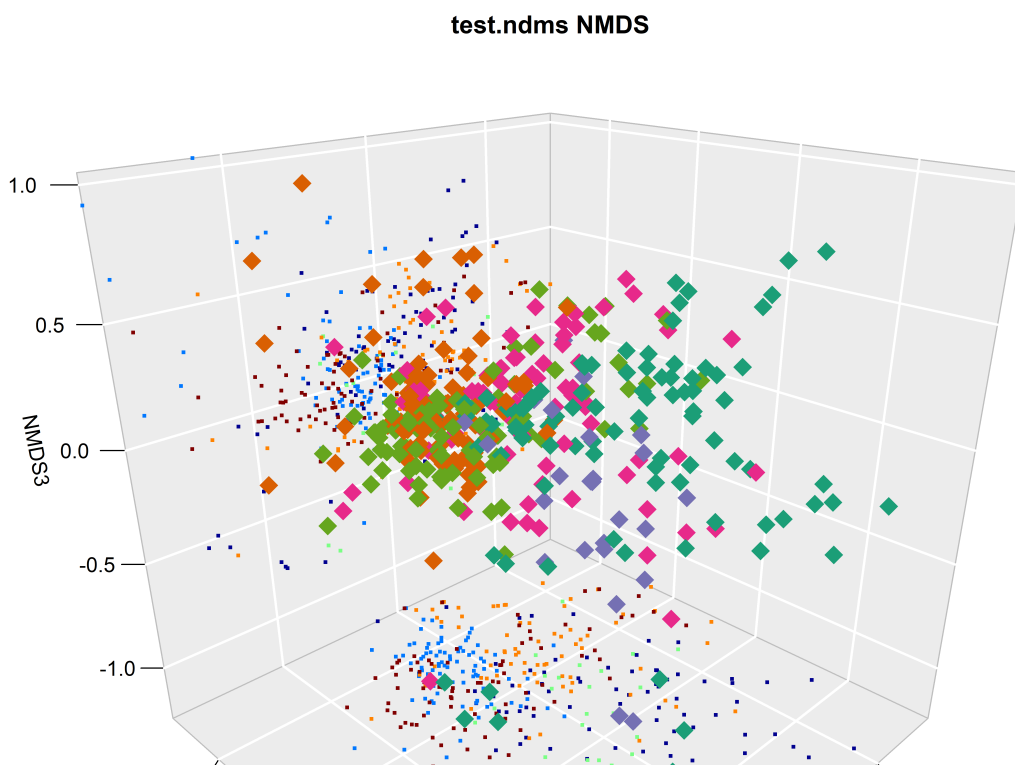
