INVESTIGATING THE IMPACT OF XENOBIOTICS ON BACTERIAL COMMUNITIES USING NEXT GENERATION SEQUENCING STRATEGIES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Vienvilay Phandanouvong-Lozano

December 2019

INVESTIGATING THE IMPACT OF XENOBIOTICS ON BACTERIAL COMMUNITIES USING NEXT GENERATION SEQUENCING STRATEGIES

Vienvilay Phandanouvong-Lozano, Ph. D.

Cornell University 2019

The decline in DNA sequencing costs and the increased efficiency of next generation sequencing platforms have led to a flood of raw sequence data that taxes the analytical capacities of the current bioinformatic and statistical tools. The number of microbiome studies and the sequencing depth of each is likely to continue to grow given their tremendous potential to uncover important roles in ecosystem and human health. Microbiome data is highly complex; it is compositional, multi-dimensional, over-dispersed, and sparse. For these reasons, the analytical and statistical pipelines used to analyze these data should be carefully chosen to address both experimental design and the intrinsic features of the expected data. Here, I present three case studies to illustrate the use of deep 16S rRNA gene and metagenomic sequence analysis of environmental microbiomes. The first study assessed the impact of triclosan with and without biochar on soil bacterial communities. The second study used 16S rRNA amplicons to examine the effect of a glyphosate-based herbicide on the intestinal bacteria of mice. The last study used a shotgun-metagenomic approach to shed light on the genetic underpinnings of flurbiprofen degradation in a bacterial consortium enriched from active sludge.

BIOGRAPHICAL SKETCH

Vienvilay Phandanouvong Lozano was born on Donetsk, Ukraine and moved with her mother to Bogotá, Colombia at very young age. While pursuing her B.Sc. in Biology at the National University of Colombia in Bogotá, Colombia, she studied *Streptococcus pneumoniae* strains responsible of invasive pneumococcal diseases. Her undergraduate research experience inspired her to attain a research position at the Colombian Corporation of Agricultural Research (CORPOICA), where she studied the effect of essential oils on poultry intestinal bacteria. She also completed her Master's degree while at CORPOICA, studying the use of essential oils as food additives and nutraceuticals. She then began a position at the National Food and Drug Surveillance Institute, where she worked to characterize, detect, and quantify genetically modified organisms in food products and beverages. Upon deciding to pursue her Ph.D., she enrolled in the Microbiology program at Cornell University. She joined Dr. Anthony Hay's lab and began studying the impact of pharmaceuticals and other xenobiotics on bacteria and the microbiome. Her future plans include pursing a research career in the environmental and human health sector.

To my mom, my grandma, my great-grandma, to the Lozano women who taught me the true meaning of resilience, fierceness and courage

Para mi mamá, mi abuela, mi bisabuela, para las mujeres Lozano quienes me han enseñado lo que es el coraje, la fuerza y la ganas de seguir adelante

ACKNOWLEDGMENTS

Anthony Hay has been the best mentor I could have asked for throughout my graduate career. He is always willing to take the time to train me and has guided me on how to independently pursue my scientific interests. Anthony has not only trained me on how to achieve my doctoral degree, but has also taught and coached me for a research career, from lab work on the bench to scientific speech and writing. I am really thankful for his willingness to hear and discuss research questions and find solutions from different perspectives. I really appreciate his advice and corrections to guide me and keep track of my goals.

I would like to thank my minor advisors Ludmilla Aristilde and Joe Peters for their advice over my research projects and professional development. I am grateful to the undergraduates and visiting scholars in the lab for their help and constant dedication to the triclosan, glyphosate, and flurbiprofen projects.

I also want to say thank you to all Hay lab members who have been with me during this journey, to Hanh Nguyen who welcomed me in the lab and showed me the reins, to Jae Lyu for our great talks about research and life, and to recent visiting scholars from whom I learned their culture and shared their customs: thank you Sara Cantera, Mengmeng Wang, Shen Yan, Xiaoxia Liu, Zhanyong Wang.

I would also like to thank my mom Mery Lozano for introducing me to biology and always supporting my education. She has always been there for me and is a constant encouragement to keep moving forward. Gracias mamá por tu inmenso amor y enseñarme que con tenacidad y trabajo siempre salimos adelante. I am very grateful for my brother Diego Ballen with whom I shared a childhood that shaped us and brought us to accomplish our life goals, and my entire family for their continued love and encouragement both during my graduate school and all my life. Thank you tía Nana, you are my big sister. I would like to thank my fiancé Timothy Robinette for his unconditional love, support, and partnership. Tim has been the greatest support always patiently listening and giving me the best encouraging words to keep going and picture the final goal.

I would also like to thank my amazing group of friends from Corpoica, with them it was the perfect storm to start a research career, a great research team: Tatiana Rodriguez, Andrea Garcia, Carlos Ospina, Sonia Rodriguez, Felipe Vogelsan.

Last but not least, I want to thank to my friends with whom I shared this amazing journey of the doctoral degree. Thanks to Johana Uribe, Karina Acosta, and Marcela Patiño for their continued friendship and the enjoyable times we spent together.

TABLE OF CONTENTS

Chapter 5. Conclusions and Future Work

CHAPTER 1

INTRODUCTION TO NEXT GENERATION DNA SEQUENCING AND BIOINFORMATIC ANALYSIS

1.1 INTRODUCTION TO MICROBIOME DATA

Bacteria contribute to many metabolic processes throughout nature and depending on the type of interaction with the ecosystem or with the host, bacteria have been pictured as "good cop-bad cop". Important metabolic pathways are carried out by bacteria such as the synthesis of secondary metabolites and short-chain fatty acids (Flint et al., 2012; Nicholson et al., 2012). Likewise, bacteria have been used in a wide variety of industrial processes including the production of pharmaceuticals (Kieslich, 1986; Sanchez-Garcia et al., 2016). On the "bad cop" side of the coin, numerous diseases are caused by bacterial infections (Al-Anazi and Al-Jasser, 2014; Loch and Faisal, 2015; Rusin et al., 1997), and even bacterial imbalances in the gastrointestinal environment have been associated with multiple illnesses like obesity, inflammatory bowel disease, and type II diabetes (Hartstra et al., 2015; Huttenhower et al., 2014; Turnbaugh and Gordon, 2009).

Traditional microbiology methods generally refer to culture-dependent assays, but as has been widely demonstrated, these approaches are not entirely suitable to recreate relevant environmental conditions, revealing only culturable organisms and therefore underestimating the complexity of the bacterial communities (Lennon and Locey, 2016; Mallick et al., 2017). Since the introduction of DNA-based technologies, however, microbial ecology studies have shifted from classical cultured microbes to automated DNA sequencing, starting with Sanger, "first-

generation" technology, to next-generation sequencing (NGS) that has increase sequence coverage while lower cost per sequenced base (Cannavan et al., 2016; Metzker, 2010).

Largely as a result of cheap and cosmopolitan NGS, the study of microbial communities from diverse habitats "microbiome" has become part of the common vocabulary, often mentioned in various sources from research journals to newspapers to fashion magazines. By definition microbiome is a community of microorganisms sharing and living together in a particular habitat, which could include commensal, symbiotic, and pathogenic interactions with hosts (Bäumler and Sperandio, 2016; Peterson et al., 2009). Microbiome studies typically result in lists of microbial taxa, metagenomic or transcriptomic data, and even assembled genomes. As a result, there are more lists of taxonomically defined microbial taxa than actual knowledge about the role of these microbes in their environment. One of the biggest challenges ahead is to translate data from the microbiome studies to defined microbial-associations and causality responses (Surana and Kasper, 2017). Insights gained from microbiome studies range from associations between the microbiome and the host, whether it is linked to dysbiosis or not, to identifying environmental factors and their mechanisms that shape or disturb microbial communities (Claus et al., 2016; Huttenhower et al., 2014; Norman et al., 2015; Roh et al., 2009).

The decline in DNA sequencing costs and the increased efficiency of NGS platforms has led to a flood of raw sequence data, stimulating the development of complex and sophisticated bioinformatic and statistical tools to analyze the wealth of data generated (Liu et al., 2012; Mataragas et al., 2018). Therefore, a clear understanding of the key concepts required to design, execute and interpret NGS experiments on microbiomes is imperative (Jovel et al., 2016). There are two main approaches for analyzing the microbiome: 16S ribosomal RNA (rRNA) gene sequencing and shotgun metagenomics.

1.2 16S rRNA GENE SEQUENCING APPROACH

The 16S rRNA gene sequencing, a form of amplicon sequencing, refers to the PCR-based amplification and sequencing of the conserved 16S rRNA gene, which is used to identify and classify microbial taxa (Franks et al., 1998; Lane et al., 1985; Pace, 2009; Schulz et al., 2017).

The 16S rRNA gene is useful for microbiome characterization due to its ubiquity and essential role in cell metabolism (Pace, 2009; Stern et al., 1989). Regardless of the known biases of the 16S rRNA gene primers and the PCR process itself, the 16S rRNA gene still offers a fair taxonomic coverage for use as a bacterial marker (Kembel et al., 2012; Pace, 2009; Rosselli et al., 2016). The 16S rRNA gene sequences that act as unique identifiers correspond to nine hypervariable regions (V1-V9). The V1-V3, V4, and V4-V5 regions are the most commonly used given its genus-level resolution (Langendijk et al., 1995; Yang et al., 2016) and allow for reasonable certainty in the placement of observed bacteria taxa on the tree of life (Pace, 2009; Schulz et al., 2017). These highly conserved regions are also optimal for the design of "universal primers" to target most, but not all bacteria in a primer-dependent fashion (Fouhy et al., 2016; Klindworth et al., 2013; Nübel et al., 1997). The massive parallel sequencing approach of the 16S rRNA gene was the first to be widely accepted and implemented, and the current databases are very well curated allowing accurate taxonomic comparisons. To date, there are three large-scale curated databases: SILVA (Quast et al., 2013), GreenGenes (McDonald et al., 2012), and RDP (Ribosomal Database Project) (Cole et al., 2014). The latter offers web-based tools to construct phylogenetic trees and alignments. The SILVA database, on the other hand, has the most complete catalogue that not only includes the bacterial small ribosomal subunit sequences (16S SSU), but also the eukaryotic large ribosomal subunit sequences (LSU). Using SILVA

database requires longer computational time than the smaller database GreenGenes; though, GreenGenes has not been updated since 2013.

1.2.1 BIOINFORMATIC ANALYSIS OF 16S rRNA GENE SEQUENCING DATA

Among the intrinsic variables to consider during processing of 16S rRNA amplicon data are the type of sample, sample collection, experimental design, and temporal variability specifically for longitudinal studies. As part of the experimental design and library preparation, the variable region should be chosen carefully confirming that it has the desired taxonomic resolution to distinguish the bacteria taxa of interest in a specific microbiome. Primers appropriate for one environment might underreport on the diversity of another. For example, the phylogenetic breadth of a soil, for example, is much broader than that a human fecal sample. Even "good" universal primers should be routinely reevaluated to ensure that they are modified to include newly characterized taxa identified via other means including metagenomics (Gantner et al., 2011; Klindworth et al., 2013; Mao et al., 2012; McDonald et al., 2012). Mock communities and/or reference data should also be included in the experimental design to ensure correct identification and abundance estimation of the bacterial communities (Barb et al., 2016; Egan et al., 2018; May et al., 2014).

Quality control (QC) analysis is the first step in the bioinformatic pipeline of the 16S rRNA amplicon sequencing approach. QC is considered a pre-processing step of the raw reads and is required to eliminate sequencing artifacts; it prevents overestimation and improves accuracy (Kozich et al., 2013; Parks et al., 2015; Xue et al., 2018). First, artificial chimeric sequences are removed; then short and low-quality reads are filtered; and lastly error-laden sequences considered as noise are separated from true sequence data (Haas et al., 2011; Lee et al., 2017; Schloss et al., 2011).

Once the reads are pre-processed, a taxonomic classification can be assigned by using an existing phylogenetic tree (supervised, taxonomy-dependent approach), or clustering into Operational Taxonomic Units (OTUs, unsupervised, taxonomy-independent approach) (Goodrich et al., 2014). While in the taxonomy-dependent approach the reads are directly assigned to the pre-existing phylogeny (phylotypes); the OTU approach clusters the reads by sequence similarity to one another first and only then assigns taxonomic classifications by comparing the clusters to a reference database (Berry et al., 2012; Caporaso et al., 2010; Schloss et al., 2009). The taxonomy classification tools include the commonly known BLAST (Basic Local Alignment Search Tool), the online GreenGenes, and multimer clustering tree-based methods (Altschul et al., 1990; Liu et al., 2008; McDonald et al., 2012). The phylotype approach is computationally more efficient and relates the sequencing data to previously identified bacteria, however the taxonomy assignment depends highly on the reference database and requires high accuracy during sequencing. Thus, the phylotype approach might prevent the detection of novel sequences of unclassified taxonomic lineages especially in poorly characterized microbiomes (Westcott and Schloss, 2015). The OTU approach, on the other hand, allows the *de novo* identification of taxonomic groups and is widely used in microbiome studies of habitats that are highly diverse, but is computationally more demanding.

1.2.2 CHALLENGES AND LIMITATIONS OF 16S rRNA GENE SEQUENCING

Since the 16S rRNA amplicon sequencing approach requires PCR amplification, OTU or bacteria taxa could be misidentified or not detected due to PCR-associated biases, e.g. priming or mismatch nucleotides, resulting in underestimation of bacterial diversity and richness (Acinas et al., 2005; Kennedy et al., 2014; Laursen et al., 2017; Mao et al., 2012). Alternatively, bacterial abundance might be overestimated because of the variation of 16S copy number among bacteria

species, or due to artifacts during sequencing or read processing like incorrect alignments to the reference databases (Kembel et al., 2012; Větrovský and Baldrian, 2013). One main limitation of the 16S rRNA amplicon sequencing approach is that the bacterial metabolic functions cannot be determined directly, the sequencing data only provide taxonomic composition. Predictive functional profiling tools such as PICRUSt and Tax4Fun have been recently developed, which infer community function by using the sequenced genomes of organisms whose 16S match those in the amplicon library (Aßhauer et al., 2015; Langille et al., 2013). While these demonstrate significant correlation with shotgun metagenomic studies of samples from well-characterized environments, they are unlikely to work as well using 16S rRNA gene sequencing data from poorly characterized environments that include many uncultured bacteria for whom no reference genomes are available (Aßhauer et al., 2015; Langille et al., 2013). Recent advances in the field of metagenome assembled genomes (discussed below), however, may help to overcome these limitations by providing reference genome assemblies that can inform functional predictions (Parks et al., 2017; Tully et al., 2018).

1.3. SHOTGUN METAGENOMIC SEQUENCING APPROACH

In contrast to the 16S rRNA gene sequencing approach, shotgun metagenomic sequencing permits analysis of all genomic DNA in the community gaining insights of the metabolism and gene function through functional classification. By direct DNA extraction, the genetic information of a bacterial community can be independently sequenced and assembled (Sharpton, 2014). Although the bacterial composition of microbiomes can also be determined from taxonomically informative genes (i.e. 16S rRNA, *rpoB*), metagenomic studies seek to answer questions related to functional roles that could be shared or be co-dependent between members of the microbiome (Dai et al., 2018; Velsko et al., 2018). Metagenomic studies provide

information that allows comparison of functional profiles and community dynamics (including growth rates) of microbiomes of such as healthy host versus disease-exposed host, exposure versus no-exposure to a specific contaminant, wild-type versus knockout mice, etc. Metagenomics is a powerful tool that provides answers to questions including who is there?, how do they relate to each other?, what do they do?, and how are they changing?. Hence, through shotgun metagenomic sequencing not only the bacterial community composition is described, but also the genetic potential at the community and individual level (Bowers et al., 2017; Noyce et al., 2016; Pedron et al., 2019).

1.3.1. BIOINFORMATIC ANALYSIS OF SHOTGUN METAGENOMIC DATA

Similarly to 16S rRNA gene sequencing, sample collection, DNA extraction, and experimental design are critical to obtain high-quality reads from the shotgun sequencing data. The extracted DNA should be of high quality and sufficient to provide enough reads that would be aligned to various genomic locations from the myriad microorganisms present in a sample, thus sampling depth is a current limitation for most studies and usually only permits analysis of the numerically dominant organisms (>1%) (Bag et al., 2016; Costa et al., 2017; Lagier et al., 2012). A quality check (QC) step is also performed with the raw reads of the whole metagenome shotgun sequencing. Then, filtered reads are assembled into longer sequences (contigs) that would facilitate the final assembly (Parks et al., 2015; Sharpton, 2014). There are two assembly approaches: reference-based and *de novo*. As in the 16S rRNA gene sequencing approach, when a reference is used (comparable to the phylotype-approach in 16S amplicon), the assembly depends on the similarity of the metagenome reads to sequences already in a database (Alneberg et al., 2018). The reference-based approach performs poorly if there are polymorphisms or large deletions or insertions in the genomes. The *de novo* assembly, however, overcomes these

genomic limitations. The biggest disadvantage of the *de novo* assembly is the large computational cost which varies depending on the microbiome complexity and the type of assembler (Nurk et al., 2017; Vollmers et al., 2017). A *de novo* assembly might take weeks to run.

The next step is binning where the assembled contigs are sorted into groups that might correspond to an individual genome, or closely related organisms based on distinctive sequence features. Several binning algorithms have been recently developed and are mostly based in similarity like hierarchical clustering, and compositional-based binning using for instance the conserved nucleotide composition such the frequency of pentanucleotide repeats for example (Sieber et al., 2018; Wang et al., 2019; Wu et al., 2014). After binning, a quality check is again performed evaluating the completeness (base on percent coverage of required single copy genes) and purity of the taxon bin (Parks et al., 2015). After another round of assembly, the metagenome-assembled genomes (MAG) are annotated using tools like RAST (Rapid Annotation using Subsystem Technology) and IMG (Integrated Microbial Genomes system) that use gene prediction and functional annotation from previously annotated and characterized genes (Keegan et al., 2016; Markowitz et al., 2012; Overbeek et al., 2014; Randle-Boggis et al., 2016).

1.3.2. CHALLENGES AND LIMITATIONS OF SHOTGUN METAGENOMIC SEQUENCING

In comparison to the 16S rRNA gene sequencing, shotgun metagenomic sequencing is more expensive, requiring more computational power and data storage. The bioinformatic pipeline is more complex and the metagenomic tools (assemblers, binning) are still in development, but various tools are become increasingly available that are easy to use and have good accuracy for

short-raw-reads like those produced by the most commonly used sequencers from Illumina (Ayling et al., 2019; Escobar-Zepeda et al., 2015; Teeling and Glöckner, 2012).

Genome assembly and binning process have trouble distinguishing between closely related genomes and individual heterogeneity which could prevent a true metagenome assembly (Ghurye et al., 2016). Sequencing depth is also one of most limiting factors in the shotgun metagenomic sequencing analysis, most assemblers required a minimal sequence coverage that can be out of reach economically for projects with many samples or those with smaller budgets (Nurk et al., 2017; Wu et al., 2014).

1.4. STRUCTURE OF MICROBIOME DATASET

At the end of the sequencing analysis pipeline, the results are organized in a table format reporting OTU's abundances which could come in the form of 16S rRNA or other gene sequence abundances from either the amplicon or the shotgun metagenomic sequencing. The features associated with these tables are not only the treatments, conditions, or temporal variabilities of the experimental design, there are also features associated to the OTUs or genes which are highly dimensional to themselves. For OTUs dimensionality is a function of nested taxonomic levels, and for the annotated genes it is a function of the metabolic process they encode. From a practical standpoint, this means that it may be more meaningful to compare a group of related organisms or metabolic functions rather than individual OTUs or genes.

1.4.1 FEATURES OF MICROBIOME DATA

Given that the microbiome data is highly dimensional and includes many poorly-defined characters (unclassified OTUs or hypothetical proteins), analysis of the microbiome as a whole and of microbiome-associations is a statistical and bioinformatics challenge (Jovel et al., 2016).

The count data for given characters (OTU or gene) is also over-dispersed, which means that the variance of the read count is larger than the predicted by traditional multinomial regression models (Sun et al., 2017). Given that this over-dispersion has several sources including the randomness of the sequencing process (DNA extraction, PCR, library preparation), in addition to the intrinsic variability of the microbiome like between intestinal samples from the same individual, it is hard for traditional statistical distributions to control for the variability (McMurdie and Holmes, 2014).

Microbiome data is considered "sparse" because, even though it is typically based on 10s or even 100s of thousands of observations (sequences), many environments are highly diverse and microbes are seldom normally dispersed in space or time. This results in the detection of unique OTUs or rare taxa in some samples but not others and an inflated number of zeros in the count tables, especially in bacterial groups at lower taxonomic levels (genus and below). The presence of an inflated number of zeros poses a challenge for the statistical analysis affecting statistical transformation methods and the variance among replicates, pulling results below that of expected data. From a statistical point of view, it is truly challenging to differentiate between a sampling zero (as result of data handling) from a structural zero (real zero) (Jonsson et al., 2018; Peng et al., 2016).

1.4.2 MODELING MICROBIOME DATA

When analyzing and modeling microbiome data, the chosen statistical tools in addition to accounting for controls and replicates of the experimental design, should address the following: (1) the dimensionality of the phylogenetic information associated to the OTUs/bacterial taxa or the metabolic processes associated to the annotated genes, (2) the variable's uniqueness

(sparsity) measured by inflated number of zeros (e.g. rare taxa), and (3) the intrinsic variability reflected in over-dispersed statistical variance.

Bioinformatic pipelines that process raw sequence reads such as Mothur and QIIME while comparing and grouping the reads by similarity, can also assign phylogenetic or functional categories to the groups of reads (Keegan et al., 2016; Kozich et al., 2013; Kuczynski et al., 2011). Then, depending on the category of interest the count table is agglomerated, and the abundances are pulled out. In addition, software like Phyloseq, which is an R package suitable to work with output files from Mothur or QIIME, permits handling the assigned categories post-raw-read-processing (McMurdie and Holmes, 2013).

The sparsity, over-dispersion, and undersampling of microbiome data challenge statistical parametric models since low-abundant taxa and zero inflation especially at lower taxonomic levels (genus and below) result in a non-normal skewed distribution of the OTU's abundances and of the OTU's occurrence probabilities. Thus, parametric models such as normal, binomial, and Poisson distribution are not appropriate to model and analyze microbiome data. Negative binomial (NB) and zero inflated models, however, are better fit and often applied in microbiome count data (McMurdie and Holmes, 2014). Analyses based on NB models have proven to be successful finding differentially abundant features via simulation as well as in longitudinal studies (Zhang et al. 2018). DESeq2, which was the R package used in Chapter 3 (*"Sex-dependent disturbance effect of glyphosate on gut microbiome"*), is based on NB distribution and corrects the data variability by using local regression within the dynamics range. In addition to 16S rRNA amplicon libraries, DESeq2 can also be used to compare sequence counts and perform differential analysis of high-throughput RNA-Seq data (Anders and Huber 2010).

White et al. (2009) also proposed and implemented a non-parametric statistical method called Metastats to detect differentially abundant taxa. In this method, the microbiome data is

transformed into proportions and the two-sample comparisons are performed using the Storey and Tibshirani's permutation and are corrected for the false discovery rate (2003). In Metastats, the differentially abundant sparse counts are compared using a Fisher's exact test, which has shown to outperform the Student's t-test and log-linear models (White et al. 2009). Improvements in the Metastats analysis using a mixed-model zero-inflated Gaussian distribution have also showed higher accuracy and reduced intrinsic covariance in simulated data and clinical metagenomic samples (Paulson et al., 2011; Jonsson et al., 2016). The Metastats analysis was used in Chapter 2 (*"Biochar does not attenuate triclosan's impact on soil bacterial communities"*, Phandanouvong-Lozano et al., 2018) to compare bacterial abundances given that our qualitative analysis suggested that triclosan impacted shared OTUs among the bacterial communities, instead of selecting unique OTUs as observed in the glyphosate's effect in Chapter 3 (*"Sex-dependent disturbance effect of glyphosate on gut microbiome"*). High number of unique OTUs results in non-normal skewed distributions, which as discussed above are implemented in bioinformatic analyses like DESeq2. For this reason, DESeq2 was a more suitable tool to compare the bacterial communities in Chapter 3.

In order to answer questions related to causality and quantify changes in longitudinal studies, prospective statistical tools should link the microbiome shifts to the independent variables such as to the host status (e.g. health or disease), or to a temporal exposure (short-term vs. long-term). The linear mixed-effects model allows for testing the probability of causal and temporal relationships with the response variable (microbiome data) adjusting for confounding factors like over-dispersion and replicates' intrinsic variability. The linear mixed-effects model was implemented in Chapters 2 and 3 through the lme4 and emmeans R packages (Bates et al. 2015). Similar linear mixed-models have been used to analyze associations of oral and gut

microbiome with temporal and clinical variables such us progression of type 1 diabetes and respiratory diseases (Kostic et al., 2015; Tipton et al., 2018).

Hence based on the challenges inherent to the microbiome data, the analytical and statistical pipelines used to analyze next generation sequencing reads should be carefully chosen to address both experimental design and the intrinsic features of the expected microbiome data.

1.4.3 ANALYZING MICROBIOME DATA AS COMPOSITIONAL BY NATURE

As discussed briefly in section 1.4.1 technical limitations on sequencing depth and data analytics make the treatment of zeros problematic. From a practical standpoint, non-exhaustive sequencing and the arbitrary number of total raw reads per sample mean that each sequencing run merely subsamples part of the DNA sequences in a sample rather than performing a true census (Gloor et al., 2017). For these reasons, an increasing number of researchers have started to analyze microbiome data as compositional in nature (Kurtz et al., 2015; Mandal et al., 2015; Rivera-Pinto et al., 2018). The microbiome as compositional count data has two geometric properties. First, the library size which is the total number of reads per sample is an artifact of the sampling procedure affected by technical and experiment-specific variability. Second, the compositional data is proportional to the sequencing capacity of the instrument and to the relative abundance underlying in the sample (Tsilimigras and Fodor, 2016).

Various statistical tools have been recently developed to implement compositional analysis on microbiome data. These tools generally use log-ratio transformation to constrain the compositional nature of the microbiome data and use proportionality for substituting spurious correlations as in the case when estimated relative abundances make independent components appear correlated (Aitchison and Egozcue, 2005; Lovell et al., 2011). Since the compositional analysis considers the raw reads as proportions, the libraries of the microbiome data are not

normalized to a specific size (Gloor et al., 2017). Then, distance metrics such as Aitchison's and ordination analysis such as PCA (Principal Component Analysis) can be implemented to ensure correct treatment of the compositional data (Aitchison and Egozcue, 2005; Gloor et al., 2017). Currently, there are available two statistical tools to evaluate differential abundances in the microbiome data, ANCOM (Analysis of Composition of Microbiomes) and ALDEx2. These tools are based on compositional log-ratios and use an ANOVA-like framework (Fernandes et al, 2013; Mandal et al., 2015; Gloor et al., 2016).

Although compositional statistical tools have been increasingly incorporated in microbiome studies, analysis of compositional data is still a challenge because of the multivariate nature of microbiome datasets because there is currently no universally accepted solution for the issue of inflated numbers of zeros associated with microbiome features like unique or rare OTUs (e.g. log-ratio of zero values) (Knight et al., 2018; Weiss et al., 2017). Some authors argue that relative abundance estimates, normalized, or rarefied counts implicitly acknowledged the compositional nature of the microbiome (Weiss et al. 2017), however, by rarefying or normalizing to an arbitrary library size some microbiome data can be discarded (McMurdie and Holmes 2014). Statistical packages like edgeR, metagenomeSeq, and DESeq (explained above) which are based on Negative Binomial and mixture models considering raw counts as absolute values appear to offer an appropriate method to infer differential abundances in the microbiome data (McMurdie and Holmes 2014).

How much does it really matter? In the end, the relative impact of more traditional analytics versus compositional approaches depends on the features of the experimental design and the nature of the microbiome that is investigated (Knight et al., 2018). From the compositional analysis perspective, microbiome data is handled as proportions, therefore, the total sum of the counts is 1 and the research interest is towards the ratios between the

components more than in the absolute difference between the observations (Fernandes et al. 2014; Martín-Fernández et al. 2015). When the absolute counts and total sum between the components are relevant for the research interest, distance metrics and statistical methods like Unifrac, PCoA and DESeq are valid and reproducible (Knight et al., 2018, Weiss et al., 2017).

It is likely that bioinformaticians and statisticians will continue to argue about the correct treatment of microbiome data for years to come, in much the same way as statistical ecologists continue to argue about appropriate dissimilarity functions (which themselves remain a topic of debate in the microbiome literature). I will therefore leave the reader with this food for thought on the topic from Ricotta and Podani (2017) in the hope that it will help them make a justifiable, though imperfect choice about their statistical treatment of microbiome data.

"Although dissimilarity [insert 'correct statistical model'] may appear an intuitively simple concept, there is no single, unequivocal way for its measurement. The literature of numerical ecology *treats many more, even hundreds of dissimilarity functions and selection among them is often arbitrary, dictated by fashion, availability in commercial software or personal preference. The choice of a dissimilarity index best suited for a specific ecological problem is a complex question which does not have clear and unambiguous answer."*

1.5. REFERENCES

Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., Polz, M.F., 2005. PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample. Appl. Environ. Microbiol. 71, 8966–8969. https://doi.org/10.1128/AEM.71.12.8966-8969.2005

Aitchison, J., Egozcue, J.J., 2005. Compositional Data Analysis: Where Are We and Where Should We Be Heading? Mathematical Geology. 37, 829–850. https://doi.org/10.1007/s11004-005-7383-7

Al-Anazi, K.A., Al-Jasser, A.M., 2014. Infections Caused by Stenotrophomonas maltophilia in Recipients of Hematopoietic Stem Cell Transplantation. Front. Oncol. 4. https://doi.org/10.3389/fonc.2014.00232

Alneberg, J., Sundh, J., Bennke, C., Beier, S., Lundin, D., Hugerth, L.W., Pinhassi, J., Kisand, V., Riemann, L., Jürgens, K., Labrenz, M., Andersson, A.F., 2018. BARM and BalticMicrobeDB, a reference metagenome and interface to meta-omic data for the Baltic Sea. Sci. Data 5, 180146. https://doi.org/10.1038/sdata.2018.146

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Aßhauer, K.P., Wemheuer, B., Daniel, R., Meinicke, P., 2015. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics 31, 2882–2884. https://doi.org/10.1093/bioinformatics/btv287

Ayling M., Clark, M.D., Leggett, R.M., 2019. New approaches for metagenome assembly with short reads. Brief. Bioinformat. bbz020. https://doi.org/10.1093/bib/bbz020

Bag, S., Saha, B., Mehta, O., Anbumani, D., Kumar, N., Dayal, M., Pant, A., Kumar, P., et al., 2016. An Improved Method for High Quality Metagenomics DNA Extraction from Human and Environmental Samples. Sci. Rep. 6, 26775. https://doi.org/10.1038/srep26775

Barb, J.J., Oler, A.J., Kim, H.-S., Chalmers, N., Wallen, G.R., Cashion, A., Munson, P.J., Ames, N.J., 2016. Development of an Analysis Pipeline Characterizing Multiple Hypervariable Regions of 16S rRNA Using Mock Samples. PLOS ONE 11, e0148047. https://doi.org/10.1371/journal.pone.0148047

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015.Fitting Linear Mixed-Effects Models using lme4. J. Statistic. Soft. 67, 1. https://doi.org./ doi: 10.18637/jss.v067.i01

Bäumler, A.J., Sperandio, V., 2016. Interactions between the microbiota and pathogenic bacteria in the gut. Nature 535, 85–93. https://doi.org/10.1038/nature18849

Berry, D., Schwab, C., Milinovich, G., Reichert, J., Ben Mahfoudh, K., Decker, T., Engel, M., Hai, B., et al., 2012. Phylotype-level 16S rRNA analysis reveals new bacterial indicators of health state in acute murine colitis. ISME J. 6, 2091–2106. https://doi.org/10.1038/ismej.2012.39

Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., et al., 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat. Biotechnol. 35, 725–731. https://doi.org/10.1038/nbt.3893

Cannavan, F.S., Nakamura, F.M., Germano, M.G., de Souza, L.F., Tsai, S.M., 2016. Chapter 5 - Next-Generation Sequencing to Elucidate Biochar-Effected Microbial Community Dynamics, in: Biochar Application. Elsevier, pp. 109–132. https://doi.org/10.1016/B978-0-12-803433-0.00005-9

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

Claus, S.P., Guillou, H., Ellero-Simatos, S., 2016. The gut microbiota: a major player in the toxicity of environmental pollutants? Npj Biofilms Microbiomes 2, 16003. https://doi.org/10.1038/npjbiofilms.2016.3

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, et al., 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42, D633–D642. https://doi.org/10.1093/nar/gkt1244

Costa, V., Rosenbom, S., Monteiro, R., O'Rourke, S.M., Beja-Pereira, A., 2017. Improving DNA quality extracted from fecal samples—a method to improve DNA yield. Eur. J. Wildl. Res. 63, 3. https://doi.org/10.1007/s10344-016-1058-1

Dai, D., Rhoads, W.J., Edwards, M.A., Pruden, A., 2018. Shotgun Metagenomics Reveals Taxonomic and Functional Shifts in Hot Water Microbiome Due to Temperature Setting and Stagnation. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.02695

Egan, C.P., Rummel, A., Kokkoris, V., Klironomos, J., Lekberg, Y., Hart, M., 2018. Using mock communities of arbuscular mycorrhizal fungi to evaluate fidelity associated with Illumina sequencing. Fungal Ecol. 33, 52–64. https://doi.org/10.1016/j.funeco.2018.01.004

Escobar-Zepeda, A., Vera-Ponce de Leon, A., Sanchez-Flores, A., 2015. The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. Front. Genet. 6, 348. https://doi.org/10.3389/fgene.2015.00348

Fernandes, A.D., Macklaim, J.M., Linn, T.G., Reid, G., Gloor, G.B., 2013. ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. PLoS ONE 8, e67019. https://doi.org/10.1371/journal.pone.0067019

Fernandes, A.D., Reid, J.N.S., Macklaim, J.M., McMurrough, A.M., Edgell, D.R., Gloor, G.B., 2014. Unifying the Analysis of High-Throughput Sequencing Datasets: Characterizing RNA-seq, 16S rRNA Gene Sequencing and Selective Growth Experiments by Compositional Data Analysis. Microbiome. 2, 15

Flint, H.J., Scott, K.P., Louis, P., Duncan, S.H., 2012. The role of the gut microbiota in nutrition and health. Nat. Rev. Gastroenterol. Hepatol. 9, 577–589. https://doi.org/10.1038/nrgastro.2012.156

Fouhy, F., Clooney, A.G., Stanton, C., Claesson, M.J., Cotter, P.D., 2016. 16S rRNA gene sequencing of mock microbial populations- impact of DNA extraction method, primer choice and sequencing platform. BMC Microbiol. 16, 123. https://doi.org/10.1186/s12866-016-0738-z

Franks, A.H., Harmsen, H.J.M., Raangs, G.C., Jansen, G.J., Schut, F., Welling, G.W., 1998. Variations of Bacterial Populations in Human Feces Measured by Fluorescent In Situ Hybridization with Group-Specific 16S rRNA-Targeted Oligonucleotide Probes. Appl. Environ. Microbiol. 64, 3336–3345.