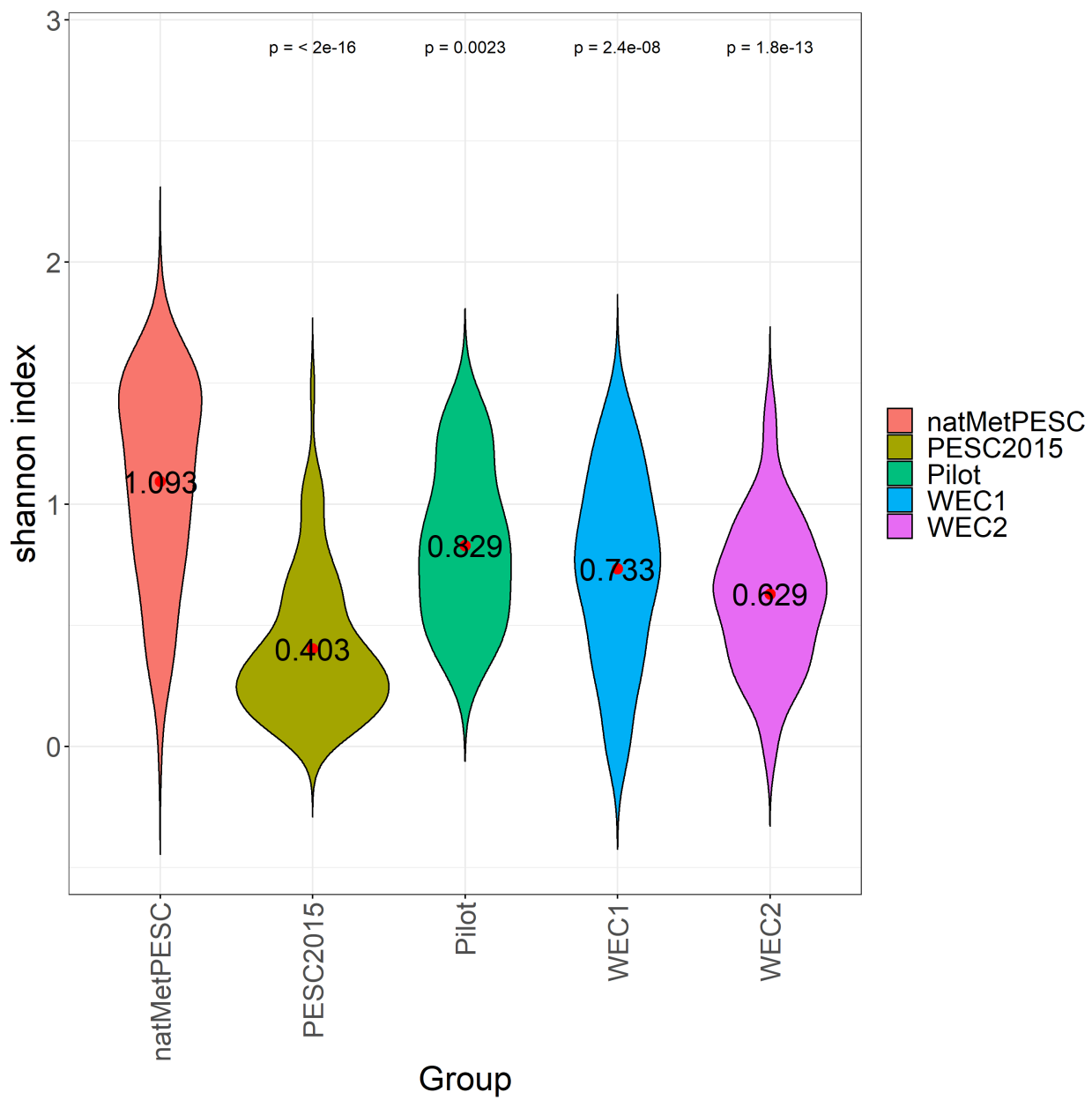
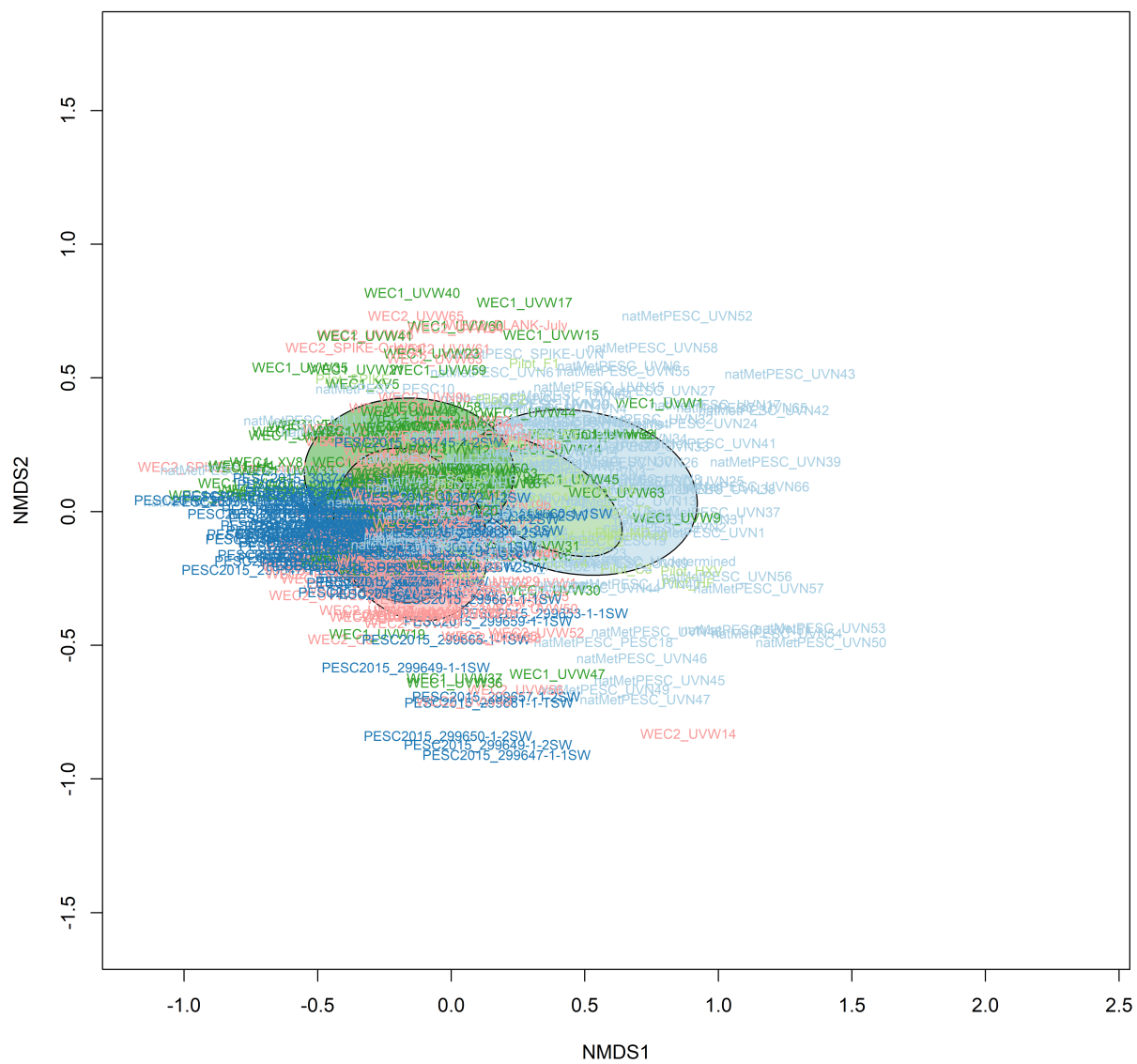
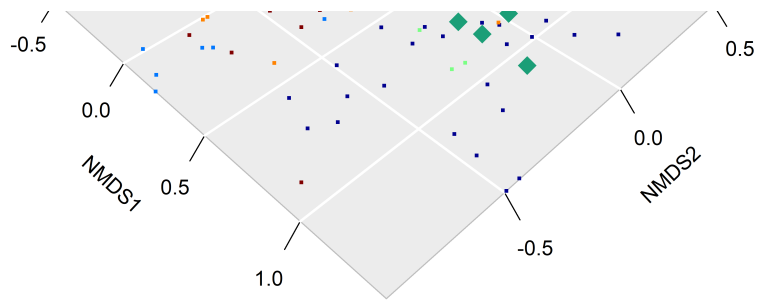