Gantner, S., Andersson, A.F., Alonso-Sáez, L., Bertilsson, S., 2011. Novel primers for 16S rRNA-based archaeal community analyses in environmental samples. J. Microbiol. Methods 84, 12–18. https://doi.org/10.1016/j.mimet.2010.10.001

Ghurye, J.S., Cepeda-Espinoza, V., Pop, M., 2016. Metagenomic Assembly: Overview, Challenges and Applications. Yale J. Biol. Med. 89, 353-362.

Gloor, G.B., Wu, J.R., Pawlowsky-Glahn, V., Egozcue, J.J., 2016. It's all relative: analyzing microbiome data as compositions. Annals of Epidemiol. 26, 322-329. https://doi.org/10.1016/j.annepidem.2016.03.003

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J., 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. Front. Microbiol. 8, 2224. https://doi.org/10.3389/fmicb.2017.02224

Goodrich, J.K., Di Rienzi, S.C., Poole, A.C., Koren, O., Walters, W.A., Caporaso, J.G., Knight, R., Ley, R.E., 2014. Conducting a microbiome study. Cell. 158, 250-262. https://doi.org/10.1016/j.cell.2014.06.037

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., et al., 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 21, 494–504. https://doi.org/10.1101/gr.112730.110

Hartstra, A.V., Bouter, K.E.C., Bäckhed, F., Nieuwdorp, M., 2015. Insights into the role of the microbiome in obesity and type 2 diabetes. Diabetes Care 38, 159–165. https://doi.org/10.2337/dc14-0769

Huttenhower, C., Kostic, A.D., Xavier, R.J., 2014. Inflammatory bowel disease as a model for translating the microbiome. Immunity 40, 843–854. https://doi.org/10.1016/j.immuni.2014.05.013

Jonsson, V., Österlund, T., Nerman, O., Kristiansson, E., 2016. Statistical evaluation of methods for identification of differentially abundant genes in comparative metagenomics. BMC Genomics 17, 78. https://doi.org/10.1186/s12864-016-2386-y

Jonsson, V., Österlund, T., Nerman, O., Kristiansson, E., 2018. Modelling of zero-inflation improves inference of metagenomic gene count data. Stat. Methods Med. Res. 962280218811354. https://doi.org/10.1177/0962280218811354

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A.L., Madsen, K.L., Wong, G.K.-S., 2016. Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.00459

Keegan, K.P., Glass, E.M., Meyer, F., 2016. MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function, in: Martin, F., Uroz, S. (Eds.), Microbial Environmental Genomics (MEG), Methods in Molecular Biology. Springer New York, New York, NY, pp. 207–233. https://doi.org/10.1007/978-1-4939-3369-3_13

Kembel, S.W., Wu, M., Eisen, J.A., Green, J.L., 2012. Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance. PLoS Comput. Biol. 8. https://doi.org/10.1371/journal.pcbi.1002743

Kennedy, K., Hall, M.W., Lynch, M.D.J., Moreno-Hagelsieb, G., Neufeld, J.D., 2014. Evaluating Bias of Illumina-Based Bacterial 16S rRNA Gene Profiles. Appl. Environ. Microbiol. 80, 5717–5722. https://doi.org/10.1128/AEM.01451-14

Kieslich, K., 1986. Production of drugs by microbial biosynthesis and biotransformation. Possibilities, limits and future developments (1st communication). Arzneimittelforschung. 36, 774–778.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner, F.O., 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 41, e1–e1. https://doi.org/10.1093/nar/gks808

Knight, R., Vrbanac, A., Taylor, B.C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., et al., 2018. Microbiome. 16, 410-422. https://doi.org/10.1038/s41579-018-0029-9

Kostic A.D., Gevers, D., Siljander, H., Vatanen T., Hyotylainen T., Hamalainen, A.M., Peet, A., Tilmann, V., 2015. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell Host Microbe. 17, 260-273. https://doi.org/ doi: 10.1016/j.chom.2015.01.001

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. Appl. Environ. Microbiol. 79, 5112–5120. https://doi.org/10.1128/AEM.01043-13

Kuckzynski, J., Stombaugh, J., Walters, W.A., Gonzalez, A., Caporaso J.G., Knight R., 2011. Using QIIME to analyze 16S rRNA gene sequences from Microbial Communities. Curr Protoc Bioinformat. 10.7.1-10.7.20. https://doi.org/10.1002/0471250953.bi1007s36

Kurtz, Z.D., Müller, C.L., Miraldi, E.R., Littman D.R., Blaser M.J., Bonneau R.A., 2015. Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput. Biol. 11, e1004226. https://doi.org/10.1371/journal.pcbi.1004226

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., Pace, N.R., 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. U. S. A. 82, 6955–6959.

Lagier J., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., Bittar, F., Fournous, G., et al., 2012. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin. Microbiol. Infect. 18: 1185-1193. https://doi.org/10.1111/1469-0691.12023

Langendijk, P.S., Schut, F., Jansen, G.J., Raangs, G.C., Kamphuis, G.R., Wilkinson, M.H., Welling, G.W., 1995. Quantitative fluorescence in situ hybridization of Bifidobacterium spp. with genus-specific 16S rRNA-targeted probes and its application in fecal samples. Appl. Environ. Microbiol. 61, 3069–3075.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., et al., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol. 31, 814–821. https://doi.org/10.1038/nbt.2676

Laursen, M.F., Dalgaard, M.D., Bahl, M.I., 2017. Genomic GC-Content Affects the Accuracy of 16S rRNA Gene Sequencing Based Microbial Profiling due to PCR Bias. Front. Microbiol. 8. https://doi.org/10.3389/fmicb.2017.01934

Lee, B., Moon, T., Yoon, S., Weissman, T., 2017. DUDE-Seq: Fast, flexible, and robust denoising for targeted amplicon sequencing. PLoS ONE 12. https://doi.org/10.1371/journal.pone.0181463

Lennon, J.T., Locey, K.J., 2016. The Underestimation of Global Microbial Diversity. mBio 7, e01298-16. https://doi.org/10.1128/mBio.01298-16

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M., 2012. Comparison of next-generation sequencing systems. J. Biomed. Biotechnol. 2012, 251364. https://doi.org/10.1155/2012/251364

Liu, Z., DeSantis, T.Z., Andersen, G.L., Knight, R., 2008. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Res. 36, e120. https://doi.org/10.1093/nar/gkn491

Loch, T.P., Faisal, M., 2015. Emerging flavobacterial infections in fish: A review. J. Adv. Res. 6, 283–300. https://doi.org/10.1016/j.jare.2014.10.009

Lovell, D., Pawlowsky-Glahn, V., Egozcue, J.J., Marguerat S., Bähler J., 2015. Proportionality: A Valid Alternative to Correlation for Relative Data. PLoS Comput. Biol. 11, e1004075. https://doi.org/10.1371/journal.pcbi.1004075

Mallick, H., Ma, S., Franzosa, E.A., Vatanen, T., Morgan, X.C., Huttenhower, C., 2017. Experimental design and quantitative analysis of microbial community multiomics. Genome Biol. 18, 228. https://doi.org/10.1186/s13059-017-1359-z

Mandal, S., Van Treuren, W., White, R.A., Eggesbø, M., Knight, R., Peddada, S.D., 2015. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb. Ecol. Health. Dis. 26, 27663. https://doi.org/10.3402/mehd.v26.27663

Mao, D., Zhou, Q., Chen, C., Quan, Z.X., 2012. Coverage evaluation of universal bacterial primers using the metagenomic datasets. BMC Microbiol. 12, 66. https://doi.org/10.1186/1471-2180-12-66

Markowitz, V.M., Chen, I.-M.A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., Huntemann, M., Anderson, I., Mavromatis, K., Ivanova, N.N., Kyrpides, N.C., 2012. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 40, D115–D122. https://doi.org/10.1093/nar/gkr1044

Martín-Fernández, J.A., Hron, K., Templ, M. Filzmoser, P., Palarea-Albaladejo, J., 2015. Bayesian-multiplicative treatment of count zeros in compositional data sets. Statistical Modelling: An International Journal. 15, 134-158

Mataragas, M., Alessandria, V., Ferrocino, I., Rantsiou, K., Cocolin, L., 2018. A bioinformatics pipeline integrating predictive metagenomics profiling for the analysis of 16S

rDNA/rRNA sequencing data originated from foods. Food Microbiol. 76, 279–286. https://doi.org/10.1016/j.fm.2018.05.009
May, A., Abeln, S., Crielaard, W., Heringa, J., Brandt, B.W., 2014. Unraveling the outcome of 16S rDNA-based taxonomy analysis through mock data and simulations. Bioinformatics 30, 1530–1538. https://doi.org/10.1093/bioinformatics/btu085
McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., Hugenholtz, P., 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 6, 610–618. https://doi.org/10.1038/ismej.2011.139
McMurdie, P.J., Holmes, S., 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLOS ONE. 8, e61217. https://doi.org/10.1371/journal.pone.0061217
McMurdie, P.J., Holmes, S., 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLOS Comput. Biol. 10, e1003531. https://doi.org/10.1371/journal.pcbi.1003531
Metzker, M.L., 2010. Sequencing technologies - the next generation. Nat. Rev. Genet. 11, 31–46. https://doi.org/10.1038/nrg2626
Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., Pettersson, S., 2012. Host-gut microbiota metabolic interactions. Science 336, 1262–1267. https://doi.org/10.1126/science.1223813
Norman, J.M., Handley, S.A., Baldridge, M.T., Droit, L., Liu, C.Y., Keller, B.C., Kambal, A., Monaco, C.L., Zhao, G., Fleshner, P., Stappenbeck, T.S., McGovern, D.P.B., Keshavarzian, A., Mutlu, E.A., Sauk, J., Gevers, D., Xavier, R.J., Wang, D., Parkes, M., Virgin, H.W., 2015. Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell 160, 447–460. https://doi.org/10.1016/j.cell.2015.01.002
Noyce, G.L., Winsborough, C., Fulthorpe, R., Basiliko, N., 2016. The microbiomes and metagenomes of forest biochars. Sci. Rep. 6, 26425. https://doi.org/10.1038/srep26425
Nübel, U., Garcia-Pichel, F., Muyzer, G., 1997. PCR primers to amplify 16S rRNA genes from cyanobacteria. Appl. Environ. Microbiol. 63, 3327–3332.
Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 27, 824–834. https://doi.org/10.1101/gr.213959.116
Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., Stevens, R., 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42, D206–D214. https://doi.org/10.1093/nar/gkt1226
Pace, N.R., 2009. Mapping the Tree of Life: Progress and Prospects. Microbiol. Mol. Biol. Rev. MMBR 73, 565–576. https://doi.org/10.1128/MMBR.00033-09
Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes. Genome Res. 25, 1043–1055. https://doi.org/10.1101/gr.186072.114
Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P., Woodcroft, B.J., Evans, P.N., Hugenholtz, P., Tyson, G.W., 2017. Recovery of Nearly 8,000 Metagenome-Assembled Genomes Substantially Expands the Tree of Life. Nature Microbiol. 2, 1533-1542. https://doi.org/10.1038/s41564-017-0012-7
Paulson, J.N., Pop, M., Corrada Bravo, H., 2011. Metastats: An Improved Statistical Method for Analysis of Metagenomic Data. Genome Biol. 12, P17. https://doi.org/ 10.1186/gb-2011-12-s1-p17

Pedron, R., Esposito, A., Bianconi, I., Pasolli, E., Tett, A., Asnicar, F., Cristofolini, M., Segata, N., Jousson, O., 2019. Genomic and metagenomic insights into the microbial community of a thermal spring. Microbiome 7, 8. https://doi.org/10.1186/s40168-019-0625-6

Peng, X., Li, G., Liu, Z., 2016. Zero-Inflated Beta Regression for Differential Abundance Analysis with Metagenomics Data. J. Comput. Biol. 23, 102–110. https://doi.org/10.1089/cmb.2015.0157

Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J.A., Bonazzi, V., McEwen, J.E., et al., 2009. The NIH Human Microbiome Project. Genome Res. 19, 2317–2323. https://doi.org/10.1101/gr.096651.109

Phandanouvong-Lozano, V., Sun, W., Sanders, J.M., Hay, A.G., 2018. Biochar does not attenuate triclosan's impact on soil bacterial communities. Chemosphere. 213, 215-225. https://doi.org/10.1016/j.chemosphere.2018.08.132

Prakash, T., Taylor, T.D., 2012. Functional assignment of metagenomic data: challenges and applications. Brief. Bioinform. 13, 711–727. https://doi.org/10.1093/bib/bbs033

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/nar/gks1219

Randle-Boggis, R.J., Helgason, T., Sapp, M., Ashton, P.D., 2016. Evaluating techniques for metagenome annotation using simulated sequence data. FEMS Microbiol. Ecol. 92. https://doi.org/10.1093/femsec/fiw095

Ricotta, C., Podani, J., 2017. On some properties of the Bray-Curtis dissimilarity and their ecological meaning. Ecological Complexity. 31, 201-205. http://dx.doi.org/10.1016/j.ecocom.2017.07.003

Rivera-Pinto, J., Egozcue, J.J., Pawlowsky-Glahn, V., Paredes, R., Noguera-Julian, M., Calleb, M.L. 2018. Balances: a New Perspective for Microbiome Analysis. mSystems 3, e00053-18. https://doi.org/10.1128/mSystems.00053-18

Roh, H., Subramanya, N., Zhao, F., Yu, C.-P., Sandt, J., Chu, K.-H., 2009. Biodegradation potential of wastewater micropollutants by ammonia-oxidizing bacteria. Chemosphere 77, 1084–1089. https://doi.org/10.1016/j.chemosphere.2009.08.049

Rosselli, R., Romoli, O., Vitulo, N., Vezzi, A., Campanaro, S., de Pascale, F., Schiavon, R., Tiarca, M., Poletto, F., Concheri, G., Valle, G., Squartini, A., 2016. Direct 16S rRNA-seq from bacterial communities: a PCR-independent approach to simultaneously assess microbial diversity and functional activity potential of each taxon. Sci. Rep. 6, 32165. https://doi.org/10.1038/srep32165

Rusin, P.A., Rose, J.B., Haas, C.N., Gerba, C.P., 1997. Risk Assessment of Opportunistic Bacterial Pathogens in Drinking Water, in: Ware, G.W. (Ed.), Reviews of Environmental Contamination and Toxicology: Continuation of Residue Reviews, Reviews of Environmental Contamination and Toxicology. Springer New York, New York, NY, pp. 57–83. https://doi.org/10.1007/978-1-4612-1964-4_2

Sanchez-Garcia, L., Martín, L., Mangues, R., Ferrer-Miralles, N., Vázquez, E., Villaverde, A., 2016. Recombinant pharmaceuticals from microbial cells: a 2015 update. Microb. Cell Factories 15. https://doi.org/10.1186/s12934-016-0437-3

Schloss, P.D., Gevers, D., Westcott, S.L., 2011. Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. PLOS ONE 6, e27310. https://doi.org/10.1371/journal.pone.0027310

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., et al., 2009. Introducing mothur: Open-Source, Platform-

Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Appl. Environ. Microbiol. 75, 7537–7541. https://doi.org/10.1128/AEM.01541-09
Schulz, F., Eloe-Fadrosh, E.A., Bowers, R.M., Jarett, J., Nielsen, T., Ivanova, N.N., Kyrpides, N.C., Woyke, T., 2017. Towards a balanced view of the bacterial tree of life. Microbiome 5, 140. https://doi.org/10.1186/s40168-017-0360-9
Sharpton, T.J., 2014. An introduction to the analysis of shotgun metagenomic data. Front. Plant Sci. 5. https://doi.org/10.3389/fpls.2014.00209
Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., Banfield, J.F., 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat. Microbiol. 3, 836. https://doi.org/10.1038/s41564-018-0171-1
Stern, S., Powers, T., Changchien, L.M., Noller, H.F., 1989. RNA-protein interactions in 30S ribosomal subunits: folding and function of 16S rRNA. Science 244, 783–790. https://doi.org/10.1126/science.2658053
Sun, S., Hood, M., Scott, L., Peng, Q., Mukherjee, S., Tung, J., Zhou, X., 2017. Differential expression analysis for RNAseq using Poisson mixed models. Nucleic Acids Res. 45, e106. https://doi.org/10.1093/nar/gkx204
Surana, N.K., Kasper, D.L., 2017. Moving beyond microbiome-wide associations to causal microbe identification. Nature 552, 244–247. https://doi.org/10.1038/nature25019
Teeling, H., Glöckner, F.O., 2012. Current opportunities and challenges in microbial metagenome analysis—a bioinformatic perspective. Brief. Bioinform. 13, 728–742. https://doi.org/10.1093/bib/bbs039
Tipton, L., Cuenco, K.T., Huang, L., Greenblatt, R.M., Kleerup E., Sciurpa, F., Duncan, S.R., Donahoe, M.P., 2018. Measuring associations between the microbiota and repeated measures of continuous clinical variables using a lasso-penalized generalized linear mised model. BioData Min. 11, 12. https://doi.org/10.1186/s13040-018-0173-9
Tsilimigras, M.C., Fodor A.A., 2016. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. Ann. Epidemiol. 26, 330-335. https://doi.org/10.1016/j.annepidem.2016.03.002
Tully, B.J., Graham, E.D., Heidelberg, J.F., 2018. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Scient. Data. 5, 170103. https://doi.org/10.1038/sdata.2017.203
Turnbaugh, P.J., Gordon, J.I., 2009. The core gut microbiome, energy balance and obesity. J. Physiol. 587, 4153–4158. https://doi.org/10.1113/jphysiol.2009.174136
Velsko, I.M., Frantz, L.A.F., Herbig, A., Larson, G., Warinner, C., 2018. Selection of Appropriate Metagenome Taxonomic Classifiers for Ancient Microbiome Research. mSystems 3. https://doi.org/10.1128/mSystems.00080-18
Větrovský, T., Baldrian, P., 2013. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PloS One 8, e57923. https://doi.org/10.1371/journal.pone.0057923
Vollmers, J., Wiegand, S., Kaster, A.-K., 2017. Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size Matters! PLOS ONE 12, e0169662. https://doi.org/10.1371/journal.pone.0169662
Wang, Ziye, Wang, Zhengyang, Lu, Y.Y., Sun, F., Zhu, S., 2019. SolidBin: Improving Metagenome Binning with Semi-supervised Normalized Cut. Bioinforma. Oxf. Engl. https://doi.org/10.1093/bioinformatics/btz253

White, J.R., Nagarajan, N., Pop, M., 2009. Statistical Methods for Detecting Differentially Abundant Features in Clinical Metagenomic Samples. PLOS Comput. Biol. 5, e1000352. https://doi.org/10.1371/journal.pcbi.1000352

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J.R., et al., 2017. Normalization and Microbial Differential Abundance Strategies Depend Upon Data Characteristics. Microbiome. 5, 27. https://doi/org/ 10.1186/s40168-017-0237-y

Westcott, S.L., Schloss, P.D., 2015. De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. PeerJ 3, e1487. https://doi.org/10.7717/peerj.1487

Wu, Y.-W., Tang, Y.-H., Tringe, S.G., Simmons, B.A., Singer, S.W., 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2, 26. https://doi.org/10.1186/2049-2618-2-26

Xue, Z., Kable, M.E., Marco, M.L., 2018. Impact of DNA Sequencing and Analysis Methods on 16S rRNA Gene Bacterial Community Analysis of Dairy Products. mSphere 3, e00410-18. https://doi.org/10.1128/mSphere.00410-18

Yang, B., Wang, Y., Qian, P.-Y., 2016. Sensitivity and Correlation of Hypervariable Regions in 16S rRNA Genes in Phylogenetic Analysis. BMC Bioinformatics 17, 135. https://doi.org/10.1186/s12859-016-0992-y

CHAPTER 2

BIOCHAR DOES NOT ATTENUATE TRICLOSAN'S IMPACT ON SOIL BACTERIAL COMMUNITIES

Vienvilay Phandanouvong-Lozano[a], Wen Sun[a,1], Jennie M. Sanders[a,2], Anthony G. Hay[a]

[a]Department of Microbiology, Cornell University, Ithaca, NY, 14853, USA

[1] Present address: Guangdong Xianmei Seeds Co., LTD, Guangdong, 510640, China

[2] Present address: Western Governors University, Salt Lake City, UT, 84107, USA

ABSTRACT

Triclosan, a broad-spectrum antimicrobial, has been widely used in pharmaceutical and personal care products. It undergoes limited degradation during wastewater treatment and is present in biosolids, most of which are land applied in the United States. This study assessed the impact of triclosan (0-100 mg Kg^{-1}) with and without biochar on soil bacterial communities. Very little ^{14}C-triclosan was mineralized to $^{14}CO_2$ (<7%) over the course of the study (42 days). While biochar (1%) significantly lowered mineralization of triclosan, analysis of 16S rRNA gene sequences revealed that biochar impacted very few OTUs and did not alter the overall structure of the community. Triclosan, on the other hand, significantly affected bacterial diversity and community structure (alpha diversity, ANOVA, $p<0.001$; beta diversity, AMOVA, $p<0.01$). Dirichlet multinomial mixtures (DMM) modeling and complete linkage clustering (CLC) revealed a dose-dependent impact of triclosan. Non-Parametric Metastats (NPM) analysis showed that 150 of 734 OTUs from seven main phyla were significantly impacted by triclosan (adjusted $p<0.05$). Genera harboring opportunistic pathogens such as *Flavobacterium* were

enriched in the presence of triclosan, as was *Stenotrophomonas*. The latter has previously been implicated in triclosan degradation via stable isotope probing. Surprisingly, *Sphingomonads*, which include well-characterized triclosan degraders were negatively impacted by even low doses of triclosan. Analyses of published genomes showed that triclosan resistance determinants were rare in *Sphingomonads* which may explain why they were negatively impacted by triclosan in our soil.

Keywords: biodegradation, xenobiotic, biosolid, pollutant, metagenome

2.1. INTRODUCTION

Triclosan is a broad-spectrum antimicrobial that interrupts fatty acid biosynthesis. It has been widely used in industrial, medical, and personal care products (Bhargava and Leonard, 1996; Jones et al., 2006; Singer et al., 2002). Some studies, however, have shown that triclosan does not enhance the antibacterial activity of products like hand soaps, dish soaps, or plastics (Aiello et al., 2007; Faoagali et al., 1995; Junker, 2004). Since its introduction in the 1960s, triclosan has been extensively discharged into the environment. It is commonly found in wastewater effluents and the aquatic environments that receive them (Higgins et al., 2011; Kolpin et al., 2002; Miller et al., 2008). Triclosan and its O-methylated derivative have also been detected in fish, human plasma, urine, and breast milk (Adolfsson-Erici et al., 2002; Balmer et al., 2004; Dann and Hontela, 2011). The potential toxicity of triclosan has not been fully investigated, but once in the environment it can be transformed into more toxic metabolites like chlorodioxins and chlorophenols (Latch et al., 2005). In vitro studies have shown that triclosan interferes with the detoxification activity of enzymes in the human liver (Wang et al., 2004), and acts as an

endocrine disruptor (Helbing et al., 2011; Hinther et al., 2011; Raut and Angus, 2010; Zorrilla et al., 2009).

Given the concern over triclosan's effectiveness, its bioaccumulation, its reported contribution to antibiotic resistance, and potential negative effects on human health, the Food and Drug Administration (FDA) recently banned the use of triclosan in over-the-counter antiseptic products (Voelker, 2016). This ban, however, did not include other products such as toothpastes, fabrics, and plastic goods, which suggest that triclosan's release into wastewaters will continue.

Though triclosan is removed with biosolids during wastewater treatment, trace amounts remain in treated effluent that is discharged to the environment where it may persist, especially in oligotrophic and anaerobic environments (Bester, 2003; Heidler and Halden, 2007; Ogunyoku and Young, 2014; Pycke et al., 2014; Singer et al., 2002). Triclosan residues in treated wastewater are degraded in constructed wetlands, though triclosan has an impact on the bacterial communities in those systems (Liu et al., 2016).

Triclosan has been widely reported in biosolids at concentrations ranging from 0.09-61 mg Kg^{-1} (Andrade et al., 2015; Cha and Cupples, 2009; Xia et al., 2010). Most of the triclosan in biosolids-amended soils has been reported to remain near the surface (Edwards et al., 2009; Sabourin et al., 2009), though Xia et al. (2010) found 49-64% of extractable triclosan between 30-120 cm depth in soils that had annually received biosolids for 33 years. Biodegradation has been reported to be the main mechanism of triclosan removal from soils, though rates may vary (Dhillon et al., 2015; Ying et al., 2007). A year after a single "sewage sludge" application, Butler et al. (2012b) recovered less than 20% of the initial triclosan. Similar results have been reported in other studies, with up 94% of triclosan dissipating from the surface of biosolid-amended soil

over two years (Cha and Cupples, 2009). Such dissipation has also been observed in soils receiving multiple applications (Lozano et al., 2010).

The impact of triclosan on soil biology is poorly understood. It has been reported to decrease microbial biomass and affect microbial respiration (Butler et al., 2011; McNamara et al., 2014; Zaayman et al., 2017). Even a low concentration of triclosan (4 mg Kg^{-1}) has been found to negatively affect microbial populations (Svenningsen et al., 2011), though subsequent re-dosing might lead to microbial acclimation and growth on triclosan as a carbon source (Butler et al., 2012a). Still, few studies have evaluated the effect of triclosan on bacterial community composition and structure (Guo et al., 2016; Harrow et al., 2011).

The goal of this study was to assess the impact of triclosan on bacterial communities in an agricultural soil with and without biochar. Biochar has been used as a soil amendment to modify xenobiotic bioavailability and enhance soil biodiversity (Anderson et al., 2011; Jenkins et al., 2017; Maurathan et al., 2015; Watzinger et al., 2014; Xu et al., 2016), yet, its ability to reduce the negative impact of antimicrobials like triclosan is unknown.

Although phylum level characterization of triclosan impacts on soil microbes have previously been reported (Guo et al., 2016; Zhao et al., 2015), little is known about its impact at the genus level or about the soil-relevance of organisms known to degrade triclosan in pure culture (Kagle et al., 2015; Lee et al., 2012). Here we report on the results of mineralization assays and Illumina deep 16S sequencing in soils exposed to 0-100 mg Kg^{-1} over the space of 42 days.

2.2. MATERIALS AND METHODS

2.2.1. Experimental Design

The soil was an Arkport silty clay cloam that had not been previously exposed to triclosan. It was taken from Dilmun Hill Student Farm at Cornell University, was dried, sieved, and then dosed with sufficient triclosan in ethyl acetate to generate soils contaminated at the desired levels. The ethyl acetate was allowed to evaporate in a fume hood. Soils with and without triclosan also received 1% (g biochar/g of dry soil x 100) of a hardwood biochar that had been pyrolysed at 450°C. Soil without either triclosan or biochar was considered as the primary control. Additional controls had either no biochar or no triclosan. Experimental groups and controls were monitored over 42 days.

2.2.2. Mineralization of Triclosan

Triclosan mineralization was measured at 1, 2, 4, 7, 14, 21, 28, 35, and 42 days as previously described (Hay et al., 2001). Briefly, 25 000 dpm of ^{14}C radiolabeled triclosan was spiked into soils with and without unlabeled triclosan (Ciba Specialty Chemicals) in order to obtain triclosan concentrations of 1, 10, and 100 mg Kg^{-1}. ^{14}C radiolabeled triclosan (^{14}C-TCS) was uniformly labeled only in the dichloro ring, and had a specific activity of 5.2 Mbq/mg. To capture $^{14}CO_2$ released as result of triclosan mineralization, the soil microcosms contained an internal vial with 1M NaOH, which was replaced at each sampling. The recovered NaOH was added to 10 ml of scintillation cocktail (ScintiSafe^{TM} Econo F cocktail, Fisher Scientific) and its $^{14}CO_2$ content was measured using a Beckman Coulter LS6500 multi-purpose scintillation counter. Experimental controls without triclosan were used to calculate the background $^{14}CO_2$ content, which averaged 4-9 cpm with no obvious temporal trend. In all cases background was less than <5% of the $^{14}CO_2$ measured in the radiolabeled triclosan treatments and in most cases was approximately 1% of the labeled treatments. After subtracting the background $^{14}CO_2$ content in the radiolabeled

treatments, the cumulative recovered $^{14}CO_2$ was compared between treatments using analysis of variance (ANOVA, *aov* function) and the emmeans package 1.2.1 for R (R version 3.4.4 and RStudio 1.1.442).

2.2.3. DNA Extraction and 16S rRNA Gene Illumina Sequencing

Samples taken at 0, 7, 21, and 42 days were used for 16S rRNA gene amplification and Illumina sequencing. DNA was extracted from 0.25 g of soil using the PowerSoil DNA isolation kit (Mo Bio) following the manufacturer's instructions, and then stored at -20°C. Concentration and quality of the extracted DNA were measured using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific).

The V6 region of the 16S rRNA gene was PCR amplified using the barcoded primers: forward primer 902F 5'- ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT-barcode-ACT YAA AKG AAT TGA CGG G -3' and reverse primer 1078R 5'- CGG TCT CGG CAT TCC TGC TGA ACC GCT CTT CCG ATC T-barcode-ACR ACA CGA GCT GAC GAC -3'. PCR was performed using a Peltier Thermal Cycler PTC-200 (MJ Research) with an initial denaturation at 95°C for 2 min; followed by 29 cycles of denaturation at 95°C for 20 s, annealing at 50°C for 15 s, and extension at 72°C for 5 min. PCR products were purified using PureLink Quick Gel Extraction Kit (Invitrogen). Then, the forward adapter 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GA- 3' and the reverse adapter 5'-CAA GCA GAA GAC GGC ATA CGA GAT CGG TCT CGG CAT TCC TGC TGA AC -3' were added via PCR with an initial denaturation at 95°C for 2 min; followed by 9 cycles of denaturation at 95°C for 20 s, annealing at 60°C for 15 s, and extension at 72°C for 5 min. PCR products were pooled in equimolar concentrations to generate the amplicon library. The library was spiked with

10% PhiX (Illumina) and sequenced on an Illumina MiSeq instrument, implementing V2 chemistry, with paired-end 250 nt reads according to the manufacturer's protocol at Cornell's Institute of Biotechnology.

2.2.4. Sequence Processing

The 16S rRNA gene sequences were processed using Mothur v1.39.5 following the Mother SOP (Mothur, 2013, https://www.mothur.org/wiki/MiSeq_SOP) and the recommendations of Kozich et al. (2013). Briefly, forward and reverse sequence reads were aligned and joined into contigs based on quality scores, and trimmed to 275 bp. Then, quality filtered sequences were aligned to the Silva 16S rRNA gene reference database (Glöckner et al., 2017; Quast et al., 2013; Yilmaz et al., 2014). Aligned sequences were preclustered by abundance having a maximum 2 bp difference over 250bp. By using the abundant sequences as reference and the VSEARCH algorithm (Rognes et al., 2016), chimeric sequences were removed. Sequences were then classified and annotated into Operational Taxonomic Units (OTUs) using the Ribosomal Database Project, RDP, release 9 (Cole et al., 2014), and following the methodology of Wang et al. (2007) with 100 bootstrap iterations and 80% confidence cutoff.

2.2.5. Bacterial Community Analysis

After processing the sequence reads, the final dataset contained 2,950,386 quality-filtered sequences ranging from 7,651 to 130,357 per sample. All bacterial community analyses were carried out using Mothur v1.39.5. The plots were generated with R version 3.4.4 and RStudio 1.1.442, and the gplot package 3.0.1 for R was used for the heatmaps.

Prior to further analysis, the samples were rarefied with 1,000 randomizations to the size of the smallest library (7,651 reads). Samples coverages were all above 99%, indicating that each library contains a sufficient number of sequences to adequately represent the diversity present in each sample. Alpha diversity was estimated with the Inverse Simpson Index, which accounts for the probability that two randomly drawn individuals from the same sample belong to different species, or in our case OTUs (Simpson, 1949). Inverse Simpson Indexes and phylum relative abundances were analyzed with the statistical model of linear mixed-effects from the lme4 package 1.1-17 for R, which corrected for the replicate effect given by the sampling times (Bates et al. 2015). Differences between groups were determined using the ANOVA function of the lme4 and emmeans packages in R.

The bacterial communities among the groups were compared by using the Yue and Clayton distance metric, Θ_{YC}, (Yue and Clayton, 2005), which measures dissimilarity between two communities based on the membership and abundance of each OTU. Distance matrices generated by Θ_{YC} coefficients were visualized using non-metric multidimensional scaling (NMDS). To test whether the spatial separation among the groups in the NMDS plots was statistically significant an Analysis of Molecular Variance (AMOVA) was performed, and the car package 2.1-5 for R was used to generate the ellipses and centroids. The ellipses predict the space in which 95% of new observations would occur.

In addition, the data clustered into distinct bacterial community types when using Dirichlet multinomial mixtures (DMM) (Holmes et al., 2012). The Laplace approximation was used to select the number of community types that best fit the data. The significance of pairwise differences in OTU abundances between the groups were determined with the Non-Parametric Metastats (NPM) analysis as incorporated in Mothur v1.39.5 (White et al., 2009). Heatmap.2 function of the gplot package 3.0.1 for R was used to generate a heatmap and dendrogram (with

complete linkage clustering (CLC)) for those OTUs identified as significant via NPM analysis. A volcano plot was used to visualize the fold change of OTUs that differed significantly between treatments.

Discreet NPM analyses were also performed on OTUs within each triclosan concentration and control in the absence (0%) and presence of biochar (1%), yielding a presence/absence ratio for significantly affected OTUs. A ratio of 1 indicated no effect of biochar, <1 that biochar negatively affected the population, and >1 that biochar had a positive effect. Relative abundances of OTUs were summed at the phylum level, then compared between biochar treatments using the Wilcoxon test in R.

2.2.6. Predictive Gene Profile

The functional gene profile of each sample was predicted using Tax4Fun package 0.3.1 for R (Aßhauer et al., 2015). The program infers function by comparing the sequenced genomes of organisms whose 16S rRNA genes match OTUs found in any given sample. We compared significant OTUs (NPM, adjusted $p<0.05$) differing by 2-fold or more. The program BLASTed 16S rRNA gene data of these OTUs against the SILVA seed 123. The 16S rRNA gene copy numbers were normalized with the copy numbers obtained from the NCBI genome annotations (Aßhauer et al., 2015). Then, functional gene profiles were generated using ultrafast protein classification (UproC) tool (Meinicke, 2015) assigning sequencing reads to KEGG Ortholog (KO) profiles from the KEGG database release 86.0.

Triclosan-resistance associated genes corresponding to: i) enoyl-acyl carrier reductases (ENRs) FabI, FabK, FabL, and FabV (Heath et al., 2000; Heath and Rock, 2000; Massengo-Tiassé and Cronan, 2008; McMurry et al., 1998a; Yazdankhah et al., 2006), ii) metagenome-

derived 7-α hydroxysteroid dehydrogenase (HSDH) ENR (Khan et al., 2016), iii) multidrug efflux pump AcrAB (McMurry et al., 1998b), and iv) aromatic degradation genes (Kagle et al., 2015; Lee et al., 2012; Lee and Chu, 2013; Meade et al., 2001; Mulla et al., 2016) were selected from the predicted functional gene profiles using the KO numbers associated by Tax4Fun.

2.2.7. Triclosan-Resistant Determinants in Bacterial Genomes

NCBI BLASTp analyses were carried out on bacterial taxa positively impacted by triclosan addition to determine the presence of putative triclosan-resistant genes in their genomes. We used the triclosan-resistant genes in the database described by Khan et al. (2018) as queries, which include the ENRs FabI, FabK, FabL, FabV, and the metagenome-derived 7-α HSDH homologue, as well as the AcrB efflux pump subunit (NCBI accession numbers: NP_415804.1, NP_357973.1, NP_388745.1, NP_233170.1, AOO54602.1, and AAC73564.1, respectively). Annotated genes with ≥ 27% predicted amino acid sequence identity with the queries were considered to be evidence of triclosan-resistance determinants in the genomes (Khan et al., 2018). When BLASTp analyses did not return any hit, tBLASTn analyses were conducted with the same % sequence identity cutoff.

2.2.8. Triclosan Sensitive ENRs in *Sphingomonas*

Predicted enoyl-acyl carrier reductases (*fabIs*) from published *Sphingomonas* genomes were aligned to the triclosan-sensitive FabI sequence of *Escherichia* coli K12 MG1655 (GenBank: AAC74370.1) using the BLAST tool in Geneious 11.0.2. The predicted FabIs from *Sphingomonas* species were characterized as either sensitive or resistant based on the presence of

nonsynonymous single point mutations in amino acids 23, 159, or 203 which have been shown to confer triclosan resistance on *E. coli's* FabI (Jiten Singh et al., 2011; Khan et al., 2016; McMurry et al., 1998a; Stewart et al., 1999).

2.3. RESULTS AND DISCUSSION

2.3.1. Mineralization of Triclosan

Triclosan concentration significantly impacted the amount of $^{14}CO_2$ released (Fig 1, ANOVA, $p=3.66x10^{-6}$). In the absence of biochar, the mineralization in 100 mg Kg^{-1} was significantly different from 1 and 10 mg Kg^{-1} (t-test, adjusted $p<0.01$), though the latter did not differ significantly from each other. Conversely, in the presence of biochar, we observed no significant difference in mineralization between 10 and 100 mg Kg^{-1}, though both were different from 1 mg Kg^{-1} (t-test, adjusted $p<0.01$). Biochar can impact the

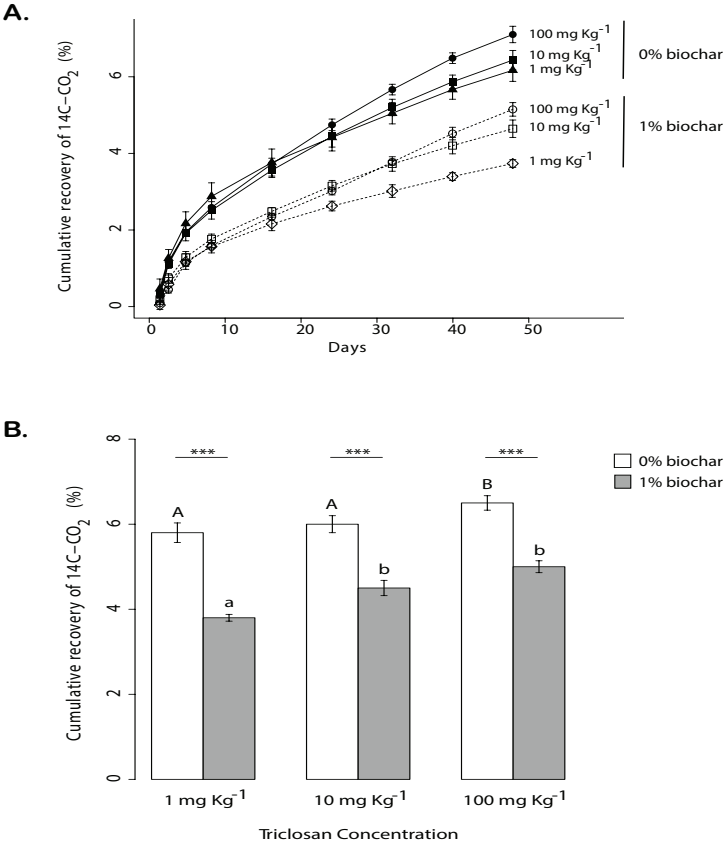


Fig. 1. Triclosan mineralization. Cumulative recovery of ^{14}C-CO_2 from ^{14}C-radiolabeled triclosan added to soil over time (**A**), and at the end of the study (**B:** 42 days). Biochar significantly reduced triclosan mineralization (*** t-test, $p<0.001$). Different letters indicate significant within-treatment differences in mineralization with (1%, grey bars) and without (0%, white bars) biochar, respectively (t-test, adjusted $p<0.01$). Error bars represent 1 S.D. of the mean, for n=3.

bioavailability of organic compounds by sorbing them; but may not be able to completely mask the effects of some compounds (Bair et al., 2016; Lehmann et al., 2011). The impact of biochar on cumulative mineralization was significant at all tested concentrations of triclosan after 42 days (Fig 1B, ANOVA, $p=1.59x10^{-10}$). In the absence of biochar, the cumulative recovery of $^{14}CO_2$ ranged from 5.8% to 6.5%, while, in presence of biochar it ranged from 3.8-4.9% (Fig 1B). Since biochar lowered the cumulative recovery of $^{14}CO_2$, we expected that it would affect the microbial community as had been previously reported in other studies in which biochar was applied to soil (Watzinger et al., 2014; Xu et al., 2016). We therefore assessed the impact of triclosan on the bacterial communities in both the presence and absence of biochar.

2.3.2. Relative Abundance of Bacteria Phyla

The 16S rRNA gene sequence analysis revealed 734 OTUs across all soil samples. The triclosan treatments (1, 10, and 100 mg Kg^{-1}) and control (0 mg Kg^{-1}) shared more than half of the total OTUs, and had few unique OTUs (Fig S1). These unique OTUs constituted less than 0.15% of the total relative abundance. Alpha diversity and community structure differed significantly with time (ANOVA $p=2.12x10^{-9}$, AMOVA $p<0.01$). The AMOVA analysis, however, showed that the bacterial communities on days 21 and 42 were not different from each other (Fig S2a, $p=0.216$). These results suggested that the communities in microcosms had stabilized by 21 days so further analyses combined days 21 and 42 when analyzing triclosan impacts.

Independent of biochar-addition, triclosan significantly affected the relative abundances of the four most abundant phyla (Fig 2A ANOVA, $p<0.01$): *Firmicutes* (22-26%), *Proteobacteria* (20-27%), *Actinobacteria* (18-22%), and *Bacteroidetes* (3-7%). In contrast,

Bacteroidetes was the only phylum positively affected by biochar-addition in the control and in the 1 mg Kg^{-1} triclosan treatment (Fig 2A, t-test, adjusted $p<0.01$).

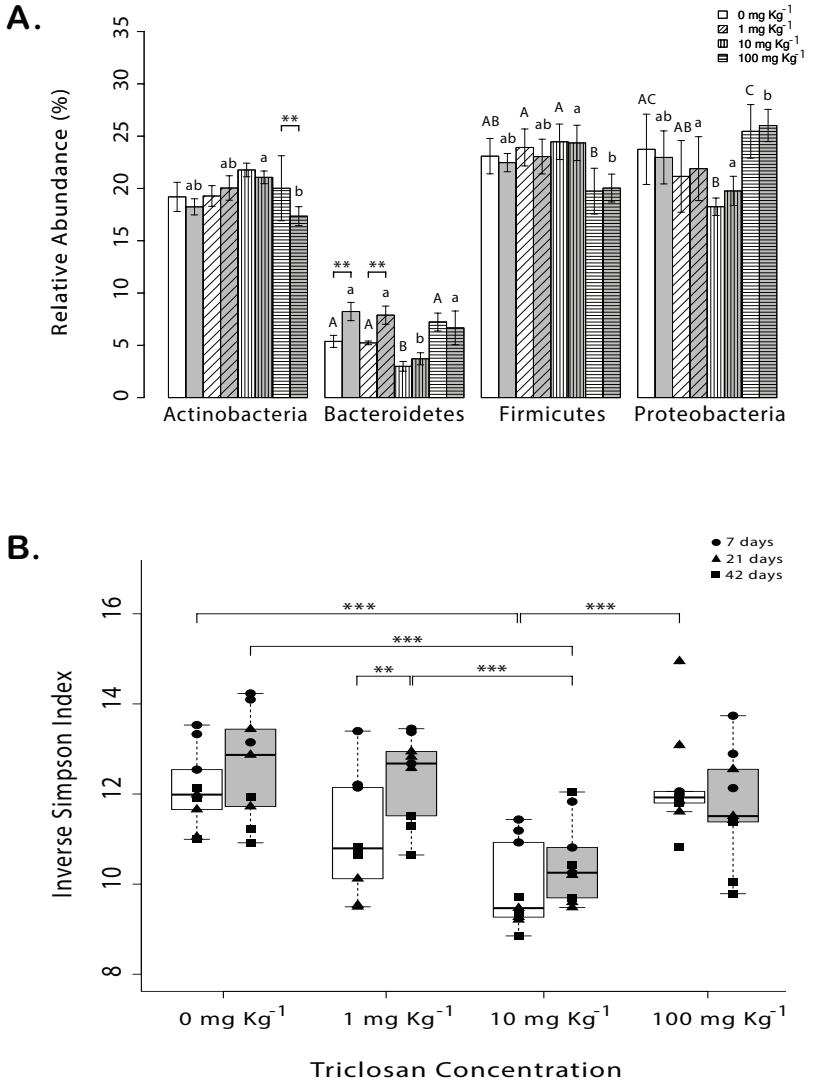


Fig. 2. A. Biochar impacts on the relative abundances of the four most abundant phyla in each treatment. Different letters indicate within-phylum differences as a function of triclosan concentration with (upper case = 1%, grey bars) and without (lower case = 0%, white bars) biochar (t-test, adjusted $p<0.01$). Error bars represent 1 S.D. of the mean, for n=3. **B.** Bacterial alpha diversity by treatment with (grey boxes) and without (white boxes) biochar measured using the Inverse Simpson Index. **: t-test, $p<0.01$, and ***: t-test, adjusted $p<0.001$.

2.3.3. Effect of Biochar on Bacterial Communities

Numerous studies on biochar-amended soils have demonstrated that biochar provides a diverse niche for microorganisms resulting in increases in microbial biomass and metabolism (Jenkins et

al., 2017; Maurathan et al., 2015; Watzinger et al., 2014; Xu et al., 2016). The bacterial response, however, is widely different depending on the soil fraction, soil type, quality of biochar and application time (Chen et al., 2015; Lehmann et al., 2011; Noyce et al., 2016).

Biochar only affected the alpha diversity in the 1 mg Kg^{-1} triclosan treatment (Fig 2B, t-test, adjusted $p<0.01$), and did not significantly impact beta diversity (Fig S2b, AMOVA, $p=0.678$). Nonetheless, NPM analyses revealed that some OTUs were significantly impacted by biochar in a triclosan-concentration dependent manner (control: 32 OTUs, 1 mg Kg^{-1}: 43 OTUs, 10 mg Kg^{-1}: 18 OTUs, and 100 mg Kg^{-1}: 32 OTUs).

When the mean abundance of biochar-impacted OTUs was summed at the phylum level, the amount of *Bacteroidetes* in the control and the 1 mg Kg^{-1} triclosan treatment more than doubled. No impact, however, was observed on *Bacteroidetes* at 10 mg Kg^{-1} or 100 mg Kg^{-1} (Table S1). Other studies have reported that *Bacteroidetes* increased more than four-fold in biochar-amended soils (Kolton et al., 2011; Zhang et al., 2017). Yet, *Bacteroidetes* were found to be sensitive to triclosan in sediments by Guo et al. (2016). Consistent with their report, our data confirms that triclosan negated the positive effects of biochar on *Bacteroidetes* and did so at a lower concentration (10 mg Kg^{-1}) than the 60 mg Kg^{-1} that was assessed by Guo et al (2016).

Conversely, *Proteobacteria*, many of which are known to be resistant to triclosan (Guo et al., 2016; Kagle et al., 2015; Mulla et al., 2016), were the only phylum positively affected by biochar at triclosan concentrations of 10 mg Kg^{-1} and 100 mg Kg^{-1}, though this phylum was not affected by biochar in the control or 1 mg Kg^{-1} triclosan (Table S1). Biochar is known to immobilize some organic compounds and reduce their rate of degradation, although in some cases it has been reported to increase overall degradation (Atkinson et al., 2010; Smebye et al., 2016). Among the *Proteobacterial* OTUs positively impacted by biochar-addition (NMP

analysis, $p<0.05$) were two taxa that include known xenobiotic degraders *Propionivibrio* and *Variovorax*, though no reports could be found detailing their ability to degrade triclosan.

2.3.4. Bacterial Diversity and Structure Affected by Triclosan

Regardless of the presence or absence of biochar, alpha diversity differed significantly by triclosan concentration (ANOVA $p=7.9x10^{-7}$). Samples exposed to 10 mg Kg^{-1} triclosan showed the greatest decrease in alpha diversity (Fig. 2B). Beta diversity, as measured by θYC dissimilarity distances and visualized via NMDS, showed statistically significant clustering of bacterial communities by triclosan concentration (Fig 3). The AMOVA confirmed these results, finding that samples exposed to triclosan (1, 10, and 100 mg Kg^{-1}) were significantly different from one another ($p<0.001$), and from the control ($p<0.01$). The Dirichlet Multinomial Mixtures (DMM) model and the dendrogram generated via complete linkage clustering (CLC) were in agreement that community composition was driven by triclosan

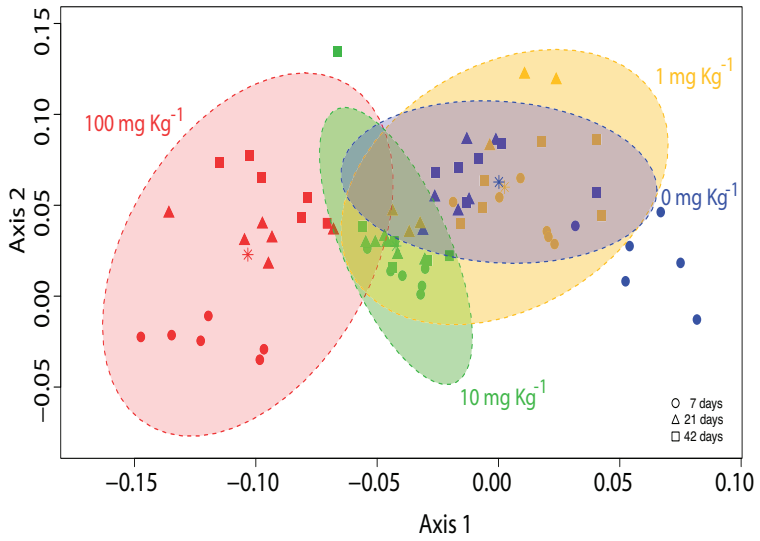


Fig 3. NMDS analysis of the bacterial community structures using θYC dissimilarity distances. Distances between symbols on the NMDS plot reflect relative dissimilarities in the community structures. The 2 axes represent 95% of the variance. The lowest stress is 0.104 with an R^2 value of 0.97. Ellipses represent the 95% confidence intervals around the centroid for each cluster. The ellipse centroids are indicated by ✳ . The spatial distances of the bacteria communities exposed to triclosan concentrations (1, 10, and 100 mg Kg^{-1}) differed significantly from one another ($p<0.001$) and from the control (0 mg Kg^{-1}) ($p<0.01$) as determined by AMOVA.

concentration, though the DMM model separated samples into three community types while CLC clustering further resolved the 10 mg Kg^{-1} and 100 mg Kg^{-1} treatments from one another (Fig. 4). A summary of the results from the analyses mentioned above is provided in Supplementary Table S2 and clearly demonstrates the statistical significance of triclosan addition on bacterial community structure, regardless of the analysis employed.

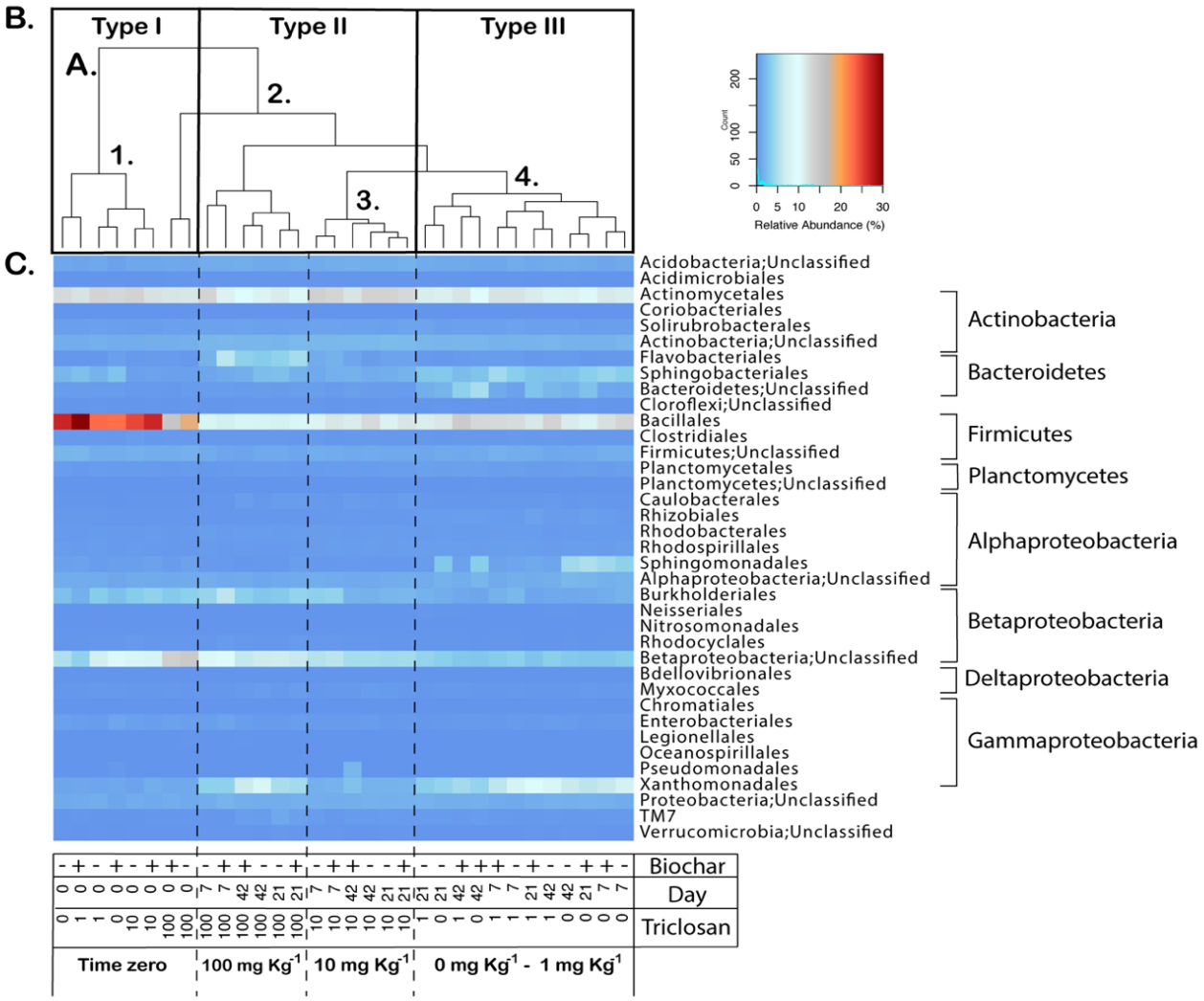


Fig 4. Three different multivariate statistical analyses support the same conclusion: triclosan significantly impacted the soil communities in a dose dependent fashion. **A.** Dendrogram with complete linkage clustering (CLC) shows four significant nodes as indicated by the dotted lines in the heat map and named at the bottom in bold. **B.** The boxes around the branches indicate the community types identified using DMM modeling: Type I: all treatments at time zero, Type II: 10 mg Kg^{-1} and 100 mg Kg^{-1}, and Type III: control and 1 mg Kg^{-1}. **C.** Heat map showing the differential relative abundance of 150 OTUs significantly impacted by triclosan based on the NPM analysis (adjusted $p<0.05$), but collapsed to the order level for ease of visualization. The treatment conditions are indicated in the imbedded table: + and – indicates the presence or absence of biochar, respectively. Sampling times are indicated by Day: 0, 7, 21, and 42. Triclosan concentrations are indicated 0, 1, 10 and 100 mg Kg^{-1}.

Several studies have demonstrated triclosan's dose-dependent effect on microbial communities; reporting decrease in microbial diversity, biomass and respiration (Butler et al., 2011; Harrow et al., 2011; Zaayman et al., 2017). Even soils irrigated with low concentrations of triclosan (2 mg Kg^{-1}) showed shifts in the relative abundance of numerous phyla and overall structure of the communities (Harrow et al., 2011). At higher doses, triclosan affected the bacterial communities very rapidly: Butler et al. (2011) found a reduction in the basal respiration of clay soils after only 16 hours of exposure to 10 mg Kg^{-1} and 100 mg Kg^{-1} triclosan. This rapid effect is consistent with the CLC dendrogram results which grouped all of the 100 mg Kg^{-1} samples together including that of day zero (Fig 4A). These results suggest that triclosan was able to impact the bacterial communities in our sample during the few hours it took to process them on the first day of our study.

We were particularly interested in understanding what was driving the decrease in alpha diversity for the 10 mg Kg^{-1} treatment even though alpha diversity did not appear to be impacted in the 100 mg Kg^{-1} treatment (Fig 2B). As shown qualitatively in the Venn diagram, richness did not differ between the triclosan treatments groups (Fig. S1). These findings suggest that triclosan disproportionately impacted a few shared OTUs. The NPM analysis revealed that 150 of the total 734 OTUs were impacted by triclosan (Fig 5A, $p<0.05$). These 150 OTUs belonged to the phyla *Acidobacteria, Actinobacteria, Bacteroidetes, Cloroflexi, Firmicutes, Planctomycetes, Proteobacteria,* and *Verrucomicrobia.* The heatmap collapses their relative abundances down to the order level for ease of visualization (Fig 4C).

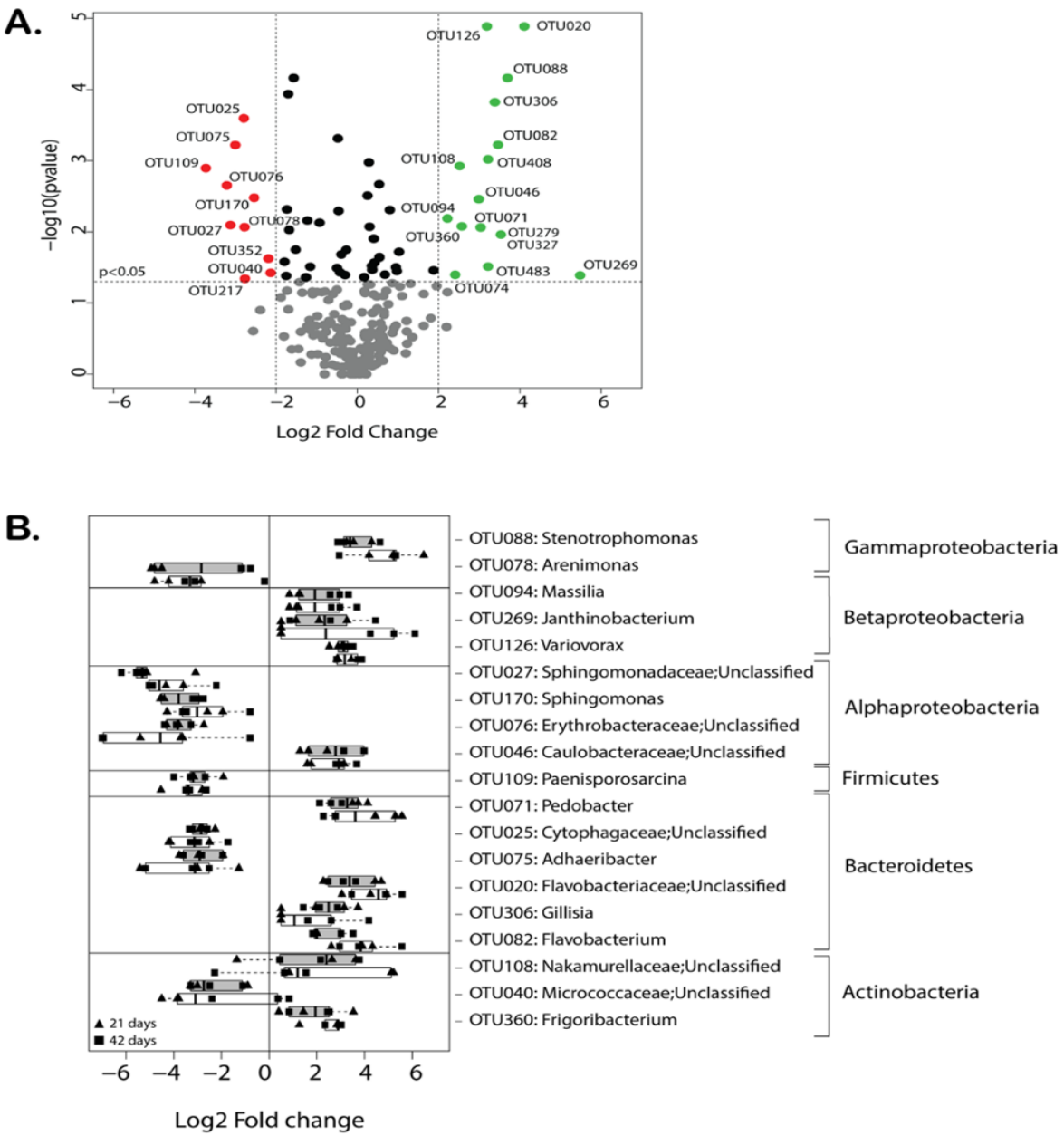


Fig 5. Fold change of differentially abundant OTUs in the 100 mg Kg^{-1} triclosan treatment relative to the control for all times points **A.** The volcano plot shows the fold change (x axis) in OTUs between 100 mg Kg-1 and the control for all time points, as well as the statistical significance of that change (y axis) as determined by NPM analysis. The relative abundance of OTUs above the dotted line were significantly impacted by triclosan (NPM, adjusted $p<0.05$). Green OTUs increased more than 2-fold in the presence of 100 mg Kg-1 triclosan, while red OTUs decreased more than 2-fold. **B.** Change in abundance of OTUs significantly impacted by triclosan from days 21 and 42 in the presence (grey boxes) and absence (white boxes) of biochar.

OTUs that significantly changed by two-fold or more in response to triclosan on days 21 and 42 were averaged and selected for further analysis since the NMDS plot showed no

community differences on these days (Fig S2a, AMOVA, p=0.216). Of the nine significant OTUs in the 10 mg Kg^{-1} treatment, seven decreased ≥ 2-fold compared to the control (Fig S3a). Only two OTUs, *Caulobacteraceae; unclassified,* and genus *Massilia,* showed ≥ 2-fold increase in that treatment (Fig S3b). The relative abundance of the latter two OTUs was also higher in the 100 mg Kg^{-1} treatment as where the abundances of three other *Proteobacteria: Variovorax, Janthinobacterium,* and *Stenotrophomonas* (Fig 5B). Other taxa positively affected by triclosan belonged to the orders *Actinomycetales, Flavobacteriales,* and *Sphingobacteriales* (Fig 4C and 5). For instance, the genus *Flavobacterium* increased 5.6-fold at 100 mg Kg^{-1} compared to the control (Fig 5B). Additionally, OTUs from the following five genera were detected in the 100 mg Kg^{-1} treatment, but not in the control: *Luteibacter, Nocardiopsis, Chitinophaga, Actinomycetospora,* and *Achromobacter.*

Not surprisingly, studies on triclosan-irrigated soils have shown that cultured bacteria known to be sensitive are significantly impacted, causing an overall diversity decrease, though some tolerant species increased in abundance (Harrow et al., 2011; Svenningsen et al., 2011). Triclosan targets some enoyl-acyl carrier protein reductases (ENRs) during fatty acid elongation, but not all ENRs are sensitive to triclosan (Heath and Rock, 2000; Zhu et al., 2010). In fact, *fabL* from *Bacillus subtilis* has been used as a selective marker in a broad host range cloning vector that confers triclosan resistance (Kagle and Hay, 2002). Furthermore, recent evidence suggests that members of the genus *Massilia,* which we found to increase in the presence of 10 and 100 mg Kg^{-1} triclosan, appear to encode a novel family of triclosan-resistant ENRs (Table S3) (Khan et al., 2016).