Figure 1: Relative Abundance of 16S Microbiomes at the Bacterial Phylum Level. The relative abundance of the top 10 bacterial phyla present across all samples represents >99% of all reads sequenced. The tadpole skin microbiome (labelled 1-23) consisted of predominately Proteobacteria (Purple). However, as life stage progresses, froglets are dominated by Verrucomicrobia. The relative distribution of phyla in the positive control (labelled POS) is as expected. Sterile swabs (labelled SSC) and negative controls (labelled NSC) have <2k reads on average.

We further explored the difference between the lifestages using an alpha diversity measure via the Shannon-Wiener index. Tadpoles exhibited the greatest alpha diversity, decreasing as it progresses to froglets (Fig. [???]). Using unweighted UniFrac distance, we showed the separation of microbiomes associated with each life stage. We captured this separation using a 2-axis PCoA, demonstrating a distinct separation of froglet and tadpoles, with samples in different stages of metamorphosis scattered in between (Fig. [???], Fig. [???]). Frog sex, mass and length of the tadpole had no impact on the community composition (Fig. [???])





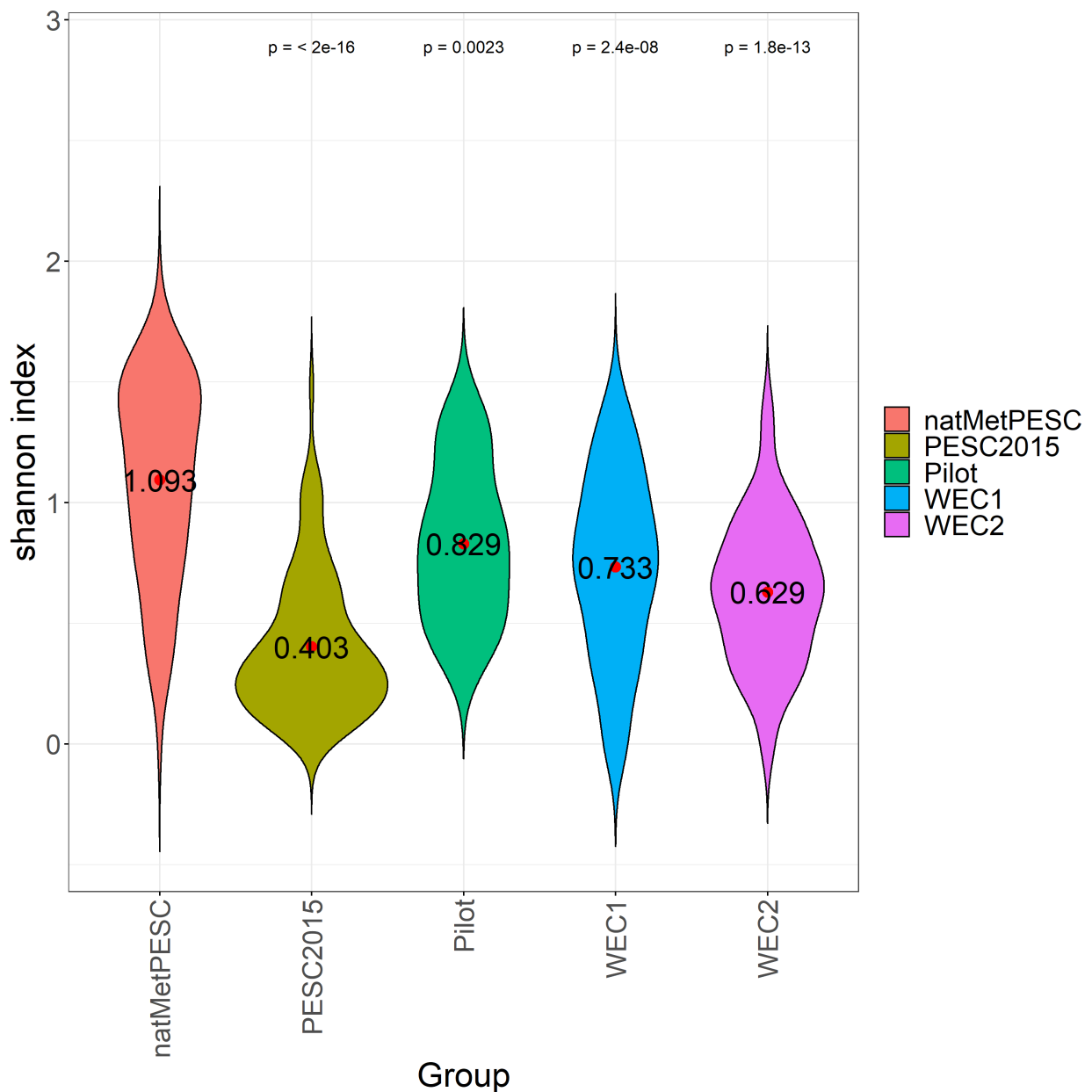
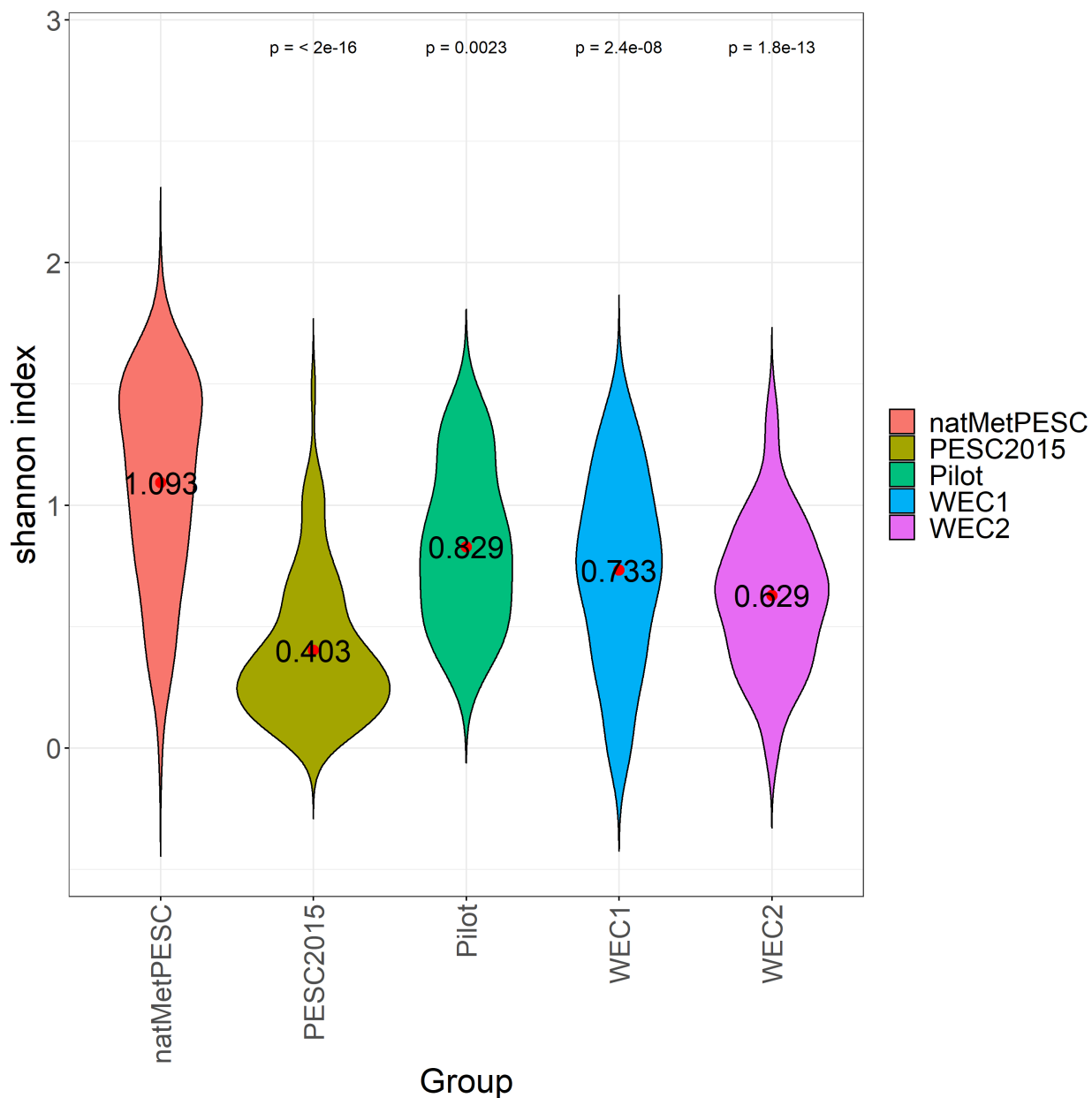