2.3.5. Predicted Impacts of Triclosan on Functional Diversity

The predicted abundance of presumptive TCS-resistance determinants (ENRs, efflux pumps, and aromatic degradation genes) increased in the presence of triclosan at the end of our study (42 days), as estimated using Tax4Fun (Fig 6) (Aßhauer et al., 2015). Of the well characterized ENRs (FabI, FabK, FabL, and FabV) and metagenome-derived 7-α-HSDH, *fabL*, which is known for its triclosan resistance, was the most abundant at 10 and 100 mg Kg^{-1} triclosan (Heath et al., 2000; Kagle and Hay, 2002; Khan et al., 2016; Massengo-Tiassé and Cronan, 2008; McMurry et al., 1998a). The predicted prevalence of *fabK* decreased in a dose response manner and was impacted by even 1 mg Kg^{-1} triclosan (Fig 6). Only the taxa *Flavobacteriaceae* and *Chitinophaga* were found to encode *fabK* by BLASTp analyses (Table S3). Although the *fabK* from *Enterococcus faecalis* and the metagenome-derived *fabK* ENR homologue have been reported to confer mild resistant to triclosan (Heath and Rock, 2000; Khan et al., 2016), *E. coli* expressing *fabL* can tolerate triclosan in agar at levels beyond its limit of solubility (Kagle and

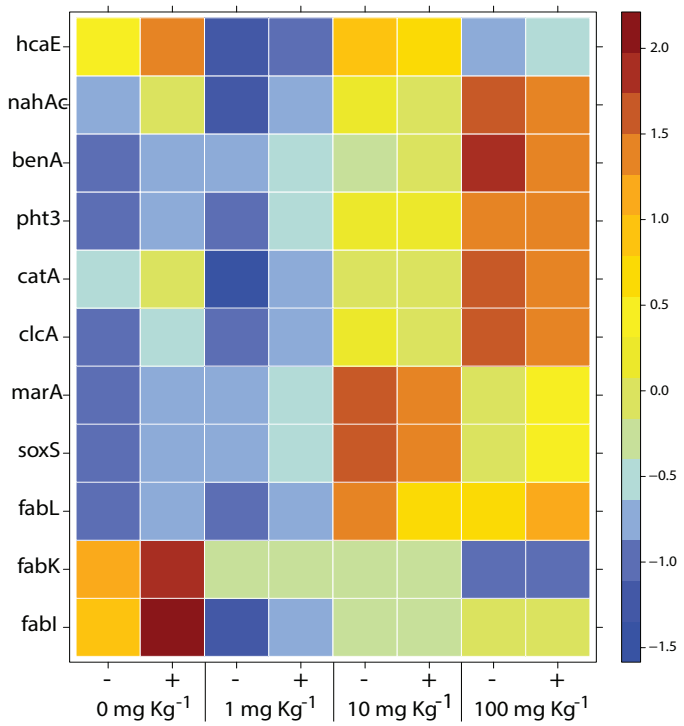


Fig 6. Scaled heatmap of the predicted abundance of triclosan-relevant genes (ENRs, efflux pumps, and aromatic degradation genes) in all triclosan treatments at 42 days in the presence (+) and absence (-) of biochar, as inferred by Tax4Fun analysis. *fabI, fabK, fabL*: enoyl-acyl-carrier reductase (KEGG#: K00208, K02371, and K10780, respectively), *marA* and *soxS*: AraC family transcriptional regulator (K13632 and K13631, respectively), *clcA*: chlorocatechol 1,2-dioxygenase (K15253), *catA*: catechol 1,2-dioxygenase (K03381), *pht3*: phthalate 4,5-dioxygenase (K07519), *benA*: benzoate 1,2-dioxygenase subunit alpha (K05549), *nahAc*: naphthalene 1,2-dioxygenase subunit alpha (K14579), *hcaE*: 3-phenylpropionate/trans-cinnamate dioxygenase subunit alpha (K05708)

Hay, 2002). The Tax4fun results also showed that *fabIs*, most of which are known to be triclosan sensitive, were less abundant in all triclosan treatments than the control (Fig 6). The ENR *fabV* was not predicted to be present by Tax4fun, which may have resulted from the lack of gene annotation and assigned KEGG Ortholog profiles in some of the pre-computed genomes used by Tax4Fun like *Flavobacterium*. Yet, BLAST analyses demonstrated the presence of a FabV-like ENR in the genomes of some taxa that increased in those soils exposed to triclosan (Table S3). FabV is known to confer triclosan-resistance in opportunistic human pathogens like *Pseudomonas aeruginosa* and *Vibrio cholera* (Huang et al., 2016; Massengo-Tiassé and Cronan, 2008; Zhu et al., 2010). Several studies on triclosan resistance have reported that multidrug efflux pumps such as AcrAB-TolC and MexAB-OprM can confer resistance on pathogens like *Salmonella enterica, E. coli* and *P. aeruginosa* (Chuanchuen et al., 2001, 2003; McMurry et al., 1998b; Webber et al., 2008). Interestingly, genome analyses showed that AcrAB-TolC efflux pumps are also encoded by most of the bacteria whose relative abundances increased in the presence of triclosan, many of which have additional types of triclosan-resistant ENRs (Table S3, Fig 5B). Tax4fun predicted that AcrAB-TolC efflux pump positive regulators *marA* and *soxS* were likely to be more abundant in the 10 mg Kg^{-1} and 100 mg Kg^{-1} triclosan treatment than in the control or 1 mg Kg^{-1} treatment (Fig 6). Overexpression of these genes in *E. coli* has been associated with modest triclosan resistance (McMurry et al., 1998b).

2.3.6 Higher Triclosan Concentrations Resulted in More Biodegradation

One explanation for alpha diversity being higher at 100 mg Kg^{-1} than at 10 mg Kg^{-1} is the possibility that xenobiotic degraders were favored at the highest triclosan concentration. In support of that hypothesis, we found that mineralization followed a dose response, being highest

at higher concentrations of triclosan (Fig 1A). Additionally, several genes encoding the degradation of aromatic compounds such as catechol- and benzoate dioxygenases were predicted to be more abundant in the 100 mg Kg^{-1} treatment by Tax4fun (Fig 6), although the phenoxybenzoate dioxygenase *hcaE*, which has been closely related with the putative triclosan-degrading dioxygenase *tcsA* (Kagle et al., 2015), was predicted to be less abundant.

Stenotrophomonas and *Variovorax*, which increased in the presence of triclosan (Fig 5B) include known xenobiotic degraders (Dejonghe et al., 2003; Lee et al., 2002; Sørensen et al., 2008), and members of both genera appear to encode the TCS-resistance determinants: FabV and AcrAB-TolC efflux pump (Table S3). Interestingly, *Stenotrophomonas* has recently been shown via stable isotope probing to use triclosan as carbon source in sewage sludge enrichments (Lee et al., 2014; Lolas et al., 2012). *Stenotrophomonas maltophilia* is known to overexpress a multridrug-efflux pump after triclosan exposure, conferring not only resistance to triclosan but to other antibiotics as well (Sanchez et al., 2005). This ubiquitous genus is also home to opportunistic human pathogens (Al-Anazi and Al-Jasser, 2014; Brooke, 2012). This is the first report, however, of *Stenotrophomonas* abundance increasing in triclosan amended soils.

Contrary to expectations based on our earlier work with triclosan degrading bacteria, none of the triclosan-enriched OTUs were related to *Sphingomonadales* (Hay et al., 2001; Kagle et al., 2015). In fact, the genera *Sphingomonas* and *Sphingopyxis* decreased four to five-fold when exposed to all tested triclosan concentrations (Fig 5B, S3b, S4). This was very surprising given that members of these genera have been reported to degrade triclosan (Kagle et al., 2015; D. G. Lee et al., 2012; Mulla et al., 2016) and *Sphingomonadales* abundance has been shown to increase in sediments receiving 60 mg Kg^{-1} of triclosan (Guo et al., 2016). In that work, however, triclosan disappearance exceeded 80%, thus high degradation correlated with increase

Sphingomonadales abundance suggesting that triclosan degrading *Sphingomonadales* may have contributed to triclosan degradation.

Not all *Sphingomonadales*, however, are resistant to triclosan or able to degrade it. Kim et al. (2011) found that the well-characterized xenobiotic degrader *Sphingomonas wittichii* RW1 could not degrade triclosan and was sensitive to it, even at very low doses (<0.01 mg L^{-1}). The triclosan sensitivity of members of these genera may be explained by both the gene dosage and the nature of the FabI encoded in the genome. The genomes of the known triclosan-degraders *Sphingomonas sp.* RD1 and *Sphingomonas sp.* YL-JM2C (Kagle et al., 2015; Mulla et al., 2016, 2015) encode three or more FabI homologs, several of which are predicted to be triclosan resistant, whereas their closest relatives encode only one FabI which is predicted to be sensitive to triclosan (Table S4).

The relatively low recovery of $^{14}CO_2$ (<6.5% of added ^{14}C-triclosan) (Fig 1) from our samples, coupled with the decrease in *Sphingomonadales* in our soil (Fig 5 and S3) is consistent with the hypothesis that triclosan resistance and degradation determinants are not universal traits among *Sphingomonadales* (Table S4). This may explain in part why triclosan degradation rates vary widely from soil to soil (Lozano et al., 2010; Ying et al., 2007; Butler et al., 2012b; Higgins et al., 2011). In addition, abiotic factors like soil pH, texture, moisture content, and organic matter have been found to affect triclosan bioavailability and degradation rate in soil (Butler et al., 2012b; Chen et al., 2008; Wu et al., 2009).

In summary, our work with 16S rRNA gene sequencing coupled with predictive functional gene profiling suggests that multiple resistant determinants are likely used by soil bacteria exposed to triclosan. Additionally, our work implies that aromatic dioxygenases with broad substrate specificity, but which have not previously been implicated in triclosan

metabolism, may contribute to its degradation in the soil. More work such as stable isotope probing, metatranscriptomics, and characterization of triclosan resistant isolates from soil will be required to test these hypotheses.

2.4. CONCLUSIONS

Even though biochar appeared to reduce triclosan bioavailability, biochar could not overcome the effect of higher concentrations of triclosan on soil bacterial communities. Triclosan addition resulted in large community shifts. Several bacteria taxa increased in presence of triclosan, including those home to opportunistic pathogens that encode triclosan resistant ENRs and multidrug efflux pumps (Khan et al., 2015; McMurry et al., 1998a). Several aromatic-degradation genes were predicted to be more abundant in our highest triclosan treatment. Xenobiotic degraders like *Variovorax* and *Stenotrophomonas* (Dejonghe et al., 2003; Lee et al., 2002) increased in our soil with triclosan addition, though the relative abundance of *Sphingomonas*, that includes well-characterized triclosan degraders, decreased even at our lowest concentration (1 mg Kg^{-1}). Our results suggest that bacterial responses to triclosan addition were concentration dependent with resistance being primary at 10 mg Kg^{-1} and with degradation, by uncharacterized triclosan degraders, being more important at 100 mg Kg^{-1}.

2.5. REFERENCES

Adolfsson-Erici, M., Pettersson, M., Parkkonen, J., Sturve, J., 2002. Triclosan, a commonly used bactericide found in human milk and in the aquatic environment in Sweden. Chemosphere 46, 1485–1489.

Aiello, A.E., Larson, E.L., Levy, S.B., 2007. Consumer antibacterial soaps: effective or just risky? Clin. Infect. Dis. 45, S137–S147.

Al-Anazi, K.A., Al-Jasser, A.M., 2014. Infections Caused by Stenotrophomonas maltophilia in Recipients of Hematopoietic Stem Cell Transplantation. Front. Oncol. 4, 232. https://doi.org/10.3389/fonc.2014.00232
Anderson, C.R., Condron, L.M., Clough, T.J., Fiers, M., Stewart, A., Hill, R.A., Sherlock, R.R., 2011. Biochar induced soil microbial community change: Implications for biogeochemical cycling of carbon, nitrogen and phosphorus. Pedobiologia 54, 309–320. https://doi.org/10.1016/j.pedobi.2011.07.005
Andrade, N.A., Lozano, N., McConnell, L.L., Torrents, A., Rice, C.P., Ramirez, M., 2015. Long-term trends of PBDEs, triclosan, and triclocarban in biosolids from a wastewater treatment plant in the Mid-Atlantic region of the US. J. Hazard. Mater. 282, 68–74. https://doi.org/10.1016/j.jhazmat.2014.09.028
Aßhauer, K.P., Wemheuer, B., Daniel, R., Meinicke, P., 2015. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics 31, 2882–2884. https://doi.org/10.1093/bioinformatics/btv287
Atkinson, C.J., Fitzgerald, J.D., Hipps, N.A., 2010. Potential mechanisms for achieving agricultural benefits from biochar application to temperate soils: a review. Plant Soil 337, 1–18. https://doi.org/10.1007/s11104-010-0464-5
Bair, D.A., Mukome, F.N.D., Popova, I.E., Ogunyoku, T.A., Jefferson, A., Wang, D., Hafner, S.C., Young, T.M., Parikh, S.J., 2016. Sorption of Pharmaceuticals, Heavy Metals, and Herbicides to Biochar in the Presence of Biosolids. J. Environ. Qual. 45, 1998–2006. https://doi.org/10.2134/jeq2016.03.0106
Balmer, M.E., Poiger, T., Droz, C., Romanin, K., Bergqvist, P.-A., Müller, M.D., Buser, H.-R., 2004. Occurrence of Methyl Triclosan, a Transformation Product of the Bactericide Triclosan, in Fish from Various Lakes in Switzerland. Environ. Sci. Technol. 38, 390–395. https://doi.org/10.1021/es030068p
Bates, D., Mächler, M., Boelker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. 67, 1-48. https://doi.org/10.18637/jss.v067.i01
Bester, K., 2003. Triclosan in a sewage treatment process—balances and monitoring data. Water Res. 37, 3891–3896. https://doi.org/10.1016/S0043-1354(03)00335-X
Bhargava, H.N., Leonard, P.A., 1996. Triclosan: applications and safety. Am. J. Infect. Control 24, 209–218.
Brooke, J.S., 2012. Stenotrophomonas maltophilia: an Emerging Global Opportunistic Pathogen. Clin. Microbiol. Rev. 25, 2–41. https://doi.org/10.1128/CMR.00019-11
Butler, E., Whelan, M.J., Ritz, K., Sakrabani, R., van Egmond, R., 2012a. The effect of triclosan on microbial community structure in three soils. Chemosphere 89, 1–9. https://doi.org/10.1016/j.chemosphere.2012.04.002
Butler, E., Whelan, M.J., Ritz, K., Sakrabani, R., van Egmond, R., 2011. Effects of triclosan on soil microbial respiration. Environ. Toxicol. Chem. 30, 360–366. https://doi.org/10.1002/etc.405
Butler, E., Whelan, M.J., Sakrabani, R., van Egmond, R., 2012b. Fate of triclosan in field soils receiving sewage sludge. Environ. Pollut. 167, 101–109. https://doi.org/10.1016/j.envpol.2012.03.036
Cha, J., Cupples, A.M., 2009. Detection of the antimicrobials triclocarban and triclosan in agricultural soils following land application of municipal biosolids. Water Res. 43, 2522–2530. https://doi.org/10.1016/j.watres.2009.03.004

Chen, J., Liu, X., Li, L., Zheng, Jinwei, Qu, J., Zheng, Jufeng, Zhang, X., Pan, G., 2015. Consistent increase in abundance and diversity but variable change in community composition of bacteria in topsoil of rice paddy under short term biochar treatment across three sites from South China. Appl. Soil Ecol. 91, 68–79. https://doi.org/10.1016/j.apsoil.2015.02.012

Chen, Z., Song, Q., Cao, G., Chen, Y., 2008. Photolytic degradation of triclosan in the presence of surfactants. Chem. Pap. 62, 608-615. https://doi.org/10.2478/s11696-008-0077-0

Chuanchuen, R., Beinlich, K., Hoang, T.T., Becher, A., Karkhoff-Schweizer, R.R., Schweizer, H.P., 2001. Cross-resistance between triclosan and antibiotics in Pseudomonas aeruginosa is mediated by multidrug efflux pumps: exposure of a susceptible mutant strain to triclosan selects nfxB mutants overexpressing MexCD-OprJ. Antimicrob. Agents Chemother. 45, 428–432. https://doi.org/10.1128/AAC.45.2.428-432.2001

Chuanchuen, R., Karkhoff-Schweizer, R.R., Schweizer, H.P., 2003. High-level triclosan resistance in Pseudomonas aeruginosa is solely a result of efflux. Am. J. Infect. Control 31, 124–127. https://doi.org/10.1067/mic.2003.11

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., Tiedje, J.M., 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42, D633–D642. https://doi.org/10.1093/nar/gkt1244

Dann, A.B., Hontela, A., 2011. Triclosan: environmental exposure, toxicity and mechanisms of action. J. Appl. Toxicol. 31, 285-311. https://doi.org/10.1002/jat.1660

Dejonghe, W., Berteloot, E., Goris, J., Boon, N., Crul, K., Maertens, S., Höfte, M., Vos, P.D., Verstraete, W., Top, E.M., 2003. Synergistic Degradation of Linuron by a Bacterial Consortium and Isolation of a Single Linuron-Degrading Variovorax Strain. Appl. Environ. Microbiol. 69, 1532–1541. https://doi.org/10.1128/AEM.69.3.1532-1541.2003

Dhillon, G.S., Kaur, S., Pulicharla, R., Brar, S.K., Cledón, M., Verma, M., Surampalli, R.Y., 2015. Triclosan: Current Status, Occurrence, Environmental Risks and Bioaccumulation Potential. Int. J. Environ. Res. Public. Health 12, 5657–5684. https://doi.org/10.3390/ijerph120505657

Edwards, M., Topp, E., Metcalfe, C.D., Li, H., Gottschall, N., Bolton, P., Curnoe, W., Payne, M., Beck, A., Kleywegt, S., Lapen, D.R., 2009. Pharmaceutical and personal care products in tile drainage following surface spreading and injection of dewatered municipal biosolids to an agricultural field. Sci. Total Environ. 407, 4220–4230. https://doi.org/10.1016/j.scitotenv.2009.02.028

Faoagali, J., Fong, J., George, N., Mahoney, P., O'Rourke, V., 1995. Comparison of the immediate, residual, and cumulative antibacterial effects of Novaderm R, Novascrub R, Betadine Surgical Scrub, Hibiclens, and liquid soap. Am. J. Infect. Control 23, 337–343. https://doi.org/10.1016/0196-6553(95)90263-5

Glöckner, F.O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G., Yarza, P., Peplies, J., Westram, R., Ludwig, W., 2017. 25 years of serving the community with ribosomal RNA gene reference databases and tools. J. Biotechnol. 261, 169–176. https://doi.org/10.1016/j.jbiotec.2017.06.1198

Guo, Q., Yan, J., Wen, J., Hu, Y., Chen, Y., Wu, W., 2016. Rhamnolipid-enhanced aerobic biodegradation of triclosan (TCS) by indigenous microorganisms in water-sediment

systems. Sci. Total Environ. 571, 1304–1311. https://doi.org/10.1016/j.scitotenv.2016.07.171
Harrow, D.I., Felker, J.M., Baker, K.H., 2011. Impacts of Triclosan in Greywater on Soil Microorganisms. Appl. Environ. Soil Sci. 2011, 646750. https://doi.org/10.1155/2011/646750
Hay, A.G., Dees, P.M., Sayler, G.S., 2001. Growth of a bacterial consortium on triclosan. FEMS Microbiol. Ecol. 36, 105–112.
Heath, R.J., Rock, C.O., 2000. Microbiology: a triclosan-resistant bacterial enzyme. Nature 406, 145–146.
Heath, R.J., Su, N., Murphy, C.K., Rock, C.O., 2000. The Enoyl-[acyl-carrier-protein] Reductases FabI and FabL from Bacillus subtilis. J. Biol. Chem. 275, 40128–40133. https://doi.org/10.1074/jbc.M005611200
Heidler, J., Halden, R.U., 2007. Mass balance assessment of triclosan removal during conventional sewage treatment. Chemosphere 66, 362–369. https://doi.org/10.1016/j.chemosphere.2006.04.066
Helbing, C.C., Van Aggelen, G., Veldhoen, N., 2011. Triclosan Affects Thyroid Hormone–Dependent Metamorphosis in Anurans. Toxicol. Sci. 119, 417–418.
Higgins, C.P., Paesani, Z.J., Abbott Chalew, T.E., Halden, R.U., Hundal, L.S., 2011. Persistence of triclocarban and triclosan in soils after land application of biosolids and bioaccumulation in Eisenia foetida. Environ. Toxicol. Chem. 30, 556–563.
Hinther, A., Bromba, C.M., Wulff, J.E., Helbing, C.C., 2011. Effects of triclocarban, triclosan, and methyl triclosan on thyroid hormone action and stress in frog and mammalian culture systems. Environ. Sci. Technol. 45, 5395–5402.
Holmes, I., Harris, K., Quince, C., 2012. Dirichlet Multinomial Mixtures: Generative Models for Microbial Metagenomics. PLOS ONE 7, e30126. https://doi.org/10.1371/journal.pone.0030126
Huang, Y.-H., Lin, J.-S., Ma, J.-C., Wang, H.-H., 2016. Functional Characterization of Triclosan-Resistant Enoyl-acyl-carrier Protein Reductase (FabV) in Pseudomonas aeruginosa. Front. Microbiol. 7, 1903. https://doi.org/10.3389/fmicb.2016.01903
Jenkins, J.R., Viger, M., Arnold, E.C., Harris, Z.M., Ventura, M., Miglietta, F., Girardin, C., Edwards, R.J., Rumpel, C., Fornasier, F., Zavalloni, C., Tonon, G., Alberti, G., Taylor, G., 2017. Biochar alters the soil microbiome and soil function: results of next-generation amplicon sequencing across Europe. GCB Bioenergy 9, 591–612. https://doi.org/10.1111/gcbb.12371
Jiten Singh, N., Shin, D., Lee, H.M., Kim, H.T., Chang, H.-J., Cho, J.M., Kim, K.S., Ro, S., 2011. Structural basis of triclosan resistance. J. Struct. Biol. 174, 173–179.
Jones, G.L., Muller, C.T., O'Reilly, M., Stickler, D.J., 2006. Effect of triclosan on the development of bacterial biofilms by urinary tract pathogens on urinary catheters. J. Antimicrob. Chemother. 57, 266–272. https://doi.org/10.1093/jac/dki447
Junker, L.M., 2004. Effects of triclosan incorporation into ABS plastic on biofilm communities. J. Antimicrob. Chemother. 53, 989–996. https://doi.org/10.1093/jac/dkh196
Kagle, J., Hay, A.G., 2002. Construction of a broad host range cloning vector conferring triclosan resistance. BioTechniques 33, 490–492.

Kagle, J.M., Paxson, C., Johnstone, P., Hay, A.G., 2015. Identification of a gene cluster associated with triclosan catabolism. Biodegradation 26, 235–246. https://doi.org/10.1007/s10532-015-9730-9

Khan, R., Kong, H.G., Jung, Y.-H., Choi, J., Baek, K.-Y., Hwang, E.C., Lee, S.-W., 2016. Triclosan Resistome from Metagenome Reveals Diverse Enoyl Acyl Carrier Protein Reductases and Selective Enrichment of Triclosan Resistance Genes. Sci. Rep. 6, 32322. https://doi.org/10.1038/srep32322

Khan, R., Roy, N., Choi, K., Lee, S.-W., 2018. Distribution of triclosan-resistant genes in major pathogenic microorganisms revealed by metagenome and genome-wide analysis. PLOS ONE 13, e0192277. https://doi.org/10.1371/journal.pone.0192277

Kim, Y.-M., Murugesan, K., Schmidt, S., Bokare, V., Jeon, J.-R., Kim, E.-J., Chang, Y.-S., 2011. Triclosan susceptibility and co-metabolism – A comparison for three aerobic pollutant-degrading bacteria. Bioresour. Technol. 102, 2206–2212. https://doi.org/10.1016/j.biortech.2010.10.009

Kolpin, D.W., Furlong, E.T., Meyer, M.T., Thurman, E.M., Zaugg, S.D., Barber, L.B., Buxton, H.T., 2002. Pharmaceuticals, Hormones, and Other Organic Wastewater Contaminants in U.S. Streams, 1999–2000: A National Reconnaissance. Environ. Sci. Technol. 36, 1202–1211. https://doi.org/10.1021/es011055j

Kolton, M., Harel, Y.M., Pasternak, Z., Graber, E.R., Elad, Y., Cytryn, E., 2011. Impact of Biochar Application to Soil on the Root-Associated Bacterial Community Structure of Fully Developed Greenhouse Pepper Plants. Appl. Environ. Microbiol. 77, 4924–4930. https://doi.org/10.1128/AEM.00148-11

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. Appl. Environ. Microbiol. 79, 5112–5120. https://doi.org/10.1128/AEM.01043-13

Latch, D.E., Packer, J.L., Stender, B.L., VanOverbeke, J., Arnold, W.A., McNeill, K., 2005. Aqueous photochemistry of triclosan: Formation of 2,4-dichlorophenol, 2,8-dichlorodibenzo-p-dioxin, and oligomerization products. Environ. Toxicol. Chem. 24, 517–525. https://doi.org/10.1897/04-243R.1

Lee, D.G., Cho, K.-C., Chu, K.-H., 2014. Identification of triclosan-degrading bacteria in a triclosan enrichment culture using stable isotope probing. Biodegradation 25, 55–65.

Lee, D.G., Chu, K.-H., 2013. Effects of growth substrate on triclosan biodegradation potential of oxygenase-expressing bacteria. Chemosphere 93, 1904–1911. https://doi.org/10.1016/j.chemosphere.2013.06.069

Lee, D.G., Zhao, F., Rezenom, Y.H., Russell, D.H., Chu, K.-H., 2012. Biodegradation of triclosan by a wastewater microorganism. Water Res. 46, 4226–4234. https://doi.org/10.1016/j.watres.2012.05.025

Lee, E.Y., Jun, Y.S., Cho, K.-S., Ryu, H.W., 2002. Degradation characteristics of toluene, benzene, ethylbenzene, and xylene by *Stenotrophomonas maltophilia* T3-c. J. Air Waste Manag. Assoc. 52, 400–406.

Lehmann, J., Rillig, M.C., Thies, J., Masiello, C.A., Hockaday, W.C., Crowley, D., 2011. Biochar effects on soil biota – A review. Soil Biol. Biochem., 43, 1812–1836. https://doi.org/10.1016/j.soilbio.2011.04.022

Liu, J., Wang, J., Zhao, C., Hay, A.G., Xie, H., Zhan, J., 2016. Triclosan removal in wetlands constructed with different aquatic plants. Appl. Microbiol. Biotechnol. 100, 1459–1467. https://doi.org/10.1007/s00253-015-7063-6

Lolas, I.B., Chen, X., Bester, K., Nielsen, J.L., 2012. Identification of triclosan-degrading bacteria using stable isotope probing, fluorescence in situ hybridization and microautoradiography. Microbiology 158, 2796–2804.

Lozano, N., Rice, C.P., Ramirez, M., Torrents, A., 2010. Fate of triclosan in agricultural soils after biosolid applications. Chemosphere 78, 760–766. https://doi.org/10.1016/j.chemosphere.2009.10.043

Massengo-Tiassé, R.P., Cronan, J.E., 2008. Vibrio cholerae FabV Defines a New Class of Enoyl-Acyl Carrier Protein Reductase. J. Biol. Chem. 283, 1308–1316. https://doi.org/10.1074/jbc.M708171200

Maurathan, N., Orr, C., Ralebitso-Senior, T.K., 2015. Biochar adsorption properties and the impact on naphthalene as a model environmental contaminant and microbial community dynamics: A triangular perspective. In: Lucas Borja, M.E., editors. Soil Management: Technological Systems, Practices and Ecological Implications. Nova Science Publishers. p 64-90.

McMurry, L.M., Oethinger, M., Levy, S.B., 1998a. Triclosan targets lipid synthesis. Nature 394, 531–532. https://doi.org/10.1038/28970

McMurry, L.M., Oethinger, M., Levy, S.B., 1998b. Overexpression of marA, soxS, or acrAB produces resistance to triclosan in laboratory and clinical strains of Escherichia coli. FEMS Microbiol. Lett. 166, 305–309.

McNamara, P.J., LaPara, T.M., Novak, P.J., 2014. The Impacts of Triclosan on Anaerobic Community Structures, Function, and Antimicrobial Resistance. Environ. Sci. Technol. 48, 7393–7400. https://doi.org/10.1021/es501388v

Meade, M.J., Waddell, R.L., Callahan, T.M., 2001. Soil bacteria Pseudomonas putida and Alcaligenes xylosoxidans subsp. denitrificans inactivate triclosan in liquid and solid substrates. FEMS Microbiol. Lett. 204, 45–48.

Meinicke, P., 2015. UProC: tools for ultra-fast protein domain classification. Bioinformatics 31, 1382–1388. https://doi.org/10.1093/bioinformatics/btu843

Miller, T.R., Heidler, J., Chillrud, S.N., DeLaquil, A., Ritchie, J.C., Mihalic, J.N., Bopp, R., Halden, R.U., 2008. Fate of Triclosan and Evidence for Reductive Dechlorination of Triclocarban in Estuarine Sediments. Environ. Sci. Technol. 42, 4570–4576.

Mothur, 2013. https://www.mothur.org/wiki/MiSeq_SOP (accessed 5 October 2018).

Mulla, S.I., Hu, A., Xu, H., Yu, C.-P., 2015. Draft Genome Sequence of Triclosan-Degrading Bacterium Sphingomonas sp. Strain YL-JM2C, Isolated from a Wastewater Treatment Plant in China. Genome Announc. 3, e00603-15. https://doi.org/10.1128/genomeA.00603-15

Mulla, S.I., Wang, H., Sun, Q., Hu, A., Yu, C.-P., 2016. Characterization of triclosan metabolism in Sphingomonas sp. strain YL-JM2C. Sci. Rep. 6, 21965. https://doi.org/10.1038/srep21965

Noyce, G.L., Winsborough, C., Fulthorpe, R., Basiliko, N., 2016. The microbiomes and metagenomes of forest biochars. Sci. Rep. 6, 26425. https://doi.org/10.1038/srep26425

Ogunyoku, T.A., Young, T.M., 2014. Removal of Triclocarban and Triclosan during Municipal Biosolid Production. Water Environ. Res. Res. Publ. Water Environ. Fed. 86, 197–203.
Pycke, B.F.G., Roll, I.B., Brownawell, B.J., Kinney, C.A., Furlong, E.T., Kolpin, D.W., Halden, R.U., 2014. Transformation Products and Human Metabolites of Triclocarban and Triclosan in Sewage Sludge Across the United States. Environ. Sci. Technol. 48, 7881–7890. https://doi.org/10.1021/es5006362
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/nar/gks1219
Raut, S.A., Angus, R.A., 2010. Triclosan has endocrine-disrupting effects in male western mosquitofish, Gambusia affinis. Environ. Toxicol. Chem. 29, 1287–1291.
Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584. https://doi.org/10.7717/peerj.2584
Sabourin, L., Beck, A., Duenk, P.W., Kleywegt, S., Lapen, D.R., Li, H., Metcalfe, C.D., Payne, M., Topp, E., 2009. Runoff of pharmaceuticals and personal care products following application of dewatered municipal biosolids to an agricultural field. Sci. Total Environ. 407, 4596–4604. https://doi.org/10.1016/j.scitotenv.2009.04.027
Sanchez, P., Moreno, E., Martinez, J.L., 2005. The Biocide Triclosan Selects Stenotrophomonas maltophilia Mutants That Overproduce the SmeDEF Multidrug Efflux Pump. Antimicrob. Agents Chemother. 49, 781–782. https://doi.org/10.1128/AAC.49.2.781-782.2005
Simpson, E.H., 1949. Measurement of Diversity. Nature 163, 688. https://doi.org/10.1038/163688a0
Singer, H., Müller, S., Tixier, C., Pillonel, L., 2002. Triclosan: Occurrence and Fate of a Widely Used Biocide in the Aquatic Environment: Field Measurements in Wastewater Treatment Plants, Surface Waters, and Lake Sediments. Environ. Sci. Technol. 36, 4998–5004. https://doi.org/10.1021/es025750i
Smebye, A., Alling, V., Vogt, R.D., Gadmar, T.C., Mulder, J., Cornelissen, G., Hale, S.E., 2016. Biochar amendment to soil changes dissolved organic matter content and composition. Chemosphere, 142, 100–105. https://doi.org/10.1016/j.chemosphere.2015.04.087
Sørensen, S.R., Albers, C.N., Aamand, J., 2008. Rapid Mineralization of the Phenylurea Herbicide Diuron by Variovorax sp. Strain SRS16 in Pure Culture and within a Two-Member Consortium. Appl. Environ. Microbiol. 74, 2332–2340. https://doi.org/10.1128/AEM.02687-07
Stewart, M.J., Parikh, S., Xiao, G., Tonge, P.J., Kisker, C., 1999. Structural basis and mechanism of enoyl reductase inhibition by triclosan1. J. Mol. Biol. 290, 859–865. https://doi.org/10.1006/jmbi.1999.2907
Svenningsen, H., Henriksen, T., Priemé, A., Johnsen, A.R., 2011. Triclosan affects the microbial community in simulated sewage-drain-field soil and slows down xenobiotic degradation. Environ. Pollut. 159, 1599–1605. https://doi.org/10.1016/j.envpol.2011.02.052
Voelker, R., 2016. Say Goodbye to Some Antibacterials. JAMA 316, 1538–1538. https://doi.org/10.1001/jama.2016.14612

Wang, L.-Q., Falany, C.N., James, M.O., 2004. Triclosan as a substrate and inhibitor of 3′-phosphoadenosine 5′-phosphosulfate-sulfotransferase and UDP-glucuronosyl transferase in human liver fractions. Drug Metab. Dispos. 32, 1162–1169.
Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl. Environ. Microbiol. 73, 5261–5267. https://doi.org/10.1128/AEM.00062-07
Watzinger, A., Feichtmair, S., Kitzler, B., Zehetner, F., Kloss, S., Wimmer, B., Zechmeister-Boltenstern, S., Soja, G., 2014. Soil microbial communities responded to biochar application in temperate soils and slowly metabolized 13C-labelled biochar as revealed by 13C PLFA analyses: results from a short-term incubation and pot experiment. Eur. J. Soil Sci. 65, 40–51. https://doi.org/10.1111/ejss.12100
Webber, M.A., Randall, L.P., Cooles, S., Woodward, M.J., Piddock, L.J.V., 2008. Triclosan resistance in Salmonella enterica serovar Typhimurium. J. Antimicrob. Chemother. 62, 83–91. https://doi.org/10.1093/jac/dkn137
White, J.R., Nagarajan, N., Pop, M., 2009. Statistical Methods for Detecting Differentially Abundant Features in Clinical Metagenomic Samples. PLOS Comput. Biol. 5, e1000352. https://doi.org/10.1371/journal.pcbi.1000352
Wu, C., Spongberg, A.L., Witter, J.D., 2009. Adsorption and Degradation of Triclosan and Triclocarban in Soils and Biosolids-Amended Soils. J. Agric. Food Chem. 57, 4900–4905. https://doi.org/10.1021/jf900376c
Xia, K., Hundal, L.S., Kumar, K., Armbrust, K., Cox, A.E., Granato, T.C., 2010. Triclocarban, triclosan, polybrominated diphenyl ethers, and 4-nonylphenol in biosolids and in soil receiving 33-year biosolids application. Environ. Toxicol. Chem. 29, 597–605. https://doi.org/10.1002/etc.66
Xu, N., Tan, G., Wang, H., Gai, X., 2016. Effect of biochar additions to soil on nitrogen leaching, microbial biomass and bacterial community structure. Eur. J. Soil Biol. 74, 1–8. https://doi.org/10.1016/j.ejsobi.2016.02.004
Yazdankhah, S.P., Scheie, A.A., Høiby, E.A., Lunestad, B.-T., Heir, E., Fotland, T.Ø., Naterstad, K., Kruse, H., 2006. Triclosan and antimicrobial resistance in bacteria: an overview. Microb. Drug Resist. 12, 83–90.
Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., Glöckner, F.O., 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. 42, D643–D648. https://doi.org/10.1093/nar/gkt1209
Ying, G.-G., Yu, X.-Y., Kookana, R.S., 2007. Biological degradation of triclocarban and triclosan in a soil under aerobic and anaerobic conditions and comparison with environmental fate modelling. Environ. Pollut. 150, 300–305. https://doi.org/10.1016/j.envpol.2007.02.013
Yue, J.C., Clayton, M.K., 2005. A Similarity Measure Based on Species Proportions. Commun. Stat. - Theory Methods 34, 2123–2131. https://doi.org/10.1080/STA-200066418
Zaayman, M., Siggins, A., Horne, D., Lowe, H., Horswell, J., 2017. Investigation of triclosan contamination on microbial biomass and other soil health indicators. FEMS Microbiol. Lett. 364, 1-6. https://doi.org/10.1093/femsle/fnx163
Zhang, C., Lin, Y., Tian, X., Xu, Q., Chen, Z., Lin, W., 2017. Tobacco bacterial wilt suppression with biochar soil addition associates to improved soil physiochemical properties

and increased rhizosphere bacteria abundance. Appl. Soil Ecol. 112, 90–96. https://doi.org/10.1016/j.apsoil.2016.12.005
Zhao, C., Xie, H., Xu, J., Xu, X., Zhang, J., Hu, Z., Liu, C., Liang, S., Wang, Q., Wang, J., 2015. Bacterial community variation and microbial mechanism of triclosan (TCS) removal by constructed wetlands with different types of plants. Sci. Total Environ. 505, 633–639. https://doi.org/10.1016/j.scitotenv.2014.10.053
Zhu, L., Lin, J., Ma, J., Cronan, J.E., Wang, H., 2010. Triclosan Resistance of Pseudomonas aeruginosa PAO1 Is Due to FabV, a Triclosan-Resistant Enoyl-Acyl Carrier Protein Reductase. Antimicrob. Agents Chemother. 54, 689–698. https://doi.org/10.1128/AAC.01152-09
Zorrilla, L.M., Gibson, E.K., Jeffay, S.C., Crofton, K.M., Setzer, W.R., Cooper, R.L., Stoker, T.E., 2009. The Effects of Triclosan on Puberty and Thyroid Hormones in Male Wistar Rats. Toxicol. Sci. 107, 56–64. https://doi.org/10.1093/toxsci/kfn225

SUPPLEMENTARY MATERIALS

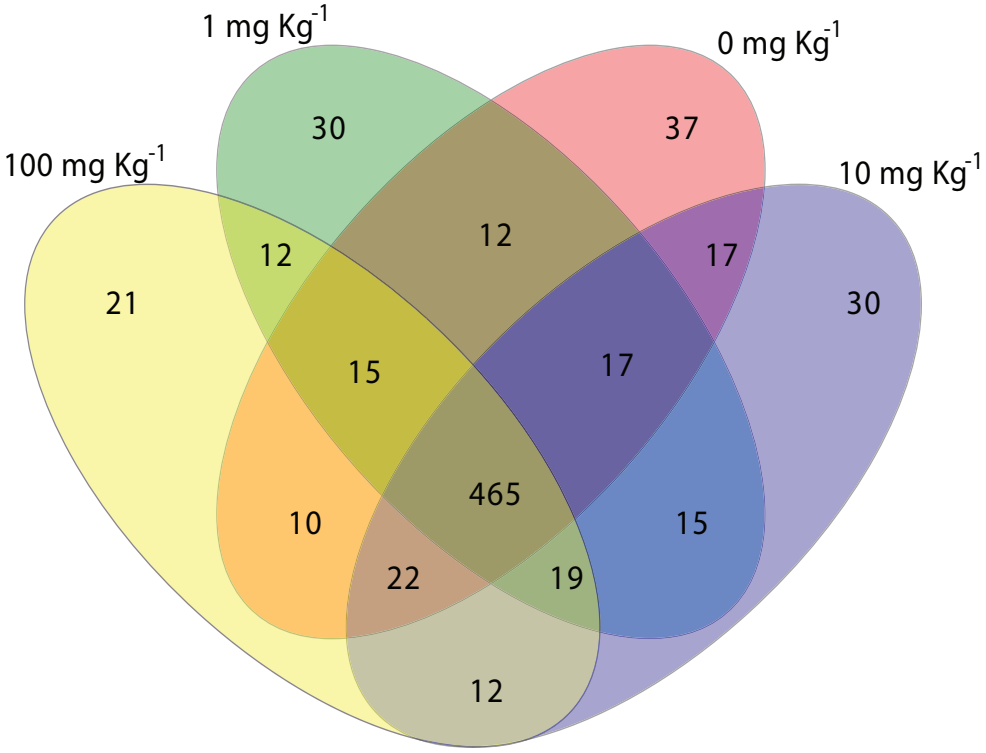


Fig S1. Number of shared and unique OTUs at each triclosan treatment (1, 10 and 100 mg Kg^{-1}) and control (0 mg Kg^{-1}). The richness was similar for all groups. The number of OTUs at 0 mg Kg^{-1}, 1 mg Kg^{-1}, 10 mg Kg^{-1}, and 100 mg Kg^{-1} were 595, 585, 597, and 576, respectively. The total richness was 734 OTUs

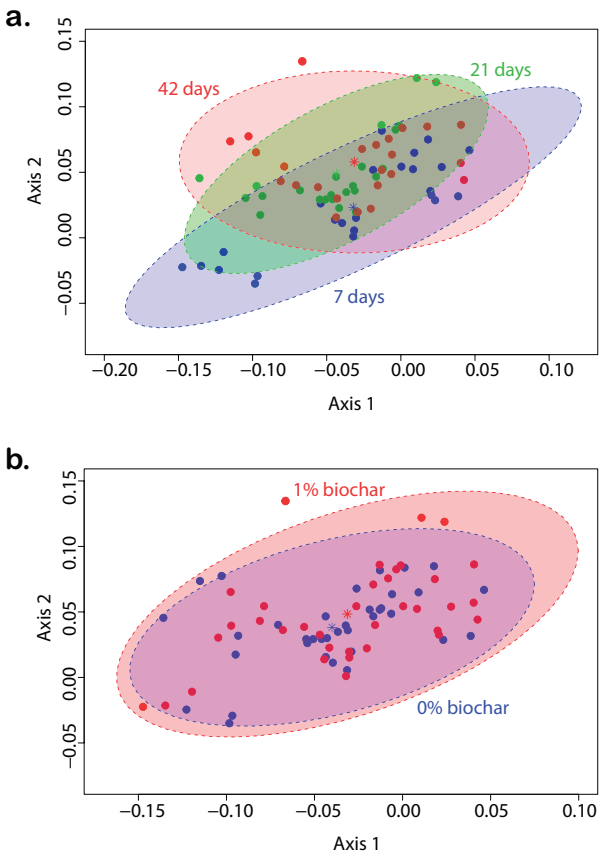


Fig S2. NMDS analysis of the bacterial community structures using θYC dissimilarity distances. Distances between symbols on the NMDS plot reflect relative dissimilarities in the community structures. The 2 axes represent 95% of the variance. The lowest stress is 0.104 with an R2 value of 0.97. Ellipses represent the 95% confidence intervals around the centroid for each cluster. The ellipse centroids are indicated with ✳. **a.** AMOVA showed that the bacterial communities sampled on days 21 and 42 were not different from each other ($p=0.216$), but were different from day 7 samples ($p<0.01$). **b.** The bacterial communities were not significantly impacted by biochar ($p=0.678$).

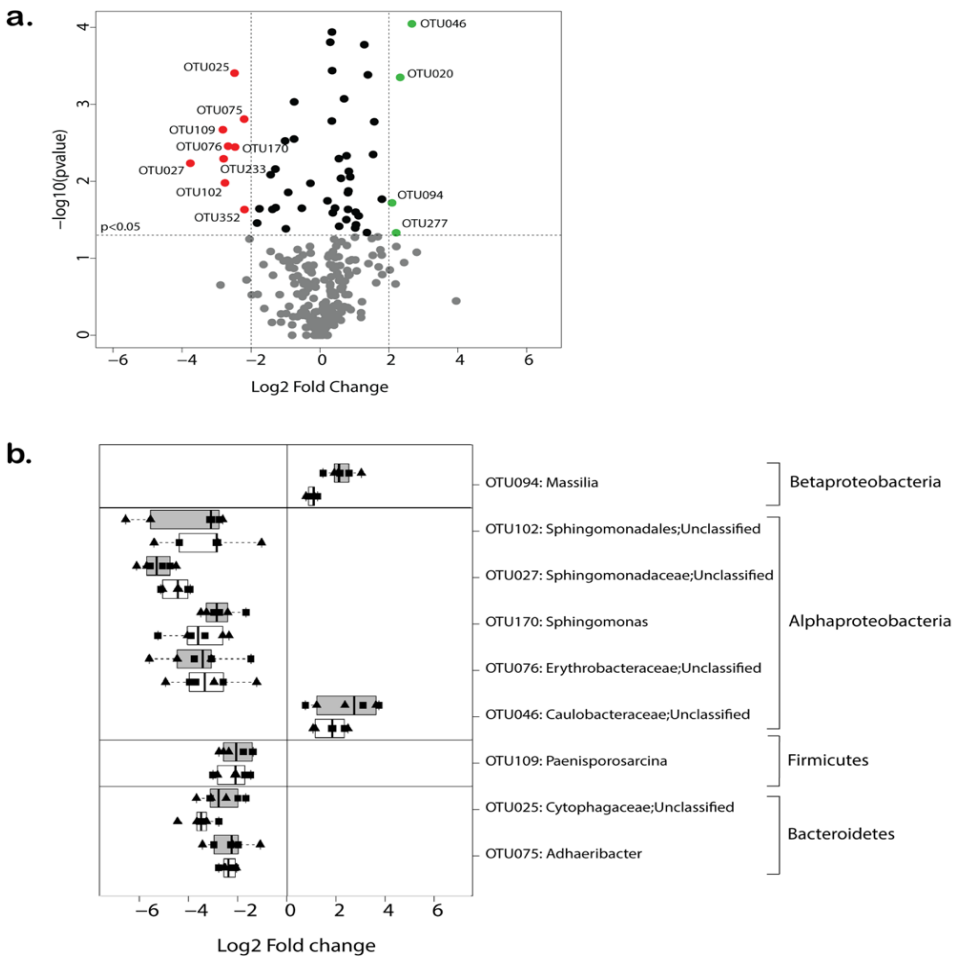


Fig S3. More OTUs decreased (7) in abundance than increased (2) after exposure to 10 mg Kg^{-1} triclosan as compared to the control. **a.** The volcano plot shows the fold change (x axis) in OTUs between 10 mg Kg^{-1} and the control for all time points, as well as the statistical significance of that change (y axis) as determined by NPM analysis. OTUs above the dotted line were differentially abundant (adjusted $p<0.05$). Green OTUs increased more than 2-fold in 10 mg Kg^{-1} triclosan treatment, while red OTUs decreased more than 2-fold. **b.** Change in abundance of OTUs significantly impacted by triclosan from days 21 and 42 in the presence (grey boxes) and absence (white boxes) of biochar.

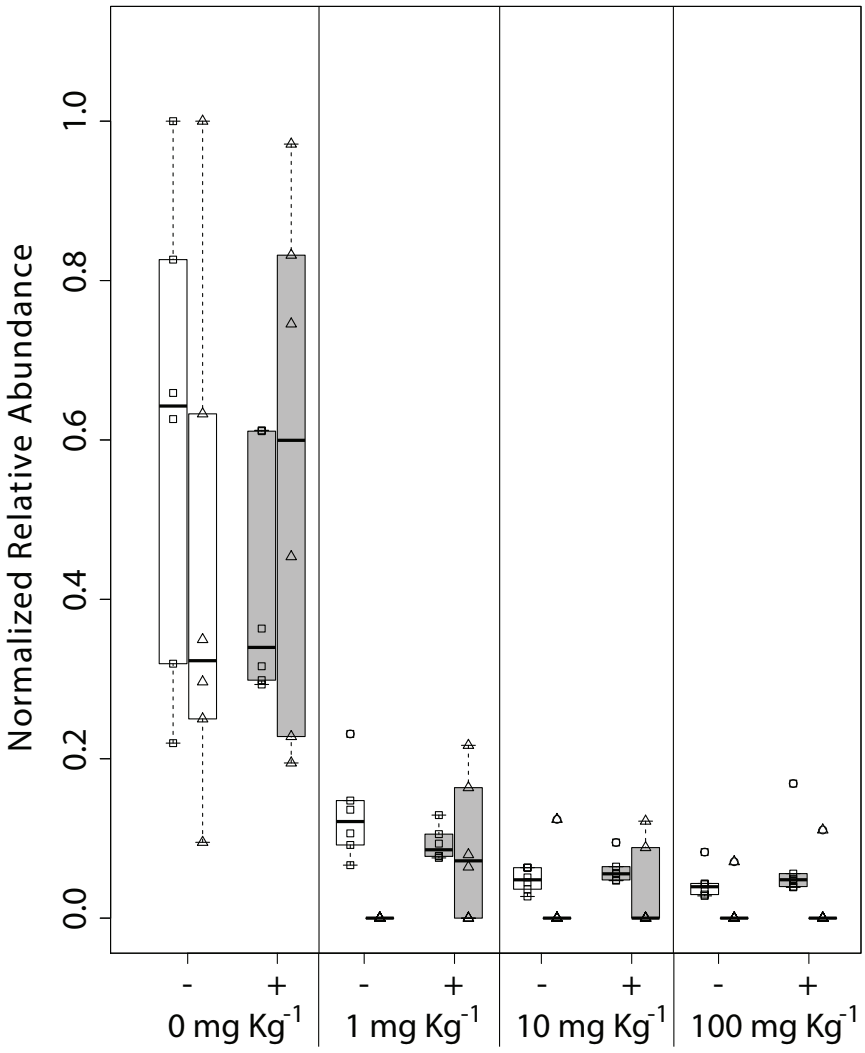


Fig S4. Triclosan exposure negatively affected the relative abundance of *Sphingomonas* (□) and *Sphingopyxis* (Δ) at 21 and 42 days by triclosan treatment in presence (+) and absence (-) of biochar. The relative abundance of each genus was normalized to the most abundant replicate in the control. The maximum relative abundances for *Sphingomonas* and *Sphingopyxis* in the control were 0.0353% and 0.0004%, respectively.

Table S1. Ratio of the relative abundance of OTUs significantly affected by biochar (NPM analysis, adjusted $p<0.05$) summed up at phylum level for each triclosan treatment. A ratio of 1 indicates no effect of biochar, <1 that biochar had a negative impact, and >1 that biochar positively affected the population.

Phylum	Triclosan concentration			
	0 mg Kg^{-1}	1 mg Kg^{-1}	10 mg Kg^{-1}	100 mg Kg^{-1}
Acidobacteria	0.9	0.5*	NA	NA
Actinobacteria	1.0	1.2*	0.8*	0.8*
Bacteroidetes	2.2*	2.3*	1.4	0.9
Firmicutes	1.2	1.3	0.9	1.0
Proteobacteria	1.0	0.9	1.6*	1.4*

* indicates significant difference of relative abundances between the presence (1%) and absence (0%) of biochar (Wilcoxon test, $p<0.05$). NA: no significant impact of biochar on any OTU belonging to that phylum (NPM analysis, $p>0.05$)

Table S2. Summary of significant treatments according to different statistical analyses

Analysis	**Time zero**	**0 mg Kg^{-1}**	**1 mg Kg^{-1}**	**10 mg Kg^{-1}**	**100 mg Kg^{-1}**
NMDS and AMOVA analysis using θYC distances		A	B	C	D
DMM modeling	I	II	II	III	III
CLC using relative abundances of significantly different OTUs found by NPM analysis	a	b	b	c	d

Table S3. The "X" indicates NCBI BLAST hits (≥ 27% predicted amino acid sequence identity) for triclosan-resistance determinants in bacteria taxa that were present in the 100 mg Kg^{-1} treatment, but not in the control (0 mg Kg^{-1}), or which increased more than 2-fold in response to triclosan.

Bacterial taxa	**NCBI taxa ID**	**FabI**	**FabK**	**FabV**	**AcrB**
Stenotrophomonas	40323	X		X	X
Massilia	149698	X			X
Janthinobacterium	29580	X			X
Variovorax	34072	X		X	X
Caulobacteraceae	76892	X		X	X
Pedobacter	84567			X	X
Flavobacteriaceae	49546	X	X	X	X
*Gillisia**	244698				X
*Flavobacterium**	237			X	X
Nakamurellaceae	85031	X			X
Frigoribacterium	96492				X
Luteibacter	242605	X		X	X
Nocardiopsis	2013	X			X
Chitinophaga	79328		X	X	X
Actinomycetospora	402649	X			
Achromobacter	222	X			X

* results from tBLASTn analyses. No hits were obtained for either FabL or 7-α-HSDH

Table S4. Total number of FabIs and their predicted sensitivity to triclosan in the genomes of *Sphingomonas* species, including triclosan degraders (*Sphingomonas sp.* RD1 and *Sphingomonas sp.* YL-JM2C: grey box).

Organism	Total predicted FabI	potentially sensitive FabI	potentially resistant FabI	NCBI Accession number
Sphingomonas sp. RD1	4	1	3	*Unpublished data*
Sphingomonas sp. YL-JM2C	3	1	2	ASTM00000000
S. wittichii RW1*	1	1	0	NC_009511
*S. histidinilytica**	1	1	0	NZ_FUYM00000000.1
*S. haloaromaticamans**	1	1	0	MIPT01000001.1
Sphingomonas sp. NIC1	1	1	0	NZ_CP015521
Sphingomonas sp. MM-1	1	1	0	NC_020561
Sphingomonas sp. LM7	1	1	0	NZ_CP019511
Sphingomonas sp. LK11	1	1	0	NZ_CP013916
Sphingomonas sp. JJ-A5	1	1	0	NZ_CP018221
Sphingomonas sp. ABOJV	1	1	0	NZ_CP018820

* closest relatives to *Sphingomonas sp.* RD1 and *Sphingomonas sp.* YL-JM2C

CHAPTER 3

SEX-DEPENDENT DISTURBANCE EFFECT OF GLYPHOSATE ON THE MOUSE GUT MICROBIOME

Vienvilay Phandanouvong-Lozano[a], Alexandra Fabian[b], Erin Daugherity[b], Anthony G. Hay[a]

[a]Department of Microbiology, Cornell University, Ithaca, NY, 14853, USA
[b]Center for Animal Resources and Education, Cornell University, Ithaca, NY, 14853, USA

ABSTRACT

Glyphosate is an herbicide used worldwide to control the growth of weeds and grasses in agricultural crops. Although glyphosate specifically targets plants with sensitive a glyphosate 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), little is known about its impact on intestinal bacteria some of which have EPSPS that are known to be sensitive. Here, we exposed male and female mice to two doses of a glyphosate-based herbicide (GBH, 0.07 mg glyphosate/L, and 0.7 mg glyphosate/L), and collected fecal samples from infancy to adulthood. Male mice receiving 0.07 mg glyphosate/L gained more weight than controls after 8 weeks. Analysis of 16S rRNA gene amplicons from mouse feces revealed a sex-dependent impact, with GBH only affecting male mice. Higher bacterial diversity and richness were found in male mice after 8 weeks of exposure to the low GBH-dose. Faith's index suggested that the affected bacteria were phylogenetically diverse. When compared with the control, 11 out of the 13 significantly altered OTUs in the low dose showed increased abundance, whereas 8 out of the 11 significantly altered OTUs in the high dose showed decreased abundance. Furthermore, weighted

Unifrac distances compared over time showed that the low GBH-dose appeared to disturb the maturation of the mouse microbiome. Our study suggests that low dose exposures to GBH might have greater impact on mammals than has previously been considered and suggests that more studies with lower glyphosate doses than the maximum contaminant level (MCL) established for the drinking water are required.

Keywords: herbicide, 16S rRNA gene deep sequencing, diversity indexes, weighted Unifrac, volatility

3.1. INTRODUCTION

The use of herbicides is a regular practice in agriculture and landscape management to control the growth of unwanted weeds and lawns (Gianessi, 2013). Glyphosate is the most widely used herbicide in the world and works by inhibiting the 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), which is an essential enzyme in plants for the biosynthesis of aromatic amino acids (Benbrook, 2016). Genetically modified crops with a glyphosate-resistant EPSPS from *Agrobacterium sp.* CP4 were introduced in the late 1990s and quickly came to dominate the US markets for corn, soy, and cotton (Funke et al., 2006; Myers et al., 2016; Swanson et al., 2014). The wide-spread occurrence of glyphosate residues in the environment has raised concerns about glyphosate's impact on ecosystems and human health (Battaglin et al., 2014; Myers et al., 2016). In fact, the World Health Organization has recently reclassified glyphosate as a probable carcinogen in humans (Bai and Ogbourne, 2016; IARC, 2015), though the carcinogenic risk posed by glyphosate is still hotly debated (Tarazona et al., 2017).

Although glyphosate-based herbicides were primarily introduced to target sensitive-glyphosate EPSPS weeds, numerous studies have demonstrated glyphosate's antimicrobial activity on non-target microorganisms (Ackermann et al., 2015; Shehata et al., 2013). The shikimate pathway is not exclusively found in plants, many bacterial and fungal species use this pathway to synthesize the aromatic amino acids: tyrosine, phenylalanine, and tryptophan (Zhi et al., 2014). Based on the molecular weight, sequence homology, and tolerance to glyphosate, the EPSPS enzymes can be classified in two phylogenetic clusters: Class I and Class II (Mir et al., 2015; Zhi et al., 2014). While the Class I enzymes are sensitive to glyphosate and are present in all plants and some bacteria like *Escherichia coli;* the Class II enzymes can tolerate high concentrations of glyphosate and can be found in bacteria like *Staphylococcus aureus* (Funke et al., 2006; Priestman et al., 2005). Mammals including humans do not synthesize aromatic amino acids *de novo* and therefore they do not have the shikimate pathway (Mir et al., 2015). For that reason, glyphosate was firstly believed to confer minimal risk to mammals; but humans and other mammals rely on their diet and on gut bacteria for aromatic amino acids (EFSA, 2015). Thus, glyphosate's impact on non-target organisms including gut microbes is being reevaluated (Myers et al., 2016).

Glyphosate's effects on non-target organisms range from changing behavioral patterns in honey bees to reducing reproduction of soil-dwelling earthworms (Balbuena et al., 2015; Gaupp-Berghausen et al., 2015; Van Bruggen et al., 2018). Glyphosate has been found responsible for inhibiting bacterial growth and disturbing microbial communities from diverse environments including soil, rhizosphere, cattle rumen, and poultry gastrointestinal tract (Ackermann et al., 2015; Nicolas et al., 2016; Shehata et al., 2013; Tsui and Chu, 2003). Furthermore, the bacterial response to glyphosate has been found to be species and even strain-specific (Aristilde et al.,

2017), and thus may be difficult to discern from amplicon based sequencing of different microbial communities which typically only provide genus level resolution.

Gut microbiota play critical roles in the wellbeing of the host by affecting maturation of the immune system, niche exclusion of pathogens, fermentation of non-digestible fibers, anaerobic metabolism of proteins, and xenobiotic biotransformation (Kho and Lal, 2018). Moreover, imbalances in the gut bacterial communities could lead to health disorders like inflammatory bowel disease, obesity and diabetes (Frank et al., 2007; Gao et al., 2015; Surana and Kasper, 2017). The gut microbiota constantly interacts with environmental chemicals or xenobiotics, metabolizing some directly upon ingestion or after metabolism by the liver (Björkholm et al., 2009; Claus et al., 2016). Some xenobiotics may inhibit bacterial growth, thereby altering the gut microbiome composition and/or interfering with the bacterial metabolic activities, which could result in deleterious effects for the host (Claus et al., 2016). Given the demonstrated antibacterial activity of glyphosate on intestinal bacteria, it has been hypothesized that glyphosate would exert a negative effect on the host if consumed in the diet (Ackermann et al., 2015; Shehata et al., 2013).

The impact of glyphosate on the gut microbiome has not been fully investigated. The few papers that have been published on mouse gut microbiome vary with respect to glyphosate concentration, formulation, and route of administration (Lozano et al., 2018; Mao et al., 2018; Nielsen et al., 2018). Here, we describe changes to the intestinal bacterial communities of mice raised on a low glyphosate, high fat diet after consuming a glyphosate-based herbicide (GBH) in their drinking water at levels below the Maximum Contaminate Level (MCL) established to protect human health in drinking water (US EPA, 2015)

3.2. MATERIALS AND METHODS

3.2.1. Mouse Feed and Chemicals

C57BI/6/J mice were fed a refined high fat diet (D12450, Research Diets, Inc), which was tested for glyphosate content using a glyphosate ELISA kit and following the manufacturer's instructions (Abraxis, Inc). The refined high fat diet contained very low levels of glyphosate (111.5 ± 7.8 ng/g), which was approximately nine times less than the standard maintenance diet Teklad LM-485 from Envigo (1000.8 ± 216.6 ng/g). GlyStar® Plus (Albaugh, LLC), a commercial glyphosate-based herbicide (GBH), was use as glyphosate source to expose mice through the drinking water.

3.2.2. Experimental Conditions

All mice used in this study belonged to the inbred strain C57BI/6/J obtained from Jackson Labs (Bar Harbor, ME). The mice were handled according to the regulations of the Institutional Animal Care and Use Committee (IACUC) of Cornell University, and kept in the facilities of the Center for Animal Resources and Education (CARE), Cornell University, Ithaca, NY. To minimize any potential effects of glyphosate present in the maintenance diet provided by the Cornell vivarium, mice were fed with the refined low fat version of the diet mentioned above (D12450, Research Diets, Inc), and were bred for two generations. The F2 generation of newborn mice stayed with their mothers until weaning, and later were separated by gender and treatment group. To minimizing confounding parental effects and mother-to-offspring microbiota vertical transmission, mice from each litter were split between treatment groups and the control.

Infant six-week old male and female mice received GBH in the drinking water *ad libitum* at 0.07 mg glyphosate/L (low dose) and 0.7 mg glyphosate/L (high dose). New batches of GBH-supplemented water were prepared weekly. There was no supplementation of GBH in the drinking water for the control group. Mouse weights and fecal samples were collected at the beginning of GBH supplementation (week 0), and weekly thereafter for eight weeks. Due to resource limitations, only week 1 and week 8 samples after GBH supplementation were analyzed by deep 16S rRNA gene sequencing. The collected fecal samples were stored at -20°C for further analysis. Pellets from five male and five female mice were chosen at random from each experimental group for further analysis.

3.2.3. DNA Extraction from Fecal Pellets

DNA was extracted from fecal pellets using a modified protocol from Yu and Forster (2005). Briefly, fecal pellets arrayed in 96-well deep plates were resuspended with lysis buffer [2% (wt/vol) SDS, 100 mM Tris HCl, 5mM EDTA, 200 mM NaCl, pH 8.0] and 2 M ammonium acetate; and then bead-beaten with 0.1 mm glass beads for 2 min at maximum speed (Mini-Beadbeater-96, BioSpec Products). Later, the DNA extracts were purified with a 96-well DNA Clean and Concentrator kit following the manufacturer's instructions (Zymo Research). DNA concentrations were measured via a fluorescence approach using the Quant-iT™ PicoGreen™ dsDNA assay kit and following the manufacturer's instructions (ThermoFisher Scientific). DNA extracts were kept at -20°C for further analysis.

3.2.4. 16S rRNA Amplification and Sequencing

Following the Mothur Wet-lab SOP and recommendations of Kozich et al. (2013), the V4 region of the 16S rRNA gene was amplified via PCR using barcoded-primers (Mothur, 2013). The V4-16S rRNA gene specific primer sequences were: forward primer 515F 5' – GTGYCAGCMGCCGCGGTAA – 3' and reverse primer 806R 5' – GGACTACNVGGGTWTCTAAT – 3'. Each PCR reaction was performed in triplicate in a Peltier Thermal Cycler PTC-200 (MJ Research) and using AccuPrime^{TM} *Pfx* DNA polymerase (ThermoFisher Scientific) and 2 ng of template DNA. A commercial mock community containing DNA from 8 different bacterial species (Zymo Research) was also amplified and prepared for sequencing as positive control. The PCR conditions were as follows: initial denaturation at 95°C for 2 min; 30 cycles of denaturation at 95°C for 20 s, annealing at 50°C for 15 s, and extension at 72°C for 5 min; and final extension at 72°C for 10 min. All PCR products were combined and visualized with gel electrophoresis. A SequalPrep^{TM} Normalization plate kit (ThermoFisher Scientific) was used to clean up and normalize the PCR products. The pooled amplicon library was sent to the Institute of Biotechnology at Cornell University to be sequenced on an Illumina instrument using a MiSeq V2 kit with paired-end 2x250 bp reads and following the manufacturer's protocol (Illumina).

3.2.5. Sequence Processing

Raw 16S rRNA gene sequence reads were analyzed using Mothur v1.40.0 according to the Mothur MiSeq SOP (Mothur, 2013) and Kozich et al. (2013). Briefly, forward and reverse sequence reads were aligned and grouped into contigs based on PANDseq quality scores. All sequence reads were screened and trimmed to 275 bp. The quality filtered sequences were

aligned to the Silva 16S rRNA gene reference database, release 132 (Quast et al., 2013; Yilmaz et al., 2014). Later, the aligned sequences were preclustered allowing up to 2 bp difference between sequences over 250bp. Chimeric sequences were removed using the VSEARCH algorithm, which identified 5.54% of the sequence reads as chimeras (Rognes et al., 2016). Unspecific sequence reads that might correspond to 18S rRNA gene, archaeal, chloroplast or mitochondrial DNA were also removed. The final dataset then contained in total 1,226,234 processed sequences, ranged from 2,138 to 42,595 per sample. Lastly, by using the Ribosomal Database Project, release 16 (Cole et al., 2014), and following the protocol of Wang et al., (2007) with 100 bootstrap iterations and 97% confidence cutoff, the processed sequences were clustered and classified into Operational Taxonomic Units (OTUs). The mock community was processed along with the entire amplicon library and was used to assess the error rate of the sequence processing. This analysis showed a very low error (<0.0005%).