Figure 2: Community composition did not correlate with mass, size, or sex.

The effect of T3 and T4 Exposure on the *Lithobates catesbeianus* skin microbiome.

The skin microbiome were analysed based on 40 sequenced samples exposed to T3 or T4 (across premet, promet and froglets life stages).

The skin microbiota of the T3 and T4 exposed samples were dominated by proteobacteria and bacteroidetes. The exposure of T4 but not T3 led to an increase in alpha diversity measured by both Shannon H and Simpson index with an enrichment in *Caulobacter henricii* and an unclassified chlamydiaceae with T4 exposure and an unclassified enterobacteriaceae with T3 exposure [Fig. ???]. PCoA of UniFrac distances based on normalised OTU abundances did not show any clustering between the samples (Fig. [???], Fig. [???]).



PCoA of the T3/T4 treated micorbiomes. PCoA of the T3/T4 treated micorbiomes.

Meta-transcriptomic analysis of E2, T3, and T4 Exposure

From the counts generated from the tank water data set, primarily proteobacteria and Bacteroidetes were counted, specifically bacteria from the flavobacterium genus. Assigning bacterial taxonomies to de novo assembled transcripts from the TF and OE of tadpoles from the E2, T4, and T3 exposure sets revealed the presence of mostly bacteria from the proteobacteria phylum, with a smaller portion of bacterial transcripts mapping to gram positive phyla actinobacteria and firmicutes in both tissues (table 2, table 4). The OE had more bacterial transcripts that were significantly represented compared to the TF. Agrobacterium and Pseudomonas species exhibited the most varieties of transcripts of all present species. TFT3 exposure resulted in the differential expression of T226 16S ribosomal RNA that is assigned to Micrococcus luteus (Table 3). The M. luteus transcript exhibits a 6.5-fold change in transcript abundance (table 9). The M. luteus transcript exhibits similarity to an M. luteus 16S rRNA gene (Accession MT611279). No other treatments results in differential expression of other bacterial transcripts. Viral taxonomies were also assigned to assembled transcripts. There are fewer viral transcript assignments than bacterial in both tissues (table 5, table 7). The distribution of viral families assigned transcripts is very similar between tissues, with the majority belonging to Iridoviridae, followed by Adenoviridae, and Siphoviridae. Three viral species were commonly assigned to transcripts from both tissues: proteus phage VB PmiS-Isfahan, Frog virus 3, and human

mastadenovirus C. In the TF, T4 exposure results in the differential expressed of a single transcript assigned to the Frog virus 3 species (Table 7) with a 5.5-fold change of transcript abundance (table 13). Through a blastn query, this transcript is confirmed to encode for a hypothetical protein FV3gorf78L from the Frog virus 3 genome which contains an RNA recognition motif. T3 and E2 exposure did not result in the differential expression of any viral assigned transcripts in the TF. In the OE, both T4 and T3 exposure results in the differential expression of viral assigned transcripts from Proteus phage VB_PmiS-Isfahan, however, these are annotated with the human mitochondrial genome and are likely not viral transcripts (table 14-17). No significant results were found from filtering with the other RefSeq databases with Kraken2.

Discussion

The 16S sequencing results demonstrate a very slight response of microbial communities to T4 exposure (figure 2, figure 3), and only a single taxonomic unit being differentially present with the T3 exposure (figure 5). The assigning of bacterial taxonomies to expressed transcripts reveals a very slight response with T3 exposure in both tissues and T4 exposure did not result in the differential expression of any bacteria-assigned transcripts (table 3, table 5).