3.2.6. Bacterial Diversity and Community Analyses

Bacterial diversity and community analyses were carried out using Mothur v1.40.0. Specific R packages for handling phylogenetic sequencing data, plotting data, and performing statistical analysis as mentioned below were installed in R version 3.4.4 and RStudio 1.1.442.

In order to perform comparative analysis, all samples were subsampled to the smallest library (2,138 sequences), and prior to each analysis they were rarified with 1,000 randomizations. Rarefaction analysis showed that under these conditions all samples had a coverage above 99%, demonstrating that the number of analyzed sequences was representative of the diversity in each sample.

Bacterial alpha diversity was assessed using three different diversity indexes: Shannon (richness and eveness), Faith (phylogenetic differences), and Chao1 estimator (unique OTUs). Results from the diversity analyses were plotted with the ggplot2 package 3.1.0 for R, and the Venn diagrams with the VennDiagram package 1.6.20 for R. The statistical analyses were performed with a linear mixed-effects model from the lme4 package 1.1-17 for R, which corrected for the effect of the intrinsic mouse variability (Bates et al., 2015).

The bacterial communities between samples were compared by using pairwise similarity matrices generated with weighted Unifrac distances. The weighted Unifrac metric not only accounts for differential abundance of OTUs between the samples, but it also considers the relatedness of the community members by calculating the phylogenetic distances between them (Lozupone et al., 2011). The non-metric multidimensional scaling (NMDS) approach was used to visualize the Weighted Unifrac distances in ordination plots. The package plot3D 1.1.1 for R was used to generate the NMDS plots. The spatial separation between the groupings in the NMDS plots was statistically evaluated with the Analysis of Molecular Variance (AMOVA) in Mothur v1.40.0.

3.2.7. Differential Bacterial Abundance among Experimental Groups

The relative abundances of OTUs that were significantly different between the GBH treatments and the control were identified using the DESeq2 package 1.18.1 implemented in Bioconductor 3.9 for R (Bioconductor, 2003). DeSeq2 uses a negative binomial distribution which not only considers the mean but also the dispersion of the variance, providing a more balanced selection of OTUs differentially abundant while controlling for the probability of false discoveries (Anders and Huber, 2010). Fold changes of differentially abundant OTUs were visualized on volcano

plots using R and the maptools package 0.9-4 for R. To calculate the fold-change in unique OTUs the rarefaction cutoff to the smallest library (2,138 reads) was used to determine the detection limit (1/2,138). Then, a minimal fold-change was estimated using the count number of sequences for each unique OTUs.

3.2.8. Temporal Volatility Analysis

The weighted Unifrac distances were also used to evaluate the glyphosate's effect on the temporal volatility of the bacterial communities within treatment groups. By comparing the shifts of the bacterial communities over the time, the volatility analysis allows measurement and tracking of disturbance and stability stages of the microbiome (DiGiulio et al., 2015; Halfvarson et al., 2017). Pairwise comparisons of the weighted Unifrac distances were performed among mice at each sampling time, and were plotted using the ggplot2 package 3.1.0 for R. The significance of the weighted Unifrac distance shifts over time was tested with a linear mixed-effects model from the lme4 package 1.1-17 for R, which not only accounted for the intrinsic mouse variability, but also for the shift of each pairwise comparison over the time (Bates et al., 2015).

3.3. RESULTS AND DISCUSSION

3.3.1. Sex-Dependent Impact of Glyphosate on Mouse Weight and Gut Bacteria

Both male and female mice gained weight as expected over the course of the experiment (Fig 1A). Male mice showed higher body weights than females at all sampling times in both GBH treatments and in the control (no GBH addition). At the end of this study, however, males

exposed to the low GBH-dose (0.07 mg glyphosate/L) had significantly higher weight and higher total weight gain than those in either the control or the high GBH-dose (0.7 mg glyphosate/L) (Fig 1A and Supporting Information Fig. S1, ANOVA $p<0.01$, TukeyHSD adjusted $p<0.01$). The female body weights did not show any differences in the GBH treatments at any of the sampling times.

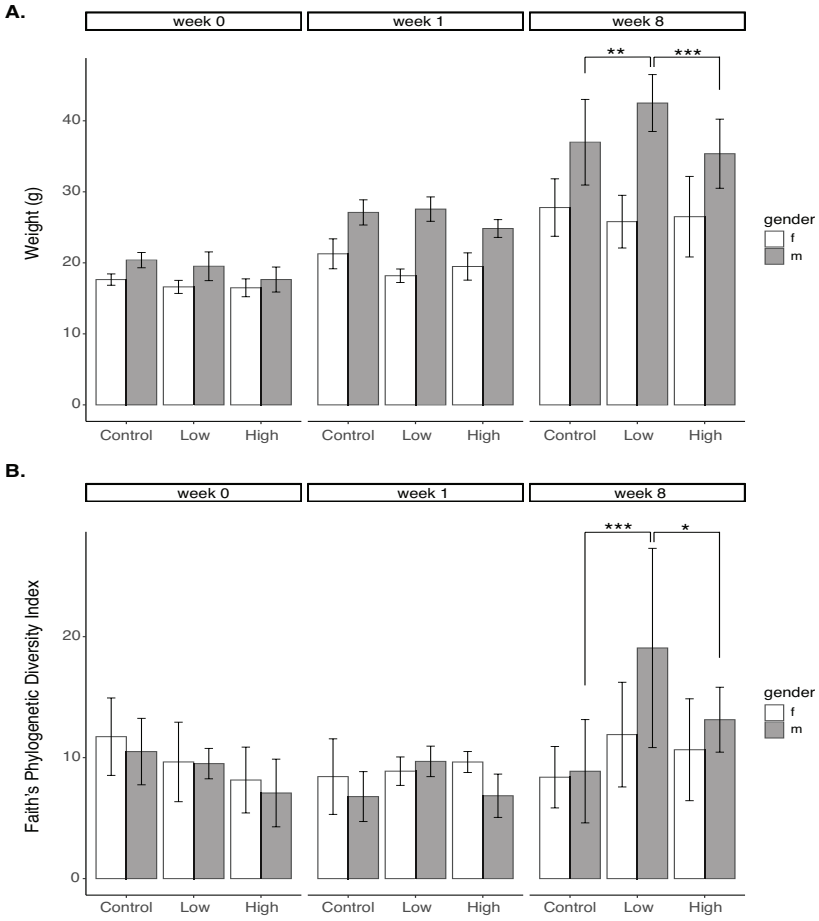


Fig 1. Body weight **(A)** and Faith's phylogenetic diversity index **(B)** over time showed significant effect of GBH on male mice at low dose (0.07 mg glyphosate/L) compared to higher dose (0.7 mg glyphosate/L) and to the control (no GBH addition) after 8 weeks of exposure (week 8) (ANOVA $p<0.001$, TukeyHSD adjusted *$p<0.05$, **$p<0.01$, ***$p<0.001$). f: female, m: male.

The low GBH-dose had a sex dependent effect on the bacterial alpha diversity of the mouse fecal samples as measured by the Faith's index which emphasizes phylogenetic breadth (Faith and Baker, 2006). Specifically, Faith's index was higher in male mice after 8 weeks of exposure to the low GBH-dose than in either the control or the high dose (Fig 1B, ANOVA $p<0.001$, TukeyHSD adjusted $p<0.05$). Faith's index was not significantly different between the female controls and either treatment group (Fig 1B). Modest, though significant differences between females in the low GBH-dose and the control were observed in the Shannon index which integrates the relative abundance and presence/absence of OTUs without respect to phylogenetic origin (Hill et al., 2003; Shannon, 1948) (Supporting Information Fig. S2, ANOVA $p<0.05$, TukeyHSD adjusted $p<0.05$). The traditional Shannon's calculation depends more on the richness and the less abundant species. Therefore, the Shannon's index is sensitive to small diversity changes, regardless of the relatedness of OTUs. The Faith's Phylogenetic Diversity index, on the other hand, considers phylogenetic distances between the species or in this case between the OTUs, meaning that samples with more distantly related OTUs would have a higher Faith's score, implying greater diversity (Faith and Baker, 2006). Our findings suggest that the GBH used in this study had a sex-dependent effect, and importantly that the low-GBH dose (0.07 mg glyphosate/L) affected phylogenetically diverse bacteria taxa.

The total number of unique OTUs and bacterial richness as calculated using the Chao1 estimator demonstrated that the changes observed in the Faith's index were likely due to proliferation of unique taxa (Fig 2.). The Chao1 estimation for male mice after 8 weeks of receiving low GBH-dose was significantly higher than in either the control or the high dose (Fig 2A, ANOVA $p<0.05$, TukeyHSD adjusted $p<0.001$). This was driven by a higher number of unique OTUs (Fig 2B).

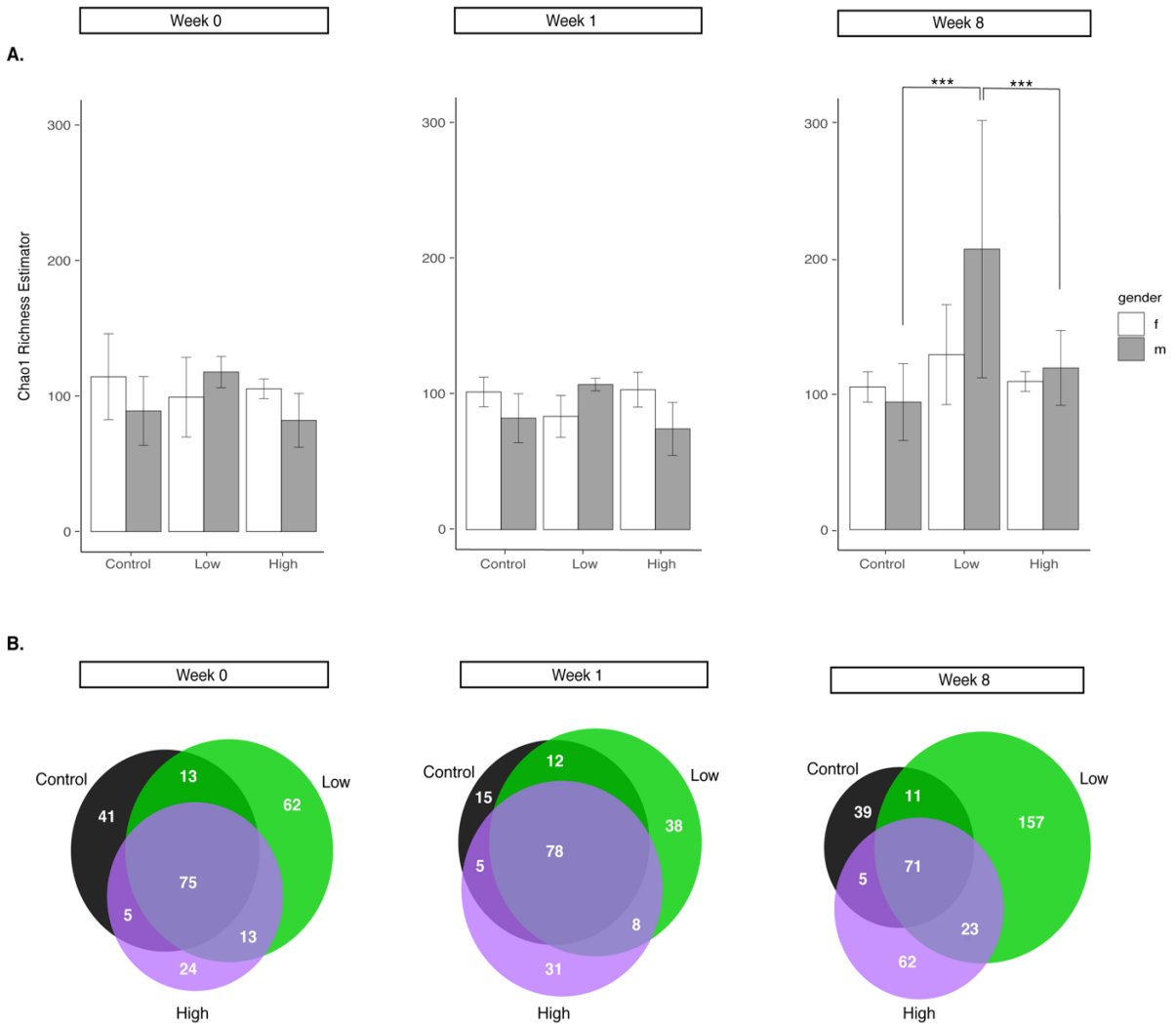


Fig 2. A. Bacterial richness measured by the Chao1 estimator increased significantly at the end of this study (week 8) on male mice receiving low dose of GBH (0.07 mg glyphosate/L) in contrast to the high dose of GBH (0.7 mg glyphosate/L) and to the control (no GBH addition) (ANOVA, $p<0.05$, TukeyHSD adjusted ***$p<0.001$). f: female, m: male. **B.** Male mice receiving low dose of GBH also showed higher number of unique OTUs at the end of this study as represented by the Venn diagrams. Numbers in each shared and unshared area indicate the number of shared and unique OTUs, respectively.

A recent study with male juvenile Sprague-Dawley rats also showed a higher number of OTUs in the colon after two weeks of exposure to a commercial glyphosate-based herbicide (25 mg/Kg of Glyfonova® 450Plus) (Nielsen et al., 2018). Interestingly, in Nielsen's *et al.* work (2018) male rats exposed to only glyphosate (2.5 mg/Kg and 25 mg/Kg) did not show differences in the number of OTUs and neither dose of glyphosate alone had an effect on bacterial diversity.

For ease of comparison, based on estimated water consumption of 6 ml/day of 0.07mg glyphosate/L (low GBH-dose), our mice would have received 21ug/Kg, or approximately 1000 times less glyphosate. Some studies have shown that the toxic effects of glyphosate-based herbicides depend not only on the glyphosate concentrations, but also on the formulation suggesting a combined effect, with some formulations being more toxic that glyphosate alone (Kurenbach *et al.*, 2015; Nicolas *et al.*, 2016; Tsui and Chu, 2003).

In the few studies about glyphosate's effect on gut bacteria, only two have reported bacterial diversity indexes and neither considered the glyphosate's effect on bacterial diversity using phylogenetically sensitive methods (Mao et al., 2018; Nielsen et al., 2018). Our study, in contrast, not only investigated GBH's effect on the phylogenetic bacterial diversity via the Faith's index, but also used the Chao1 estimator to assess bacterial richness. Thus, we observed a significant sex-dependent impact of GBH on phylogenetically distant bacteria, driven largely by proliferation at levels greater than the control.

Comparison of bacterial diversity between communities (beta diversity) as a function of sex, time, and treatment using weighted Unifrac distances showed, as expected, that the bacterial communities grouped in a sex-dependent fashion (Kim et al., 2019; Markle et al., 2013) at all times and treatments (Supporting information Fig S3A, AMOVA $p<0.05$). The pairwise similarity matrices were also consistent with our alpha diversity findings of a sex-dependent impact of GBH on bacterial diversity and richness. While the pairwise similarity matrices of female bacterial communities showed no effect of GBH exposure, males exposed to GBH doses grouped separately from each other and from the control (Fig 3 and Supporting information Fig S3B and Fig S3C, AMOVA $p<0.01$ in males).

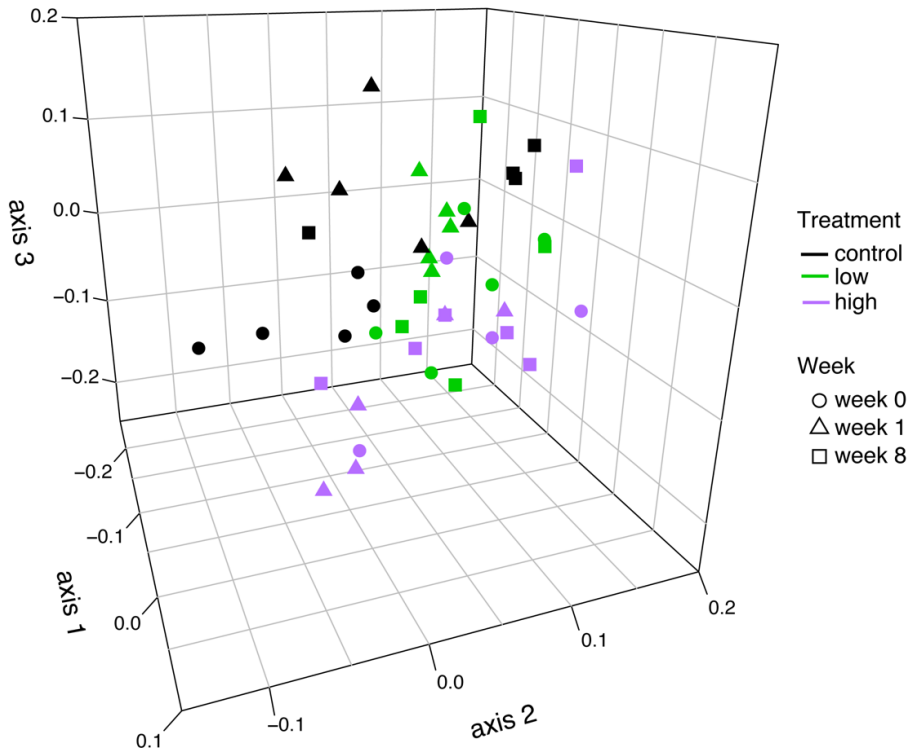


Fig 3. Weighted Unifrac distances and NMDS analysis of male mice showed that the bacterial communities grouped by treatment group (AMOVA, $p<0.01$). Control: no GBH addition, Low: 0.07 mg glyphosate /L, High: 0.7 mg glyphosate/L. Spatial distances between symbols reflect relative similarities between the bacterial communities. The threes axes represent 91% of the variance with a lowest stress of 0.130 and an R^2 value of 0.92.

A recent study with a glyphosate-based herbicide also found a sex-dependent impact on gut microbiota using Sprague-Dawley rats (Lozano et al., 2018). Their results contrasted with ours in that they only found effects on the bacterial communities of female rats long-term exposed (2 years) to a glyphosate-based herbicide. In Lozano's *et al.* study (2018), the female bacterial communities (collapsed at the phylum and family levels) grouped separately from the control, but there were no differences between the glyphosate-based herbicide doses (50 ng/L, 0.1 g/L, 2.25 g/L of estimated glyphosate). Importantly, they only compared the communities of three animals, did not report controlling for litter effects, and their low glyphosate-based herbicide doses would have been swamped by the levels of glyphosate the same group previously found in laboratory mouse chow (Mesnage et al., 2015). The contrasting results

suggest that the sex-dependent effect of glyphosate not only varies among animals, but also depends on the dose and time of glyphosate exposure.

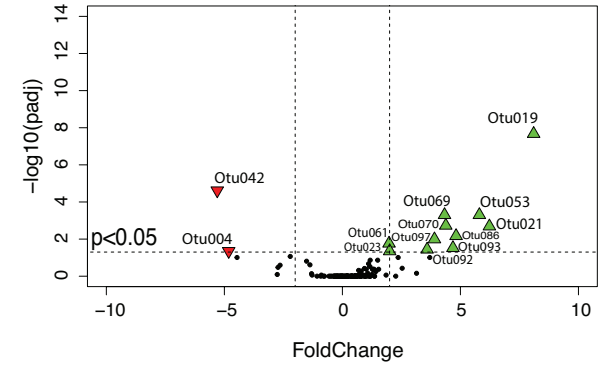
3.3.2. Differential Relative Abundance of Glyphosate-Impacted Bacteria Taxa

Since our results showed a GBH's effect only on male mice, further analyses of the bacterial abundances and changes over the time focused exclusively on males.

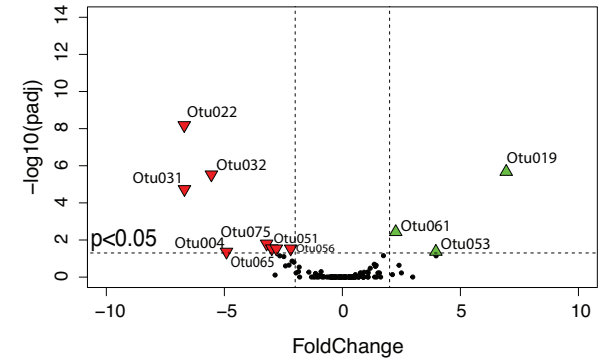
Analysis of 16S rRNA gene amplicons revealed a cumulative total of 712 OTUs in the male fecal samples. The most abundant phylum was *Firmicutes* (67.2 ± 11.1%) followed by *Verrucomicrobia*, *Bacteroidetes*, and *Proteobacteria* (16.3 ± 7.2%, 11.8 ± 4.4%, and 3.3 ± 1.4%, respectively). Of the 712, 41 OTUs had a significantly differential relative abundance in the GBH treatments when compared to the control after 8 weeks of exposure (DeSeq2, Benjamini-Hochberg correction, adjusted $p<0.05$), of these 20 OTUs changed more than 2-fold (Fig 4).

Most of the OTUs with more than 2-fold change in the high GBH-dose showed decreased relative abundances (8 out of 11), while in the low dose the majority of the differentially abundant OTUs increased in relative abundance (11 out of 13) (Fig 4A and 4B). The 8 OTUs with more than 2-fold decreased relative abundance in the high GBH-dose belonged to genera from the families: *Lachnospiraceae, Ruminococcaceae,* and *Erysipelotrichaceae* (Fig 4C). In the low GBH-dose, eight OTUs with more than 2-fold increase relative abundance belonged to the family *Lachnospiraceae*, two to the family *Ruminococcaceae,* and one to *Peptococcaceae* (Fig 4C). These findings are consistent with our bacterial diversity analyses suggesting that the two doses of GBH used in this study had different effects on mouse gut bacteria.

A.



B.



C.

Phylum	Order	Family	Genus	OTUid	FoldChange	
					Low vs. Control	High vs. Control
Firmicutes	Clostridiales	Lachnospiraceae	A2	OTU086	4.82	
			Blautia	OTU042	-5.30	
			GCA-900066575	OTU070	4.38	
			Lachnoclostridium	OTU075		-3.21
			NK4A136_group	OTU019	8.10	6.94
				OTU032		-5.55
			unclas	OTU053	5.80	3.96
				OTU065		-2.99
				OTU093	4.68	
				OTU097	3.90	
			Tyzzerella	OTU023	2.00	
			uncul	OTU061	1.98	2.26
		Peptococcaceae	uncul	OTU092	3.59	
		Ruminococcaceae	Ruminiclostridium_5	OTU022		-6.70
				OTU056		-2.19
			Ruminiclostridium_9	OTU051		-2.80
			Ruminococcaceae_ge	OTU021	6.23	
			NK4A214_group	OTU069	4.33	
			UCG-014	OTU031		-6.69
	Erysipelotrichales	Erysipelotrichaceae	Erysipelatoclostridium	OTU004	-4.82	-4.91

Fig 4. Pairwise comparisons of OTU relative abundances between the control and the GBH treatments found 20 differentially abundant OTUs with more than 2-fold change in male mice after 8 weeks (DeSeq2, Benjamini-Hochberg correction, adjusted $p<0.05$). The volcano plots show the OTU fold changes when the low and high GBH doses (0.07 mg glyphosate/L and 0.7 mg glyphosate/L, respectively) were compared against the control (**A** and **B**, respectively). OTUs above the dotted line had significant fold change. Up green triangles correspond to OTUs that increased more than 2-fold, and down red triangles to OTUs that decreased more than 2-fold when exposed to GBH. **C.** Taxonomic classification of the differentially abundant OTUs with more than 2-fold change. "unclas" and "uncul" in the genus name signify unclassified and uncultured, respectively.

Given these observations and the bacterial diversity results, we were particularly interested in examining the unique OTUs that proliferated in the low GBH-dose. As shown in Fig 2B, after 8 weeks of exposure the low GBH-dose had more unique OTUs than the control and the high dose. Of the 157 unique OTUs in the low dose, 16 showed more than 2-fold estimated change and belonged to diverse taxonomic groups. While all represented less than 0.5% of the relative abundance (Table 1), when combined, these unique OTUs represented approximately 6.5% of the total relative abundance, however, their cumulative impact on the host remains unknown.

Members of the clostridia *Lachnospiraceae* family have been reported to provide protection against colitis and show reduced abundance during inflammatory diseases in mice and humans (Frank et al., 2007; Surana and Kasper, 2017). In the bee midgut, on the other hand, *Lachnospiraceae* showed higher relative abundance after a 20 mg/L glyphosate treatment, though at lower glyphosate concentrations (0.8 mg/L and 4 mg/L) there was not a significant change (Dai et al., 2018). *Ruminococcaceae,* another clostridia family, also increased in the bee midgut after the 20 mg/L glyphosate treatment (Dai *et al.* 2018). In the rumen of cows, however, Ackerman *et al.* (2015) found that glyphosate had an inhibitory effect on the *Ruminococcaceae* genus *Ruminococcoccus*, which are known to be important for fiber degradation. Several of our *Ruminococcaceae* OTUs that decreased more than 2-fold in the high GBH-dose (0.7 mg glyphosate/L) were *Ruminiclostridium* and other poorly characterized genera (Fig 4B and 4C). These contrasting findings suggest that glyphosate's impact depends on host physiology and dose, but are not surprising given that even closely related species have been shown to be differentially susceptible to glyphosate (Ackermann et al., 2015; Aristilde et al., 2017; Shehata et al., 2013). Minimal inhibitory concentration (MIC) assays with pure glyphosate and glyphosate-

Table 1. Taxa and their associated OTUs that occurred at the end of this study (week 8) in the low-dose GBH treatment, but absent in the control or vice versa (grey highlighted). For the sake of comparison, minimal estimated fold-change was calculated based on the count number of sequences for each unique OTU divided by 1 which would represent a detection limit of 1/2,138 (rarefaction cutoff) or 0.04.

Phylum	Class	Order	Family	Genus	OTUid	Minimal estimated fold-change	Relative abundance (%)
Actinobacteria	Coriobacteriia	Coriobacteriales	Eggerthellaceae	Parvibacter	OTU098	9-fold	0.42
Bacteroidetes	Bacteroidia	Bacteroidales	Muribaculaceae	unclas	OTU188	4-fold	0.19
		Chitinophagales	Chitinophagaceae	uncul	OTU248	2-fold	0.09
		Cytophagales	Cytophagaceae	Cytophaga	OTU204	3-fold	0.14
			Microscillaceae	Chryseolinea	OTU175		0.14
Firmicutes	Bacilli	Bacillales	Listeriaceae	Listeria	OTU123	> 10-fold	0.46
			Bacillaceae	unclas	OTU125	10-fold	0.46
			Staphylococcaceae	Staphylococcus	OTU099	4-fold	0.19
			unclas	unclas	OTU118	3-fold	0.14
	Clostridia	Clostridiales	Lachnospiraceae	A2	OTU368	> 10-fold	0.46
				FCS020_group	OTU106	6-fold	0.28
				Lachnoclostridium	OTU090	> 10-fold	0.46
				NK4A136_group	OTU496	3-fold	0.14
				unclas	OTU097	> 10-fold	0.46
					OTU134		0.46
				uncul	OTU112	5-fold	0.23
			Ruminococcaceae	Oscillibacter	OTU080	7-fold	0.33
				Ruminococcaceae _ge	OTU021	> 10-fold	0.46
	Erysipelotrichia	Erysipelotrichales	Erysipelotrichaceae	Candidatus_ Stoquefichus	OTU071	> 10-fold	0.46
Proteobacteria	Alpha proteobacteria	Rhizobiales	Hyphomicrobiaceae	Pedomicrobium	OTU227	3-fold	0.14
			Xanthobacteraceae	Pseudolabrys	OTU105	4-fold	0.19
				uncul	OTU155	2-fold	0.09
Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Akkermansiaceae	Akkermansia	OTU212	3-fold	0.14

*unclas and uncul in the genus name signify unclassified and uncultured, respectively

-based herbicides have shown, for instance, that *Salmonella* serovar Typhimurium had higher MICs than the serovars Enteritidis and Infantis. Host-specific differences have also been found in the glyphosate susceptibility of *Salmonella* isolates from pigs and poultry (Poppe *et al.* 2019).

Several studies have investigated bacterial susceptibility and resistance to glyphosate using *in vitro* assays with pure cultures as well as with samples from the intestinal environment (Ackermann et al., 2015; Clair et al., 2012; Shehata et al., 2013). Those results have shown that glyphosate's impact depends not only on the bacteria taxa, but also on the formulation of the glyphosate based-herbicide (Clair et al., 2012; Nielsen et al., 2018). For example, the MICs to a glyphosate-based herbicide were higher than to pure glyphosate for commensal and pathogenic *Escherichia coli* strains, with pathogenic strains having higher MICs on average (Bote *et al.* 2019b).

We did not see any differential effect of our GBH on *Enterobacteria*. In our samples only two OTUs belonged to the *Enterobacteriaceae* family*;* one identified at the genus level as *Escherichia-Shigella* and the other unclassified. The relative abundance of the *Escherichia-Shigella* OTU did not change significantly in response to the GBH (0.7 and 0.3 -fold change in the low and high GBH treatments, respectively). Neither *E. coli* nor *S. enterica* serovar Typhimurium were affected by daily exposure to a glyphosate based-herbicide (1mg/L glyphosate) after a week of exposure in an *in vitro* ruminal simulation system (Bote *et al.* 2019a). Those *in vitro* findings are consistent with our *in vivo* observations.

Given that glyphosate inhibits the biosynthesis of aromatic amino acids, some authors have proposed to use it to control bacterial infections (Clair et al., 2012; Du et al., 2000; Kurenbach et al., 2015). Some pathogenic bacteria, however, are resistant to glyphosate (Ackermann et al., 2015; Kurenbach et al., 2015; Priestman et al., 2005; Shehata et al., 2013).

Shehata *et al.* (2013) found that *Salmonella* and *Clostridium* from poultry were more resistant to glyphosate than beneficial bacteria like *Lactobacillus* and *Bifidobacterium*. *Clostridia* also showed increased abundance after glyphosate exposure in the rumen of cows eating fodder from glyphosate treated fields (Ackermann et al., 2015).

Although we did not detect OTUs from the pathogenic bacteria reported above, two of the unique OTUs with more than 2-fold estimated change in the low GBH-dose were identified at the genus level as *Staphylococcus* and *Listeria* (Table 1), to which pathogenic species such as *Staphylococcus aureus* and *Listeria monocytogenes* belong. To our knowledge, our study is the first to report the differential effects of glyphosate on these genera. *S. aureus* is a commensal bacterium responsible for respiratory diseases and bacteremia, while *L. monocytogenes* is a well-known foodborne pathogen that causes gastroenteritis and can migrate to other organs triggering systemic infections (Aureli et al., 2000; Drevets and Bronze, 2008; Tong et al., 2015). Recent studies have demonstrated that commensal bacteria can prevent bacterial infections, providing colonization resistance against pathogenic bacteria such as *L. monocytogenes* and other opportunistic pathogens (Becattini et al., 2017; Zhang et al., 2017). Thus, our findings suggest that the low GBH-dose used in our work (0.07 mg glyphosate/L) might have disturbed the gut commensal microbiome thereby opening up an opportunity for possible proliferation of opportunistic pathogens, though no obvious health effects other than weight gain were noted for the mice receiving that dose.

The glyphosate-susceptibility of microorganisms depends on both the sensitivity of the 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) to glyphosate inhibition and the organism's ability to scavenge aromatic amino acids from their environment. For instance, *Snodgrassella alvi* of the bee gut microbiome possess a class I EPSPS, which is sensitive to glyphosate, and therefore after glyphosate exposure *S. alvi* abundance significantly decreased, suggesting that

there were insufficient aromatic amino acids in the bees' diet to satisfy *S. alvi's* growth demands for these compounds (Motta et al., 2018). It has been speculated that disturbance of core members of the gut microbiome like *S. alvi* in bees might allow the colonization of opportunistic pathogens and affect bee health (Motta et al., 2018). In contrast, it has been suggested that there is sufficient supply of aromatic amino acids in the intestinal environment of rats on conventional chow to supplement the loss of endogenous aromatic amino acids by glyphosate-sensitive bacteria which may explain the limit impact of glyphosate in the recent work of Nielsen *et al.* (2018).

We observed that male mice gained weight after 8 weeks of 0.07 mg glyphosate/L of GBH supplementation in the drinking water. Though we did not find increased abundances of bacteria associated with obesity and inflammatory diseases like the S24-7 *Bacteroidetes* family (Lozano et al. 2018), we were able to identify enriched and unique OTUs in the low GBH-dose (Table 1, Fig 2, and Fig 4) whose interactions in the gut microbiome might be affecting male mouse physiology and metabolism, though further work is required to elucidate the direct impacts of these shifts in the microbial population.

3.3.3. Glyphosate's Impact on Maturation of Male Mouse Gut Microbiome

Changes in the bacterial communities are expected as animals and the microbiome matures (Koenig et al., 2011). Three sampling times were considered in this study once the mice were exposed to GBH: end of weaning (week 0), juvenile (week 1), and adulthood (week 8). Based on the NMDS plots and pairwise similarity matrices generated from the weighted Unifrac distances, in male mice the control group showed bacterial community shifts over the time (Fig 5), having a more stable community composition as it progressed to adulthood (Koenig et al., 2011). In the

control the NMDS analysis showed that the bacterial communities of week 1 and week 8 grouped separately from week 0 (Fig 5A, AMOVA $p<0.05$). In the GBH treatments, to the contrary, there were no significant differences in the bacterial communities over time (Fig 5A). These findings suggest that GBH might be interfering with the maturation of the mouse gut microbiome.