Assigning bacterial taxonomies to assemble transcripts revealed similar bacteria phyla as the 16S sequencing result. Both techniques capture mostly bacteria belonging to the phylum proteobacteria, with firmicutes captured as well. 16S sequencing uniquely captures the presence of Bacteroidetes and fusobacteria, where RNA-Seq methods capture the presence of actinobacteria.

From the 16S rRNA sequencing results, the two taxonomic units that are differentially present with T4 exposure are the bacteria species *Caulobacter henricii* and the family Chlamydiaceae. The Chlamydiaceae family are part of the phylum Chlamydiae which may only grow by infecting a eukaryotic host cell [4]. Bacteria from the genus *Caulobacter* oligotrophic bacterium commonly from fresh water lakes and streams with a dependency on phosphorus as a limiting nutrient [5]. The selective response of these bacteria to T4 exposure is not expected and potentially incidental as the counts of both taxonomic units are quite low.

Although no change in overall microbial diversity is detected with T3 exposure (figure 2), the 16S rRNA sequencing results do reveal a depletion of bacteria under the bacterial family Enterobacteriaceae (figure 5). This is a very large family that contains many common disease-causing bacteria such as *Salmonella*, *Escherichia coli*, and *Shigella*, as well as some non-pathogenic bacteria. The abundance of this taxonomic unit is quite substantial indicating that the response of this family may be an indirect response from T3 exposure to the tadpole host. From the microbial RNA-Seq method, only a single transcript is present in significant abundance in the OEE2, TFE2, and TFT4 exposure sets.

The microbial RNA-Seq only detected a difference in bacteria-assigned transcript response with the T3 exposure (table 3, table 5). The transcript assigned to *C. botulinum* transcript exhibits significant abundance in all exposure sets and appears to be responsive to T3 exposure in both tissues. Since the blastn annotation of this gene returns *X. laevis* rRNA as the top hits, it is unclear whether this transcript is really from a bacterial rRNA gene and has been misannotated in NCBI's database or is a frog gene and happens to have at least 40% of its k-mers being identical matches to *C. clostridium*. The transcript assigned to *M. luteus* is confirmed to be a segment of the 16S rRNA gene from that bacterial species with the blastn annotation. This bacteria is commonly found in soil, dust, water and air, and as part of the normal microbiota of the mammalian skin [???].

The microbial RNA-Seq method captured a variety of virus-assigned transcripts that cannot be captured with 16S sequencing methods (table 6, table 8). In this dataset, there is likely a high false positive from human contamination due to retroviruses in the virus database which may have gene segments that

match the human genome. The presence of mostly ranavirus represented in these data supports that this method is actually capturing the presence of viral transcripts (table 7). The differential expression of the frog virus 3 annotated transcript is the most interesting result from this investigation, although there is not much information available about this gene. The transcript is confirmed to be a viral-protein encoding gene by blastn annotation and exhibits a 5.5-fold change with T3 exposure in the TF (table 13). Frog virus 3 can be commonly isolated from healthy frogs and its role in frog disease is unknown [6].

Three different transcripts assigned to Proteus phage VB_PmiS-Isfahan are differentially expressed in either the OET3 or OET4. All these transcripts exhibit a roughly 0.5-fold change with either T3 or T4 exposure and are present in quite high abundance of 600 to 1800 median counts (table 14 - 17). These transcripts are likely false positives as they are annotated to be from human mitochondrial genomes following a blastn query. A transcript assigned to Human endogenous retrovirus K is also differentially expressed in the OE, with a 3-fold change in transcript abundance with T3 exposure (table 18). This transcript is in much lower abundance and although it was determined to be the HERV-K gene from this retrovirus, it is still likely that this is a false positive from human contamination.

Supplementary Information

Acknowledgements

The authors would like to thank their funders and the Simon Fraser University (SFU) Research Computing Group and Compute Canada for compute resource support.