By coupling volatility analyses with a linear random-fixed effects model, we were able to evaluate how an individual mouse's microbiome changed over time and corrected for the intrinsic variability among male mice at each sampling time. By doing so, we found that in the control group the weighted Unifrac distances were significantly higher in week 1 than in week 0 (Fig 5B, ANOVA, $p<0.05$, TukeyHSD adjusted $p< 0.01$). This finding indicates that the dissimilarity between bacterial communities in the same animal increased in response to maturation from infant to juvenile. A similar response was observed on the high GBH-dose, where the weighted Unifrac distances of week 1 were higher than in week 0. Between week 1 and week 8, although there was no difference in the control, there was a statistically significant difference in the bacterial communities of individual mice on the high dose (Fig 5B, ANOVA, $p<0.05$, TukeyHSD adjusted $p< 0.05$). Of note, there were no differences of the weighted Unifrac distances over time in the low-GBH dose (0.07 mg glyphosate/L), consistent with our hypothesis that GBH exposure affected maturation of the mouse microbiome, though there appeared to be dose-dependent differences in the impact of GBH.

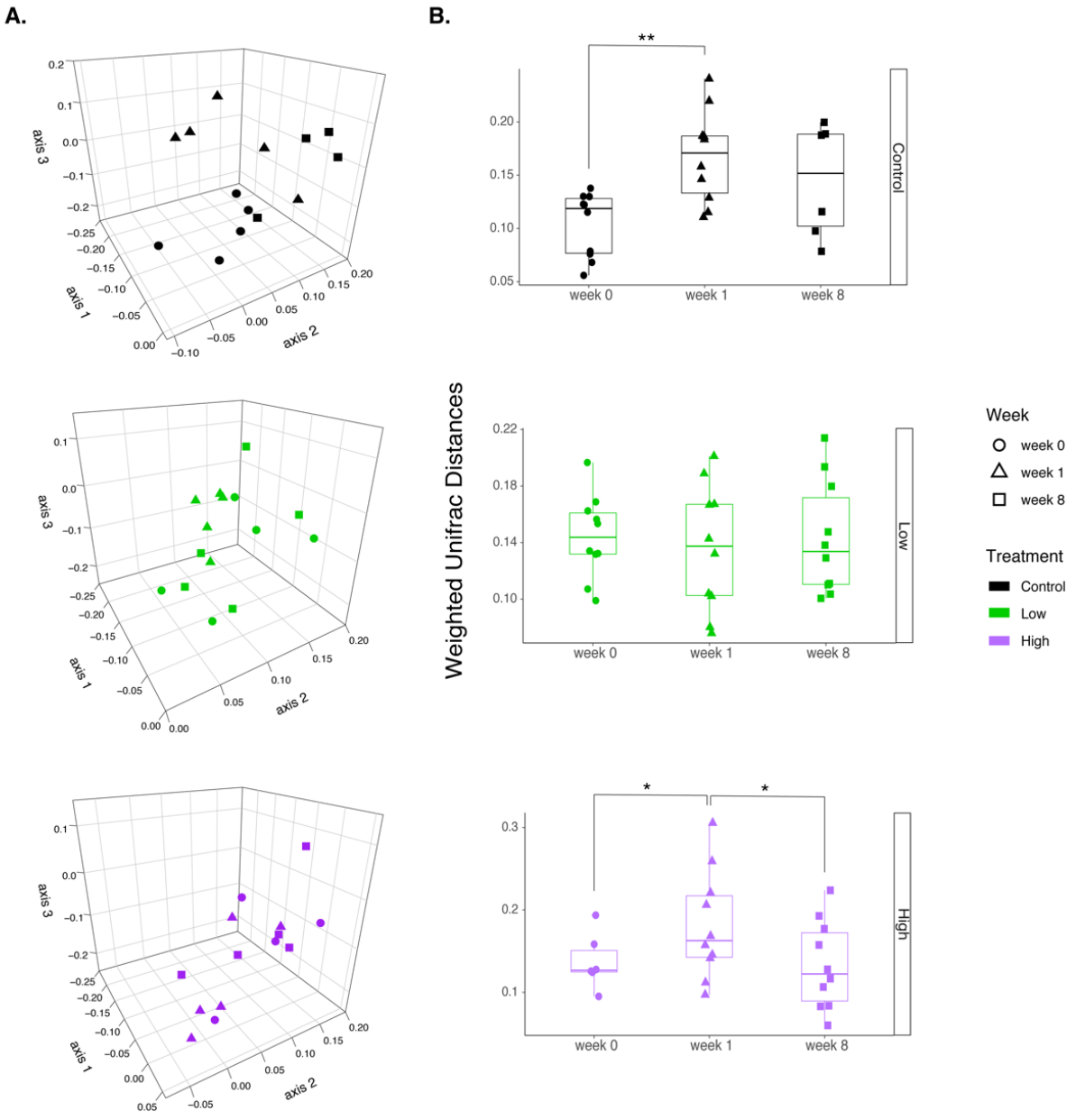


Fig 4. A. 3D-NMDS plots of the weighted Unifrac distances from male mice showed that only the infant mice (week 0) from the control grouped separately from the rest of the sampling times (week 1: juvenile, week 8: adult) (AMOVA $p<0.05$). In the GBH treatments, there were no differential groupings of the bacterial communities with time. Relative spatial distances between the symbols represent bacterial community similarities. The three axes show 91% of the variance with a lowest stress of 0.130 and R^2 value of 0.92. **B.** Temporal volatility analysis using weighted Unifrac distances found significant differences between infant and juvenile mice (week 0 and week 1, respectively) in the control group and in the high dose-GBH treatment (ANOVA, $p<0.05$, TukeyHSD adjusted *$p<0.05$, **$p<0.01$), but there were no differences in the low dose-GBH treatment.

Disruptions of the gut microbiome during the colonization process have proven to be detrimental to the host, affecting metabolism and increasing the risk to diseases associated to immunologic disorders (Koenig et al., 2011; Penders et al., 2007; Stokholm et al., 2018). Sprague Dawley rats exposed to low doses of glyphosate in utero (1.75 mg/Kg/day) showed changes in the gut microbiome composition at a very young age. One-month old rat pups exhibited higher abundance of the genus *Prevotella* and decreased abundance of the beneficial bacterium *Lactobacillus* (Mao et al., 2018). Although the long-term effects of glyphosate have not been fully investigated, bacterial imbalances in the gut microbiome caused by glyphosate have been hypothesized to impact nutritional deficiencies and intestinal diseases like celiac disease (Samsel and Seneff, 2013a, 2013b) though there is no direct experiment evidence to support this claim to our knowledge. Our temporal volatility analyses do, however, demonstrate that the low-GBH dose used in this study (0.07 mg glyphosate/L) altered maturation of the mouse microbiome, though the long-term consequences of such changes remain unknown.

3.4. CONCLUSIONS

There are growing concerns about glyphosate's potential carcinogenic effect in humans, and the U.S. Environmental Protection Agency has published a drinking water Maximum Contaminate Level (MCL) for glyphosate of 0.7 mg/L (US EPA, 2015). In our study, mice exposed to the low GBH-dose (0.07 mg glyphosate/L) would have received 21 ug/Kg of glyphosate (based on 6 ml/day of water consumption). A 70 Kg human consuming 2 L/day of water with glyphosate at our low-GBH dose would have been exposed to 10X less glyphosate (2 ug/Kg) than in mice; however, with a human drinking water at the MCL would have received approximately the same glyphosate dose (20 ug/Kg) as the mice in our low dose treatment.

Our findings, in addition to revealing a sex-dependent effect of glyphosate on mouse gut bacteria, showed that the low GBH-dose used in this study (21 ug/Kg of glyphosate) impacted the bacterial diversity and richness measured by the Faith's phylogenetic diversity index and Chao1 estimator. In contrast to the control group (no GBH addition), the low GBH-dose did not show normal changes in the bacterial communities over time suggesting interference with the expected maturation of the gut microbiome as mice transition to adulthood. Specifically, after 8 weeks, bacteria taxa mostly belonging to *Lachnospiraceae* and *Ruminococcaceae* families were enriched by exposure to the low GBH-dose (unique OTUs), and most of the differentially abundant OTUs increased in this GBH treatment when compared to the control. Since glyphosate's effect is species and even strain-specific, several members of these families should be included in future glyphosate-susceptibility assays, especially considering that these families harbor commensal bacteria whose imbalances might trigger intestinal illnesses (Koenig et al., 2011; Penders et al., 2007).

On aggregate, the evidence presented here demonstrates that low dose exposures to GBH had a previously unappreciated impact on microbiome maturation. This work suggests that additional studies with larger sample sizes and both glyphosate and GBH doses around or below the MCL should be performed. Future work should also consider using fecal transplants to unexposed germ-free mice to assess if the changes we measured in the gut microbiome were responsible for the observed weight gain in the male mice or whether they were a consequence of the weight gain.

3.5. REFERENCES

Ackermann, W., Coenen, M., Schrödl, W., Shehata, A.A., Krüger, M., 2015. The Influence of Glyphosate on the Microbiota and Production of Botulinum Neurotoxin During Ruminal Fermentation. Curr. Microbiol. 70, 374–382. https://doi.org/10.1007/s00284-014-0732-3

Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome Biol. 11, R106. https://doi.org/10.1186/gb-2010-11-10-r106

Aristilde, L., Reed, M.L., Wilkes, R.A., Youngster, T., Kukurugya, M.A., Katz, V., Sasaki, C.R.S., 2017. Glyphosate-Induced Specific and Widespread Perturbations in the Metabolome of Soil Pseudomonas Species. Front. Environ. Sci. 5. https://doi.org/10.3389/fenvs.2017.00034

Aureli, P., Fiorucci, G.C., Caroli, D., Marchiaro, G., Novara, O., Leone, L., Salmaso, S., 2000. An outbreak of febrile gastroenteritis associated with corn contaminated by Listeria monocytogenes. N. Engl. J. Med. 342, 1236–1241. https://doi.org/10.1056/NEJM200004273421702

Bai, S.H., Ogbourne, S.M., 2016. Glyphosate: environmental contamination, toxicity and potential risks to human health via food contamination. Environ. Sci. Pollut. Res. 23, 18988–19001. https://doi.org/10.1007/s11356-016-7425-3

Balbuena, M.S., Tison, L., Hahn, M.-. L., Greggers, U., Menzel, R., Farina, W.M., 2015. Effects of sublethal doses of glyphosate on honeybee navigation. J Exp Biol.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01

Battaglin, W.A., Meyer, M.T., Kuivila, K.M., Dietze, J.E., 2014. Glyphosate and its degradation product AMPA occur frequently and widely in U.S. soils, surface water, groundwater, and precipitation. J Am Water Resour Assoc 50. https://doi.org/10.1111/jawr.12159

Becattini, S., Littmann, E.R., Carter, R.A., Kim, S.G., Morjaria, S.M., Ling, L., Gyaltshen, Y., Fontana, E., Taur, Y., Leiner, I.M., Pamer, E.G., 2017. Commensal microbes provide first line defense against Listeria monocytogenes infection. J. Exp. Med. 214, 1973–1989. https://doi.org/10.1084/jem.20170495

Benbrook, C.M., 2016. Trends in glyphosate herbicide use in the united states and globally: Supporting data. Env. Sci Eur.

Bioconductor, 2003. https://www.bioconductor.org/ (accessed 11 May 2019).

Björkholm, B., Bok, C.M., Lundin, A., Rafter, J., Hibberd, M.L., Pettersson, S., 2009. Intestinal Microbiota Regulate Xenobiotic Metabolism in the Liver. PLOS ONE 4, e6958. https://doi.org/10.1371/journal.pone.0006958

Bote, K., Pöppe, J., Merle, R., Makarova, O., Roesler, U., 2019a. Minimum Inhibitory Concentration of Glyphosate and of Glyphosate-Containing Herbicide Formulation for *Escherichia coli* Isolates - Differences Between Pathogenic and Non-pathogenic Isolates and Between Host Species. Front Microbiol. 10, 932. https://doi.org/10.3389/fmicb.2019.00932

Bote, K., Pöppe, J., Riede, S., Breves, G., Roesler, U., 2019b. Effect of a Glyphosate-Containing Herbicide on *Escherichia coli* and *Salmonella* Ser. Typhimurium in an In Vitro Rumen Simulation System. Eur Jour Microbiol Immunol. https://doi.org/10.1556/1886.2019.00010

Clair, E., Linn, L., Travert, C., Amiel, C., Séralini, G.-E., Panoff, J.-M., 2012. Effects of Roundup(®) and glyphosate on three food microorganisms: Geotrichum candidum, Lactococcus lactis subsp. cremoris and Lactobacillus delbrueckii subsp. bulgaricus. Curr. Microbiol. 64, 486–491. https://doi.org/10.1007/s00284-012-0098-3

Claus, S.P., Guillou, H., Ellero-Simatos, S., 2016. The gut microbiota: a major player in the toxicity of environmental pollutants? Npj Biofilms Microbiomes 2, 16003. https://doi.org/10.1038/npjbiofilms.2016.3
Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., Tiedje, J.M., 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42, D633–D642. https://doi.org/10.1093/nar/gkt1244
Dai, P., Yan, Z., Ma, S., Yang, Y., Wang, Q., Hou, C., Wu, Y., Liu, Y., Diao, Q., 2018. The Herbicide Glyphosate Negatively Affects Midgut Bacterial Communities and Survival of Honey Bee during Larvae Reared in Vitro. J. Agric. Food Chem. 66, 7786–7793. https://doi.org/10.1021/acs.jafc.8b02212
DiGiulio, D.B., Callahan, B.J., McMurdie, P.J., Costello, E.K., Lyell, D.J., Robaczewska, A., Sun, C.L., Goltsman, D.S.A., Wong, R.J., Shaw, G., Stevenson, D.K., Holmes, S.P., Relman, D.A., 2015. Temporal and spatial variation of the human microbiota during pregnancy. Proc. Natl. Acad. Sci. U. S. A. 112, 11060–11065. https://doi.org/10.1073/pnas.1502875112
Drevets, D.A., Bronze, M.S., 2008. Listeria monocytogenes : epidemiology, human disease, and mechanisms of brain invasion. FEMS Immunol. Med. Microbiol. 53, 151–165. https://doi.org/10.1111/j.1574-695X.2008.00404.x
Du, W., Wallis, N.G., Mazzulla, M.J., Chalker, A.F., Zhang, L., Liu, W.-S., Kallender, H., Payne, D.J., 2000. Characterization of Streptococcus pneumoniae 5-enolpyruvylshikimate 3-phosphate synthase and its activation by univalent cations. Eur. J. Biochem. 267, 222–227. https://doi.org/10.1046/j.1432-1327.2000.00994.x
EFSA, 2015. Conclusion on the peer review of the pesticide risk assessment of the active substance glyphosate. EFSA J. 13, 4302. https://doi.org/10.2903/j.efsa.2015.4302
Faith, D.P., Baker, A.M., 2006. Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. Evol. Bioinforma. Online 2, 121.
Fernandez-Cornejo, J. and McBride, W., 2002. Adoption of Bioengineered Crops. USDA Economic Research Service Agricultural Economic Report No. (AER-810) USDA Agricultural Economic Report No. (AER-810) 67 pp
Frank, D.N., St. Amand, A.L., Feldman, R.A., Boedeker, E.C., Harpaz, N., Pace, N.R., 2007. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc. Natl. Acad. Sci. U. S. A. 104, 13780–13785. https://doi.org/10.1073/pnas.0706625104
Funke, T., Han, H., Healy-Fried, M.L., Fischer, M., Schönbrunn, E., 2006. Molecular basis for the herbicide resistance of Roundup Ready crops. Proc. Natl. Acad. Sci. U. S. A. 103, 13010–13015. https://doi.org/10.1073/pnas.0603638103
Gao, Z., Guo, B., Gao, R., Zhu, Q., Qin, H., 2015. Microbiota disbiosis is associated with colorectal cancer. Name Front. Microbiol. 6, 20.
Gaupp-Berghausen, M., Hofer, M., Rewald, B., Zaller, J.G., 2015. Glyphosate-based herbicides reduce the activity and reproduction of earthworms and lead to increased soil nutrient concentrations. Sci Rep 5. https://doi.org/10.1038/srep12886
Gianessi, L.P., 2013. The increasing importance of herbicides in worldwide crop production. Pest Manag. Sci. 69, 1099–1105. https://doi.org/10.1002/ps.3598
Halfvarson, J., Brislawn, C.J., Lamendella, R., Vázquez-Baeza, Y., Walters, W.A., Bramer, L.M., D'Amato, M., Bonfiglio, F., McDonald, D., Gonzalez, A., McClure, E.E., Dunklebarger, M.F., Knight, R., Jansson, J.K., 2017. Dynamics of the human gut

microbiome in inflammatory bowel disease. Nat. Microbiol. 2, 17004. https://doi.org/10.1038/nmicrobiol.2017.4
Hill, T.C.J., Walsh, K.A., Harris, J.A., Moffett, B.F., 2003. Using ecological diversity measures with bacterial communities. FEMS Microbiol. Ecol. 43, 1–11. https://doi.org/10.1111/j.1574-6941.2003.tb01040.x
IARC, 2015. IARC Monographs Volume 112: evaluation of five organophosphate insecticides and herbicides. https://www.iarc.fr/news-events/iarc-monographs-volume-112-evaluation-of-five-organophosphate-insecticides-and-herbicides/
Kho, Z.Y., Lal, S.K., 2018. The Human Gut Microbiome – A Potential Controller of Wellness and Disease. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.01835
Kim, Y.S., Unno, T., Kim, B.Y., Park, M.S., 2019. Sex Differences in Gut Microbiota. World J. Mens Health. https://doi.org/10.5534/wjmh.190009
Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J., Knight, R., Angenent, L.T., Ley, R.E., 2011. Succession of microbial consortia in the developing infant gut microbiome. Proc. Natl. Acad. Sci. 108, 4578–4585. https://doi.org/10.1073/pnas.1000081107
Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. Appl. Environ. Microbiol. 79, 5112–5120. https://doi.org/10.1128/AEM.01043-13
Kurenbach, B., Marjoshi, D., Amábile-Cuevas, C.F., Ferguson, G.C., Godsoe, W., Gibson, P., Heinemann, J.A., 2015. Sublethal Exposure to Commercial Formulations of the Herbicides Dicamba, 2,4-Dichlorophenoxyacetic Acid, and Glyphosate Cause Changes in Antibiotic Susceptibility in Escherichia coli and Salmonella enterica serovar Typhimurium. mBio 6, e00009-15. https://doi.org/10.1128/mBio.00009-15
Lozano, V.L., Defarge, N., Rocque, L.-M., Mesnage, R., Hennequin, D., Cassier, R., de Vendômois, J.S., Panoff, J.-M., Séralini, G.-E., Amiel, C., 2018. Sex-dependent impact of Roundup on the rat gut microbiome. Toxicol. Rep. 5, 96–107. https://doi.org/10.1016/j.toxrep.2017.12.005
Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., Knight, R., 2011. UniFrac: an effective distance metric for microbial community comparison. ISME J. 5, 169–172. https://doi.org/10.1038/ismej.2010.133
Mao, Q., Manservisi, F., Panzacchi, S., Mandrioli, D., Menghetti, I., Vornoli, A., Bua, L., Falcioni, L., Lesseur, C., Chen, J., Belpoggi, F., Hu, J., 2018. The Ramazzini Institute 13-week pilot study on glyphosate and Roundup administered at human-equivalent dose to Sprague Dawley rats: effects on the microbiome. Environ. Health 17. https://doi.org/10.1186/s12940-018-0394-x
Markle, J.G.M., Frank, D.N., Mortin-Toth, S., Robertson, C.E., Feazel, L.M., Rolle-Kampczyk, U., Bergen, M. von, McCoy, K.D., Macpherson, A.J., Danska, J.S., 2013. Sex Differences in the Gut Microbiome Drive Hormone-Dependent Regulation of Autoimmunity. Science 339, 1084–1088. https://doi.org/10.1126/science.1233521
Mesnage, R., Defarge, N., Rocque, L.-M., Vendômois, J.S. de, Séralini, G.-E., 2015. Laboratory Rodent Diets Contain Toxic Levels of Environmental Contaminants: Implications for Regulatory Tests. PLOS ONE 10, e0128429. https://doi.org/10.1371/journal.pone.0128429
Mir, R., Jallu, S., Singh, T.P., 2015. The shikimate pathway: review of amino acid sequence, function and three-dimensional structures of the enzymes. Crit. Rev. Microbiol. 41, 172–189. https://doi.org/10.3109/1040841X.2013.813901

Mothur, 2013. https://www.mothur.org/wiki/MiSeq_SOP (accessed 5 October 2018).
Motta, E.V.S., Raymann, K., Moran, N.A., 2018. Glyphosate perturbs the gut microbiota of honey bees. Proc. Natl. Acad. Sci. 115, 10305–10310. https://doi.org/10.1073/pnas.1803880115
Myers, J.P., Antoniou, M.N., Blumberg, B., Carroll, L., Colborn, T., Everett, L.G., Hansen, M., Landrigan, P.J., Lanphear, B.P., Mesnage, R., Vandenberg, L.N., vom Saal, F.S., Welshons, W.V., Benbrook, C.M., 2016. Concerns over use of glyphosate-based herbicides and risks associated with exposures: a consensus statement. Environ. Health 15, 19. https://doi.org/10.1186/s12940-016-0117-0
Nicolas, V., Oestreicher, N., Vélot, C., 2016. Multiple effects of a commercial Roundup® formulation on the soil filamentous fungus Aspergillus nidulans at low doses: evidence of an unexpected impact on energetic metabolism. Environ. Sci. Pollut. Res. Int. 23, 14393–14404. https://doi.org/10.1007/s11356-016-6596-2
Nielsen, L.N., Roager, H.M., Casas, M.E., Frandsen, H.L., Gosewinkel, U., Bester, K., Licht, T.R., Hendriksen, N.B., Bahl, M.I., 2018. Glyphosate has limited short-term effects on commensal bacterial community composition in the gut environment due to sufficient aromatic amino acid levels. Environ. Pollut. 233, 364–376. https://doi.org/10.1016/j.envpol.2017.10.016
Penders, J., Thijs, C., van den Brandt, P.A., Kummeling, I., Snijders, B., Stelma, F., Adams, H., van Ree, R., Stobberingh, E.E., 2007. Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study. Gut 56, 661–667. https://doi.org/10.1136/gut.2006.100164
Pöppe, J., Bote, K., Merle, R., Makarova, O., Roesler, U., 2019. Minimum Inhibitory Concentration of Glyphosate and a Glyphosate-Containing Herbicide in *Salmonella enterica* Isolates Originating from Different Time Periods, Hosts, and Serovars. Eur Jour Microbiol Immunol. 2, 35-41. https://doi.org/10.1556/1886.2019.00005
Priestman, M.A., Funke, T., Singh, I.M., Crupper, S.S., Schönbrunn, E., 2005. 5-Enolpyruvylshikimate-3-phosphate synthase from Staphylococcus aureus is insensitive to glyphosate. FEBS Lett. 579, 728–732. https://doi.org/10.1016/j.febslet.2004.12.057
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/nar/gks1219
Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4. https://doi.org/10.7717/peerj.2584
Samsel, A., Seneff, S., 2013a. Glyphosate, pathways to modern diseases II: Celiac sprue and gluten intolerance. Interdiscip. Toxicol. 6, 159–184. https://doi.org/10.2478/intox-2013-0026
Samsel, A., Seneff, S., 2013b. Glyphosate's Suppression of Cytochrome P450 Enzymes and Amino Acid Biosynthesis by the Gut Microbiome: Pathways to Modern Diseases. Entropy 15, 1416–1463. https://doi.org/10.3390/e15041416
Shannon, C.E., 1948. A Mathematical Theory of Communication. Bell Syst. Tech. J. 27, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
Shehata, A.A., Schrödl, W., Aldin, Alaa.A., Hafez, H.M., Krüger, M., 2013. The Effect of Glyphosate on Potential Pathogens and Beneficial Members of Poultry Microbiota In Vitro. Curr. Microbiol. 66, 350–358. https://doi.org/10.1007/s00284-012-0277-2
Stokholm, J., Blaser, M.J., Thorsen, J., Rasmussen, M.A., Waage, J., Vinding, R.K., Schoos, A.-M.M., Kunøe, A., Fink, N.R., Chawes, B.L., Bønnelykke, K., Brejnrod, A.D., Mortensen,

M.S., Al-Soud, W.A., Sørensen, S.J., Bisgaard, H., 2018. Maturation of the gut microbiome and risk of asthma in childhood. Nat. Commun. 9, 141. https://doi.org/10.1038/s41467-017-02573-2

Surana, N.K., Kasper, D.L., 2017. Moving beyond microbiome-wide associations to causal microbe identification. Nature 552, 244–247. https://doi.org/10.1038/nature25019

Swanson, N.L., Leu, A., Abrahamson, J., Wallet, B., 2014. Genetically engineered crops, glyphosate and the deterioration of health in the United States of America 32.

Tarazona, J.V., Court-Marques, D., Tiramani, M., Reich, H., Pfeil, R., Istace, F., Crivellente, F., 2017. Glyphosate toxicity and carcinogenicity: a review of the scientific basis of the European Union assessment and its differences with IARC. Arch. Toxicol. 91, 2723–2743. https://doi.org/10.1007/s00204-017-1962-5

Tong, S.Y.C., Davis, J.S., Eichenberger, E., Holland, T.L., Fowler, V.G. Jr., 2015. *Staphylococcus aureus* Infections: Epidemiology, Pathophysiology, Clinical Manifestations, and Management. Clin Microbiol Rev. 28, 603-661. https://doi.org/10.1128/CMR.00134-14

Tsui, M.T.K., Chu, L.M., 2003. Aquatic toxicity of glyphosate-based formulations: comparison between different organisms and the effects of environmental factors. Chemosphere 52, 1189–1197. https://doi.org/10.1016/S0045-6535(03)00306-0

USDA:NASS, 2013. Acreage planted and/or harvested. USDA Economics, Statistics and Market Information System. Albert R. Mann Library. Cornell University. http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1000 (accessed 3 June 2019)

US EPA, O., 2015. National Primary Drinking Water Regulations. https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations (accessed 13 July 2019).

Van Bruggen, A.H.C., He, M.M., Shin, K., Mai, V., Jeong, K.C., Finckh, M.R., Morris, J.G., 2018. Environmental and health effects of the herbicide glyphosate. Sci. Total Environ. 616–617, 255–268. https://doi.org/10.1016/j.scitotenv.2017.10.309

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl. Environ. Microbiol. 73, 5261–5267. https://doi.org/10.1128/AEM.00062-07

Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., Glöckner, F.O., 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. 42, D643–D648. https://doi.org/10.1093/nar/gkt1209

Yu, Z., Forster, R.J., 2005. Nucleic acid extraction, oligonucleotide probes and PCR methods, in: Makkar, H.P.S., McSweeney, C.S. (Eds.), Methods in Gut Microbial Ecology for Ruminants. Springer Netherlands, Dordrecht, pp. 81–104. https://doi.org/10.1007/1-4020-3791-0_7

Zhang, T., Abel, S., Wiesch, P.A. zur, Sasabe, J., Davis, B.M., Higgins, D.E., Waldor, M.K., 2017. Deciphering the landscape of host barriers to Listeria monocytogenes infection. Proc. Natl. Acad. Sci. U. S. A. 114, 6334–6339. https://doi.org/10.1073/pnas.1702077114

Zhi, X.-Y., Yao, J.-C., Li, H.-W., Huang, Y., Li, W.-J., 2014. Genome-wide identification, domain architectures and phylogenetic analysis provide new insights into the early evolution of shikimate pathway in prokaryotes. Mol. Phylogenet. Evol. 75, 154–164. https://doi.org/10.1016/j.ympev.2014.02.015

SUPPLEMENTARY MATERIALS

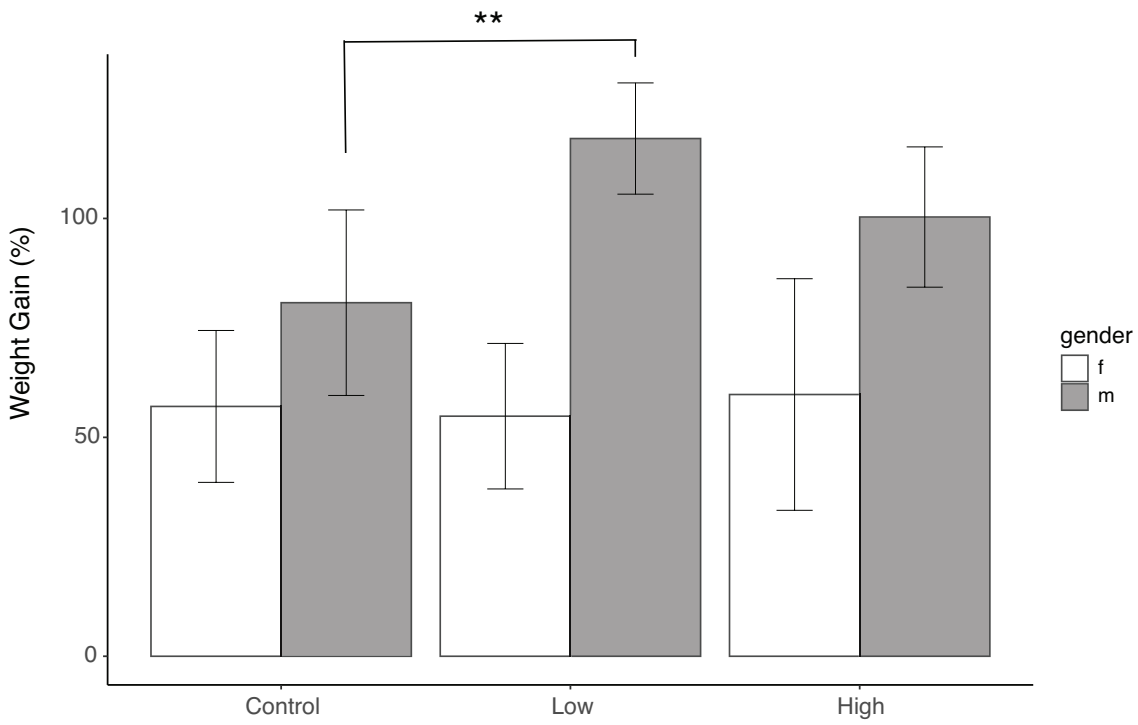


Fig S1. Total body weight gain (%) after 8 weeks of GBH exposure showed that male mice receiving low dose of GBH (0.07 mg glyphosate/L) gained more total weight than the control (no GBH) (ANOVA $p<0.01$, TukeyHSD adjusted $^{**}p<0.01$). Female mice did not show differences between the treatment groups (ANOVA $p>0.05$). f: female, m: male.

1
2

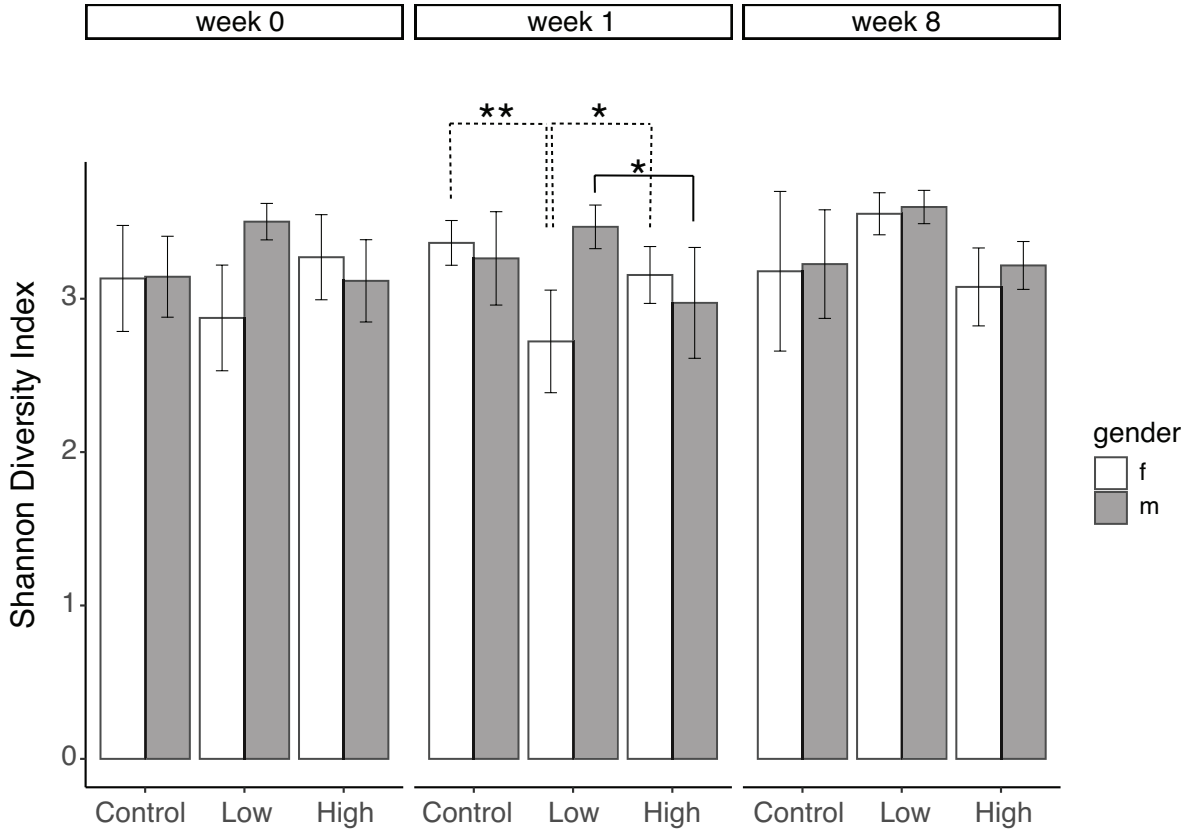


Fig S2. Bacterial alpha diversity measured by the Shannon index did not show an evident effect of GBH on mice. In females, however, there was a modest effect after one week of exposure (week 1) to the low dose of GBH (ANOVA $p<0.05$, TukeyHSD adjusted *$p<0.05$, **$p<0.01$). Control: no GBH addition, Low: 0.07 mg glyphosate/L, High: 0.7 mg glyphosate/L. f: female, m: male. Dotted and continuous lines indicate significant pairwise comparisons among female and male mice, respectively.