Funding Information

B.J. hold Canadian Institutes of Health Research (CIHR) doctoral scholarships and Simon Fraser University (SFU) Omics and Data Sciences fellowships. F.S.L.B. holds an SFU Distinguished Professorship. Additionally, this work was partially supported by Genome Canada and NSERC grants to R.G.B. and F.S.L.B.

Conflict of Interest

The authors declare no competing interests.

References

References

1. **Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2**
Evan Bolyen, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian C. Abnet, Gabriel A. Al-Ghalith, Harriet Alexander, Eric J. Alm, Manimozhiyan Arumugam, Francesco Asnicar, ... J. Gregory Caporaso
Nature Biotechnology (2019-07-24) <https://doi.org/gf5292>
DOI: [10.1038/s41587-019-0209-9](https://doi.org/10.1038/s41587-019-0209-9) · PMID: [31341288](https://pubmed.ncbi.nlm.nih.gov/31341288/) · PMCID: [PMC7015180](https://pubmed.ncbi.nlm.nih.gov/PMC7015180/)
2. **DADA2: High-resolution sample inference from Illumina amplicon data**
Benjamin J Callahan, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy Jo A Johnson, Susan P Holmes
Nature Methods (2016-05-23) <https://doi.org/f84fxp>
DOI: [10.1038/nmeth.3869](https://doi.org/10.1038/nmeth.3869) · PMID: [27214047](https://pubmed.ncbi.nlm.nih.gov/27214047/) · PMCID: [PMC4927377](https://pubmed.ncbi.nlm.nih.gov/PMC4927377/)
3. **The SILVA ribosomal RNA gene database project: improved data processing and web-based tools**
Christian Quast, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, Frank Oliver Glöckner
Nucleic Acids Research (2012-11-27) <https://doi.org/gfb6mr>
DOI: [10.1093/nar/gks1219](https://doi.org/10.1093/nar/gks1219) · PMID: [23193283](https://pubmed.ncbi.nlm.nih.gov/23193283/) · PMCID: [PMC3531112](https://pubmed.ncbi.nlm.nih.gov/PMC3531112/)
4. **Metabolic Features of Protochlamydia amoebophila Elementary Bodies – A Link between Activity and Infectivity in Chlamydiae**
Barbara S. Sixt, Alexander Siegl, Constanze Müller, Margarete Watzka, Anna Wultsch, Dimitrios Tziotis, Jacqueline Montanaro, Andreas Richter, Philippe Schmitt-Kopplin, Matthias Horn
PLoS Pathogens (2013-08-08) <https://doi.org/f3snzh>
DOI: [10.1371/journal.ppat.1003553](https://doi.org/10.1371/journal.ppat.1003553) · PMID: [23950718](https://pubmed.ncbi.nlm.nih.gov/23950718/) · PMCID: [PMC3738481](https://pubmed.ncbi.nlm.nih.gov/PMC3738481/)
5. **Phylogeny and polyphasic taxonomy of Caulobacter species. Proposal of Maricaulis gen. nov. with Maricaulis maris (Poindexter) comb. nov. as the type species, and emended description of the genera Brevundimonas and Caulobacter**
Wolf-Rainer Abraham, Carsten Strömpl, Holger Meyer, Sabine Lindholst, Edward R. B. Moore, Ruprecht Christ, Marc Vancanneyt, B. J. Tindall, Antonio Bennisar, John Smit, Michael Tesar
International Journal of Systematic and Evolutionary Microbiology (1999-07-01) <https://doi.org/dm7dt5>
DOI: [10.1099/00207713-49-3-1053](https://doi.org/10.1099/00207713-49-3-1053) · PMID: [10425763](https://pubmed.ncbi.nlm.nih.gov/10425763/)
6. **Ecology of Viruses of Cold-Blooded Vertebrates**
V. GREGORY CHINCHAR
Elsevier BV (2000) <https://doi.org/bk6h2f>
DOI: [10.1016/b978-012362675-2/50012-4](https://doi.org/10.1016/b978-012362675-2/50012-4)