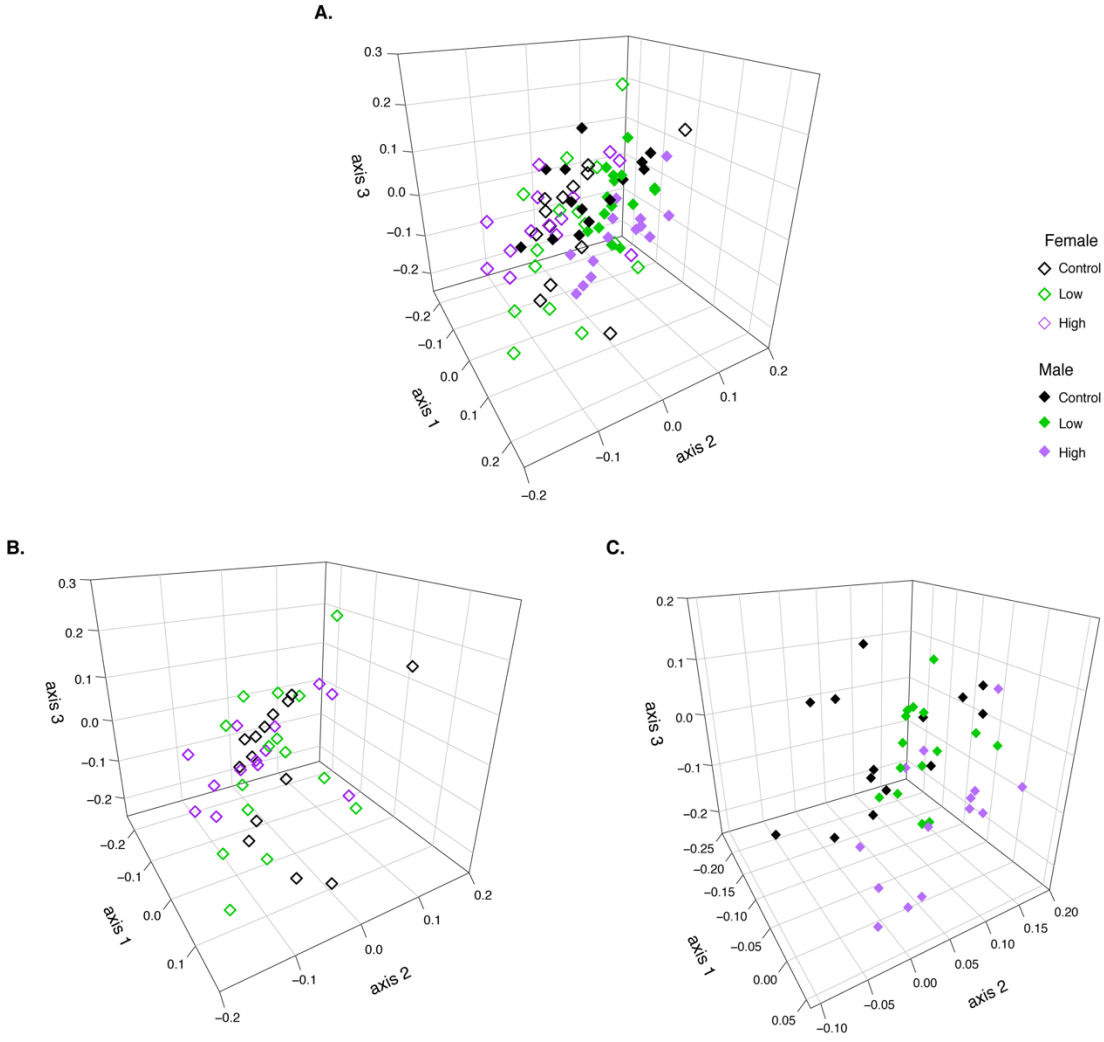


Fig S3. Weighted Unifrac distances and NMDS analysis showed that the bacterial communities grouped significantly by gender **(A) (**AMOVA, $p<0.05$). While there was no impact of GBH addition in the bacterial communities of female mice **(B)** (AMOVA, $p>0.05$); in males the bacterial communities grouped by treatment group **(C)** (AMOVA, $p<0.01$). Control: no GBH addition, Low: 0.07 mg glyphosate /L, High: 0.7 mg glyphosate/L. Spatial distances between symbols reflect relative similarities between the bacterial communities. The threes axes represent 91% of the variance with a lowest stress of 0.130 and an R^2 value of 0.92.

CHAPTER 4

CHARACTERIZATION OF A FLURBIPROFEN-DEGRADING BACTERIAL CONSORTIUM

Vienvilay Phandanouvong-Lozano[a], Mingjun Liao[a,1], Anthony G. Hay[a]

[a]Department of Microbiology, Cornell University, Ithaca, NY, 14853, USA

[1] Present address: College of Resources and Environmental Engineering, Hubei University of Technology, Wuhan, 430068, China

ABSTRACT

Flurbiprofen is a nonsteroidal, anti-inflammatory drug (NSAID). No bacterial enrichments or pure cultures have been previously reported to grow on flurbiprofen, but in an initial enrichment from sewage sludge, named Flur-4, we detected a putative ring-cleavage metabolite as well as dihydroxyflurbiprofen (DHF) and 3-fluoro-4-(1-carboxyethyl)benzoic (FCB), consistent with a biphenyl-like degradation pathway. We initially isolated a *Phenylobacterium* strain that appear pure when grown on flurbiprofen as a sole source of carbon and energy. Genomic analysis of this strain, however, revealed a mixed culture that lacked an obvious biphenyl-like-degradation operon. The strain itself was unstable and failed to survive long-term storage at -80°C. The frozen stock, however, did yield an *Ochrobactrum* species we named BC that was able to grow on flurbiprofen and produced a ring-cleavage metabolite from it. BC readily gave rise to small colony variants we called PZ that could not produce the ring-cleavage metabolite. Although the metagenome assembled genomes (MAGs) of both BC and PZ encoded genes for degradation of

aromatic compounds, neither encoded genes for an intact biphenyl-like pathway. Both, however, encoded a novel putative aromatic ring-hydroxylating dioxygenase (OfdA: putative flurbiprofen dioxygenase α-subunit from *Ochrobactrum*). PZ, however, lacked a *benR*-like transcriptional regulator that may have been required for *ofdA* expression, which may explain why this variant could no longer grow on flurbiprofen. The fact that *benR* was not contiguous with *ofdA* may also explain why no BC fosmid library clones degraded flurbiprofen. Future work should focus on putative function of OfdA in flurbiprofen metabolism and the potential role BenR in expression of *ofdA.* MG-RAST analysis of all the sequenced reads suggested contamination from other strains and the presence of a plasmid not included in the MAG assemblies which may also play an as yet, undetermined role in flurbiprofen degradation. Though our work has provided some insights about a novel flurbiprofen-degrader, the genomic characterization has proven to be challenging. Future work should also include reconstruction of plasmids from metagenome data, and reexamination of the default annotation parameters that may have mislabeled genes encoding potential flurbiprofen-degrading enzymes.

Keywords: nonsteroidal anti-inflammatory drug, NSAID, biodegradation, aromatic ring-hydroxylating dioxygenase, putative flurbiprofen dioxygenase, metagenome, *Ochrobactrum*

4.1. INTRODUCTION

Flurbiprofen (2-(3-fluoro-4-phenyl-phenyl)propanoic acid) is a nonsteroidal anti-inflammatory drug (NSAID) routinely prescribed to reduce pain and inflammation associated to osteoarthritis and rheumatoid arthritis (Davies, 1995; Richy et al., 2007). NSAIDs are a class of aromatic acidic drugs widely used because of their analgesic, antipyretic and anti-inflammatory properties.

They accomplish this by inhibiting prostaglandin synthesis by cyclooxygenase-1 (COX-1) and cycloxygenase-2 (COX-2) isoenzymes (Gentili, 2007; Gierse et al., 1999; Ricciotti and FitzGerald, 2011). Globally, NSAIDs are one of the most-consumed classes of pharmaceuticals and have limited removal during wastewater treatment (Camacho-Muñoz et al., 2012; Cycoń et al., 2016). While flurbiprofen is currently a prescription-only medication, recent reports suggest that flurbiprofen holds promise as cancer therapeutic, and even for anti-obesity purposes (Duncan et al., 2012; Hosoi et al., 2016; Wynne and Djakiew, 2010). If such applications prove to be highly effective, it is likely that flurbiprofen's popularity will increase in coming decades which will lead to great environmental release.

Environmental relevant concentrations of other NSAIDs like flurbiprofen such as diclofenac and ibuprofen have demonstrated to have toxic effects on several aquatic organisms including mussels and fish (Bickley et al., 2017; Gonzalez-Rey and Bebianno, 2014; Mezzelani et al., 2018; Parolini et al., 2011). Diclofenac and ibuprofen have also shown to disturb soil metabolic activities and associated microbial communities (Cycoń et al., 2016). Based on the compiled evidence from ecotoxicity studies, the Global Water Research Coalition (GWRC) which gathers research organizations from around the world including the USA and several European countries, listed diclofenac and ibuprofen in the top 10 priority water contaminants (de Voogt et al., 2009; Richardson and Ternes, 2011).

More than 95% of orally consumed flurbiprofen is excreted through the kidney within 24h, with 65% and 85% are glucuronide and sulfate flurbiprofen conjugates (Abdel-Aziz et al., 2012). From the very few studies that have examined flurbiprofen environmental concentrations, flurbiprofen was detected in wastewater treatment plant (WWTP) effluents from France and Italy (0.21 and 0.34 μg/L flurbiprofen, respectively); but not in Swedish WWTP effluents (Andreozzi et al., 2003; Bendz et al., 2005). Based on flurbiprofen's poor water solubility (K_{ow} of 4.2), it is

likely that flurbiprofen binds to organic matter (Abdel-Aziz et al., 2012). The freely available web-based software, PBT profiler (Persistence, Bioaccumulation, and Toxicity assessments) developed for the US Environmental Protection Agency (US EPA), predicts flurbiprofen to be moderately toxic to fish and to be persistent in sediment (US EPA, 2015).

Although the fungus *Cunninghamella* and the bacterium *Streptomyces* are capable of flurbiprofen biotransformation, the resulting metabolites are suggestive of cytochromes P450 involvement (Amadio et al., 2010; Domaradzka et al., 2015). In fact, *Cunninghamella elegans* DSM 1908 is used as a model of mammalian drug metabolism (Moody et al., 2002). Recently, production of a putative ring-cleavage metabolite from flurbiprofen was detected in activated sludge exposed to flurbiprofen in the laboratory, but no stable enrichment culture was obtained (Yanaç and Murdoch, 2019). Thus, to date, no flurbiprofen-degrading consortium or isolate have been identified or characterized.

Our study sought to investigate the potential capabilities of environmental bacteria to degrade flurbiprofen. Through flurbiprofen-enrichments from activated sludge, we were able to identify a bacterial enrichment capable of using flurbiprofen as sole carbon source, and characterized two bacteria capable of degrading flurbiprofen, though maintaining these bacteria in pure culture proved to be difficult.

4.2. MATERIALS AND METHODS

4.2.1. Enrichment Conditions and Bacterial Isolates

A 500 mL sample of activated sludge from Ithaca Area Wastewater Treatment Facility (IAWTF, Ithaca, NY) was amended with 200 mg/L of flurbiprofen (98% purity, Acros Organics), and incubated with shaking (150 rpm) at room temperature. After a week, 20 mL of the enrichment

culture were transferred into 80 mL of fresh Minimal Salts Medium (MSM) containing 200 mg/L of flurbiprofen. The MSM medium was composed of 1 mM of $MgSO_4$, 10 mM of K_2HPO_4, 3 mM of NaH_2PO_4, 10 mM of $(NH_4)_2SO_4$, 10 µM of $Fe(NO_3)_3$, and 100 µM of $Ca(NO_3)_2$. Five weekly serial transfers (1:10) were performed in total. After the fifth transfer, the enriched culture was streaked onto MSM agar plates (16 g/L of noble agar) containing 200 mg/L of flurbiprofen. Then, bacterial colonies with distinctive morphology were isolated after serial transfers onto MSM agar plates containing 400 mg/L of flurbiprofen (Fisher Scientific).

An apparent isolate we named Flur-4 when grown on flurbiprofen transiently accumulated a bright yellow color indicating the appearance of a putative ring fission metabolite. Additional isolates eventually obtained from Flur-4 frozen stocks were selected using MSM agar plates supplemented with flurbiprofen (400 mg/L) and yeast extract (0.05 g/L) (MSMFY medium). Two distinctive colonies were observed, a bigger raised-rounded colony named here BC (big colony), and a smaller translucent pinpoint colony named PZ (phantom zone).

4.2.2. Analyses of Flurbiprofen-Amended Cultures

Flur-4 was grown in triplicate in liquid MSM containing 400 mg/L of flurbiprofen. Growth was measured by optical density at 600 nm (OD600) using a Synergy™ HTX Multi-Mode Microplate Reader (BioTek instruments). The appearance and accumulation of the yellow color was monitored via optical density at 410 nm (OD410, Synergy™ HTX Multi-Mode Microplate Reader, BioTek instruments) since it was found to be the wavelength for maximum absorbance of this putative ring fission metabolite. The cultures were allowed to grow until late stationary phase when the yellow color started to disappear.

Flurbiprofen disappearance and putative metabolites were monitored via gas chromatography-mass spectrometry (GC-MS). Culture supernatants were harvested by centrifugation and subjected to aqueous acetylation before extraction with ethyl acetate (1:1 vol/vol). Ethyl acetate extracts were dried over sodium chloride, methylated using diazomethane, and then subjected to GC-MS analysis as described by Murdoch and Hay (2005). Briefly, GC-MS analyses were performed in an HP 6890 GC equipped with a HP-5MS column (5 % phenyl methyl siloxane, 30 m by 0.25 mm, 0.25 μm film thickness) using helium as the carrier gas at a flow rate of 1 mL/min. The injector temperature was 250°C. The initial oven temperature of 40°C was held for 1.5 min and ramped at a rate of 5°C/min to 180°C; then ramped up to 200°C at a rate of 10°C/min. The temperature was lastly ramped up to 300°C at a rate of 20 °C/min and held for 10 min. The detector was an HP 5973 MSD with quadrapole and source set at 150°C and 230°C respectively.

4.2.3. Detection of Metabolites During Flurbiprofen Degradation

Flur-4 cultures (400 mL) were grown on MSM containing 400 mg/L of flurbiprofen until late exponential phase. Cells were harvested by centrifugation at 7000 x g for 15 min, then resuspended in 120 mL of MSM containing 400 mg/L of flurbiprofen. The resuspended cells were equally divided into 6 glass vials. Three vials received 50 mg/L of 3-fluorocatechol (3-FC, 98% purity, Acros Organics) and the other three remained as controls (no fluorocatechol addition). After incubation for 30 min at room temperature, the cultures were centrifuged at 12000 g for 15 min, and the supernatants were collected and derivatized for GC-MS analysis as described above.

4.2.4. 16S rRNA Gene PCR and Metagenome Analysis

Direct PCR with 16S rRNA gene primers (27F: 5'- GAGAGTTTGATCMTGGCTCA -3' and 1492R: 5'- TACGGYTACCTTGTTACGACTT -3') was performed using colonies grown on MSM agar plates containing 400 mg/L of flurbiprofen. PCR products were purified using a Wizard® SV gel and PCR clean-up system (Promega), quantified with a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific), and sequenced at Cornell's Institute of Biotechnology. The 16S rRNA gene sequences were analyzed using the BLASTn tool from NCBI (Basic Local Alignment Search Tool).

DNA from Flur-4, and from the BC and PZ cultures was extracted following Yu and Forster's protocol (2005) with modifications. Briefly, cells harvested from 20 mL cultures grown in liquid MSM containing 400 mg/L of flurbiprofen were resuspended with lysis buffer [2% (wt/vol) SDS, 100 mM Tris HCl, 5mM EDTA, 200 mM NaCl, pH 8.0] and vortexed vigorously with 0.1 mm glass beads. DNA extracts were precipitated with 2 M ammonium acetate and purified with isopropanol (1:1 vol/vol). A final DNA clean-up was carried out with 70% ethanol. DNA concentrations were determined using the Quant-iT^{TM} PicoGreen^{TM} dsDNA assay kit (ThermoFisher Scientific). Then, purified DNAs were prepared by Cornell's Institute of Biotechnology for shotgun sequencing.

The web-based platform Kbase (The U.S. Department of Energy Systems Biology Knowledgebase) was used to process the raw reads and perform further genome/metagenome analyses including *de novo* assembly and binning (Arkin et al., 2018). Raw reads were pre-analyzed using FastQC v0.11.5 and processed with Trimmomatic v0.36. Later, trimmed reads were *de novo* assembled with metaSPAdes v3.12.0, and the resulting contigs were binned with MaxBin2 v2.2.4. CheckM v1.0.8 was used to assess the binned contigs quality (Parks et al.,

2015). Thereafter, individual assemblies were extracted using BinUtil v1.0.2 and annotated via RAST 2.0 (Rapid Annotations using Subsystems Technology) (Aziz et al., 2008). Taxonomic classification was performed using KAIJU v1.5.0, which uses protein sequence similarities. Additional phylogenetic and functional analysis of the metagenomes were generated via RAST 2.0 and MG-RAST 4.0.3 (Metagenomic-RAST) (Aziz et al., 2008; Keegan et al., 2016).

Annotations of the assemblies were examined using Geneious 11.0.2. Nucleotide and amino acid sequence alignments of genes and enzymes of interest were generated with ClustalW 2.1 and blosum62 as cost matrix (Larkin et al., 2007). Phylogenetic trees were constructed with Geneious tree builder that uses global alignment with free end gaps and blossum62 to generate a distance matrix. The Jukes-Cantor model and neighbor-joining method were used to build the tree. BLAST tools from the NCBI (tBLASTn, BLASTn, BLASTp) allowed comparisons of the genome annotations against repository databases.

4.2.5. Screening of Flurbiprofen Ring-Fission Metabolism in Fosmid Library

DNA from the BC isolate was extracted as mentioned above. The fosmid library was created using the CopyControl™ Fosmid Library Production Kit with pCC1FOS™ Vector following the manufacturer's instructions and hosted in the provided *Escherichia coli* EPI300 strain (Epicentre Biotechnologies).

After initial selection on LB plates containing chloramphenicol (12 mg/L), fosmid library colonies were grown in ¼ LB (2.5 g/L tryptone, 1.25 g/L NaCl, and 1.25 g/L yeast extract) with 200 mg/L of flurbiprofen in 96-well plates and screened for the appearance of a yellow ring-fission metabolite. The fosmid autoinduction solution provided in the CopyControl™ Fosmid kit was used to promote high copy number of the fosmids. In separate replicate plates, ferric

chloride (1.5 mM) was used to screen for ring dioxygenation due to its reaction with catecholic metabolites to produce a brown oxidation product (Murdoch and Hay, 2005; Soloway and Wilen, 1952).

4.3. RESULTS AND DISCUSSION

4.3.1. A Bacterial Enrichment Capable of Growing on Flurbiprofen

Flur-4 culture was selected from the flurbiprofen enrichments based on its ability to use flurbiprofen as a sole carbon source. Flur-4 entered exponential growth phase about 30 h after inoculation into liquid MSM amended with 400 mg/L of flurbiprofen (MSMF). Stationary phase was reached after approximately 60 h with a maximum OD600 of 0.15 ± 0.03 (Fig. 1A). When amended with flurbiprofen, Flur-4 cultures accumulated a bright yellow color in the supernatant, which is indicative of aromatic ring-cleavage as has been observed during the metabolism of other aromatic compounds such as catechols and the NSAID ibuprofen (Murdoch and Hay, 2005; Riegert et al., 1998). The yellow color produced by Flur-4 was found to absorb maximally at 410 nm (OD410). The OD410 was found to peak at the end of the exponential growth phase (approximately at 55 h) and then decreased by about 75% after 100 h (Fig. 1A). The yellow color disappeared when Flur-4 culture was acidified and reappeared upon neutralization. pH-dependent yellow coloration that absorbs in the range of 375-425 nm are typical of meta-cleavage products of aromatic rings, presumptively indicating dioxygenase activity (Harayama and Rekik, 1989; Suenaga et al., 2014). Flurbiprofen concentration, as monitored via GC/MS, sharply decreased as Flur-4 entered exponential growth phase, with approximately 99% of the initial flurbiprofen disappearing by 60 h (Fig. 1B) coincident with maximal yellow color accumulation.

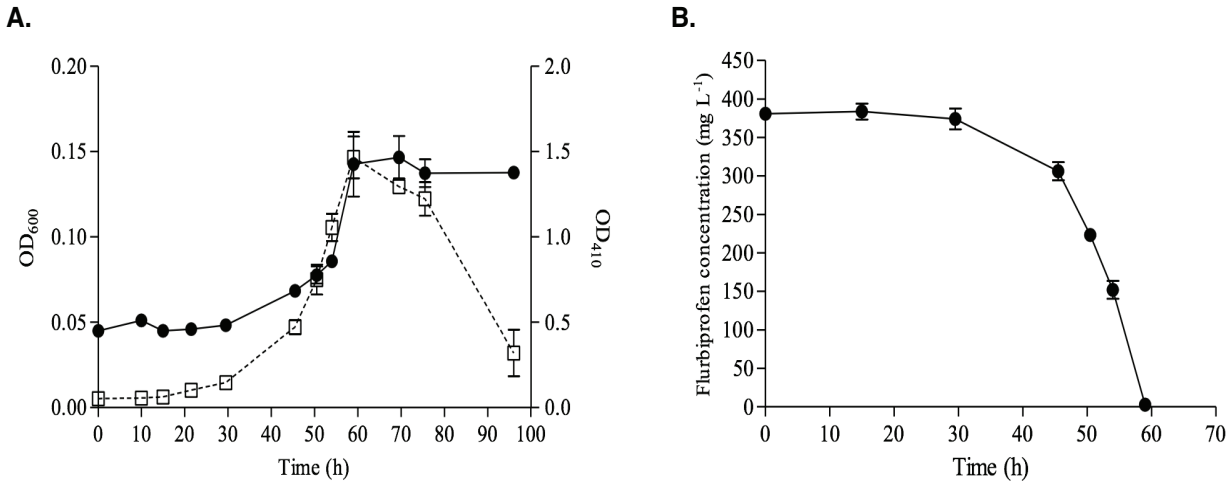


Fig 1. A. Growth of Flur-4 on flurbiprofen as measured by increased culture density, OD_{600} (black dot), and concomitant accumulation of presumptive yellow ring-fission metabolite measured via OD_{410} (white square). **B.** Flurbiprofen concentration measured via GC/MS during Flur-4 growth until reaching exponential phase. Error bars represent standard deviation.

4.3.2. Flurbiprofen degradation by Flur-4 cultures via a biphenyl-like pathway

When flurbiprofen-amended Flur-4 cultures were exposed to 3-flurocatechol (3FC), there was no accumulation of yellow color; but a new peak for which no standard was available appeared in the total ion chromatogram of acetylated and methylated ethyl-acetate supernatant extracts during the GC/MS analysis. The mass spectrum of this peak yielded a weak molecular ion of m/z 374, a fragment of m/z 332 and a base peak at m/z 290 (Fig 2). The consecutive losses of m/z 42 which yielded these two fragments are consistent with the loss of acetyl groups. Since aqueous acetylation is specific for aromatic hydroxyl groups (Boyd, 1994; Fritz and Schenk, 1959; Itoh et al., 2005), these losses suggest that the compound detected after 3FC-exposure, referred to as compound II (Fig 2), contained two aromatic hydroxyls. In addition, the base peak of m/z 290 is m/z 32 heavier than the methylated spectrum of flurbiprofen (m/z 258), which is also consistent

with the addition of two hydroxyl groups to flurbiprofen methyl ester. These results suggest that compound II correspond to dihydroxyflurbiprofen (DHF).

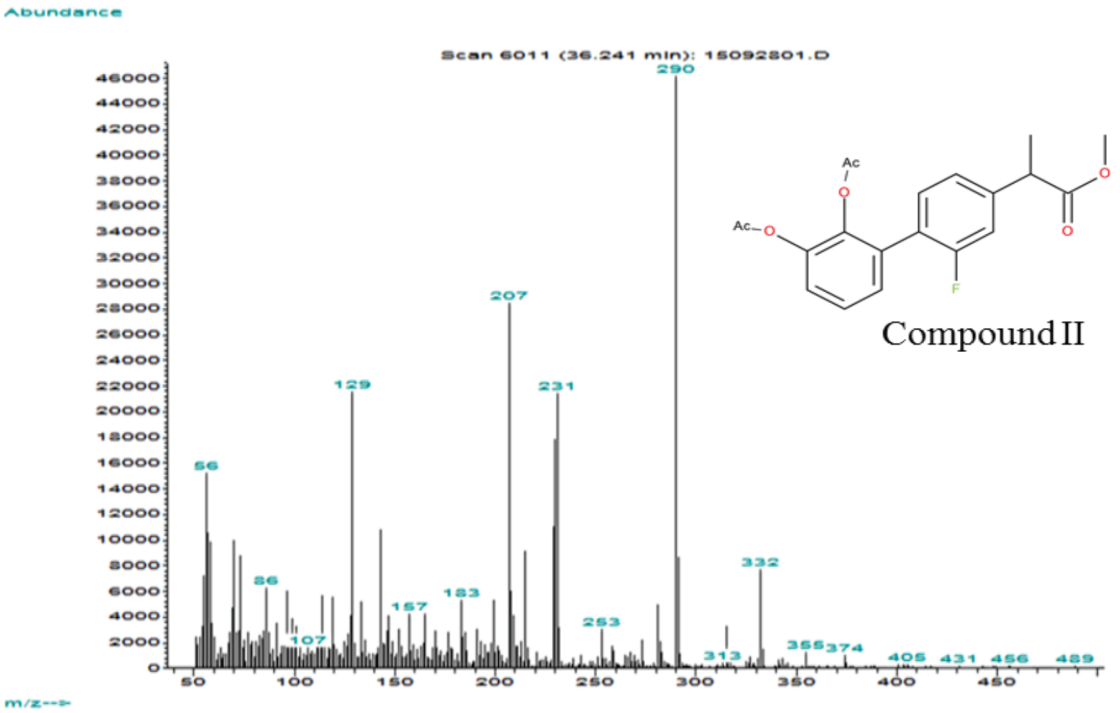


Fig 2. Mass spectrum of the putative acetylated dihydroxy-flurbiprofen methyl ester (Compound II: DHF) identified via GC/MS. This spectrum was only detected after 3-fluorocatechol-poisoning of flurbiprofen-amended Flur-4 cultures followed by aqueous acetylation and methylation. Ac: acetyl group.

Further evidence in support of compound II being DHF comes from two related observations: 1) the compound II GC/MS spectrum was not detected in the absence of 3FC, a known poison of meta-cleavage dioxygenases (Dorn and Knackmuss, 1978; Engesser et al., 1988; Mars et al., 1997); and 2) no meta-cleavage product was detected when 3FC was added. Hence, our findings are consistent with flurbiprofen being metabolized via a biphenyl-like degradation pathway that begins with initial dioxygenation of the unsubstituted aromatic ring followed by ring fission and production of a meta-cleavage metabolite (Compound III in Fig. 3) (Mars et al., 1997; Mondello, 1989; Seo et al., 2009).

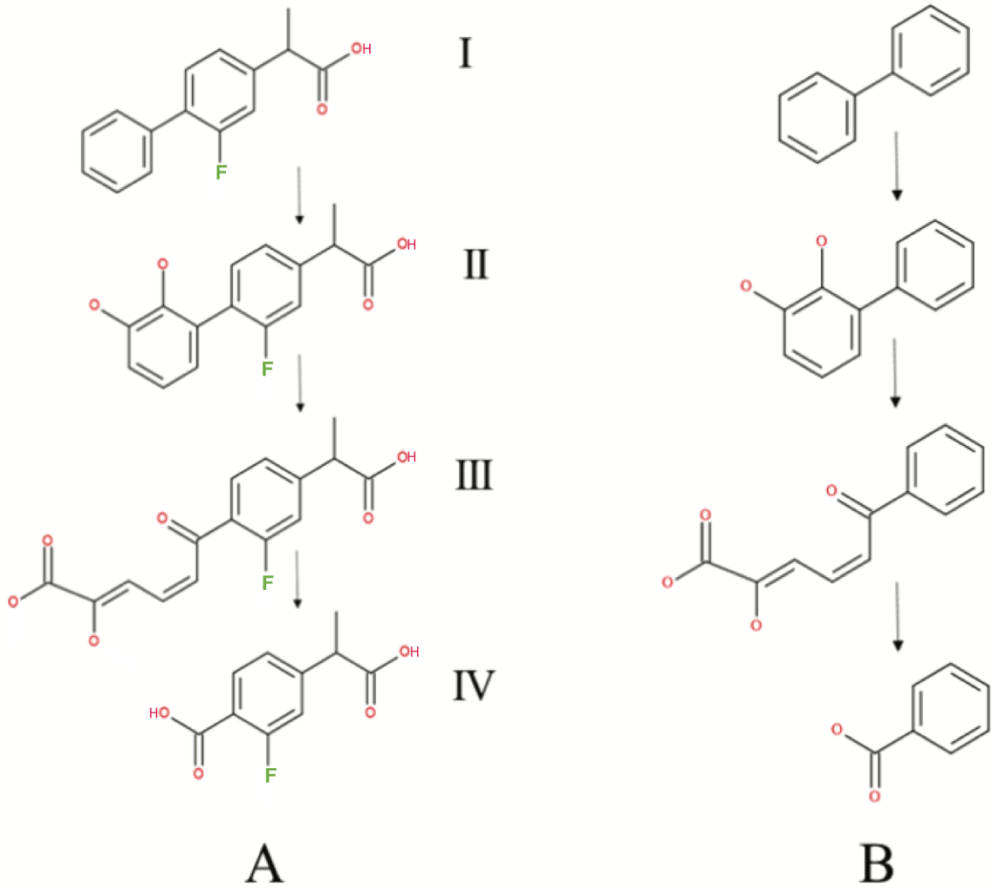


Fig 3. Putative pathway of flurbiprofen metabolism by Flur-4 (A) is similar to the biphenyl degradation pathway provided for comparison (B) (Mondello 1989). In A, compound II and compound IV were both detected via GC/MS. Compound III was inferred due to the transient accumulation of yellow color in the supernatant. I: flurbiprofen; II: dihydroxyflurbiprofen (DHF); III: presumptive meta-cleavage product of flurbiprofen; IV: 3-fluro-4-(1-carboxyethyl)benzoic acid (FCB)

A third observation supporting the hypothesis that Flur-4 employed a biphenyl-like pathway for flurbiprofen metabolism, was the accumulation of the putative flurbiprofen metabolite referred to as compound IV in Fig 3A. Although there was no standard available for compound IV, we were able to characterize its mass spectrum and showed its accumulation in the supernatant during Flur-4 growth in liquid MSMF (Fig 4). The mass spectrum of compound IV from methylated supernatant extracts showed a molecular ion of m/z 240 and yielded fragments with m/z 209 and m/z 181. These losses of m/z 31 and 59 respectively are consistent with a compound containing two methyl esters and with the structure of 3-fluro-4-(1-

carboxyethyl)benzoic acid (FCB). Production of FCB is analogous to the production of benzoic acid during biphenyl degradation (Mondello, 1989; Triscari-Barberi et al., 2012) and is in line with our detection of DHF and the putative meta-cleavage product (Fig. 3). Importantly, FCB was not detected when cells were poisoned by 3FC, providing circumstantial evidence that ring-cleavage was essential for the production of FCB and its precursor.

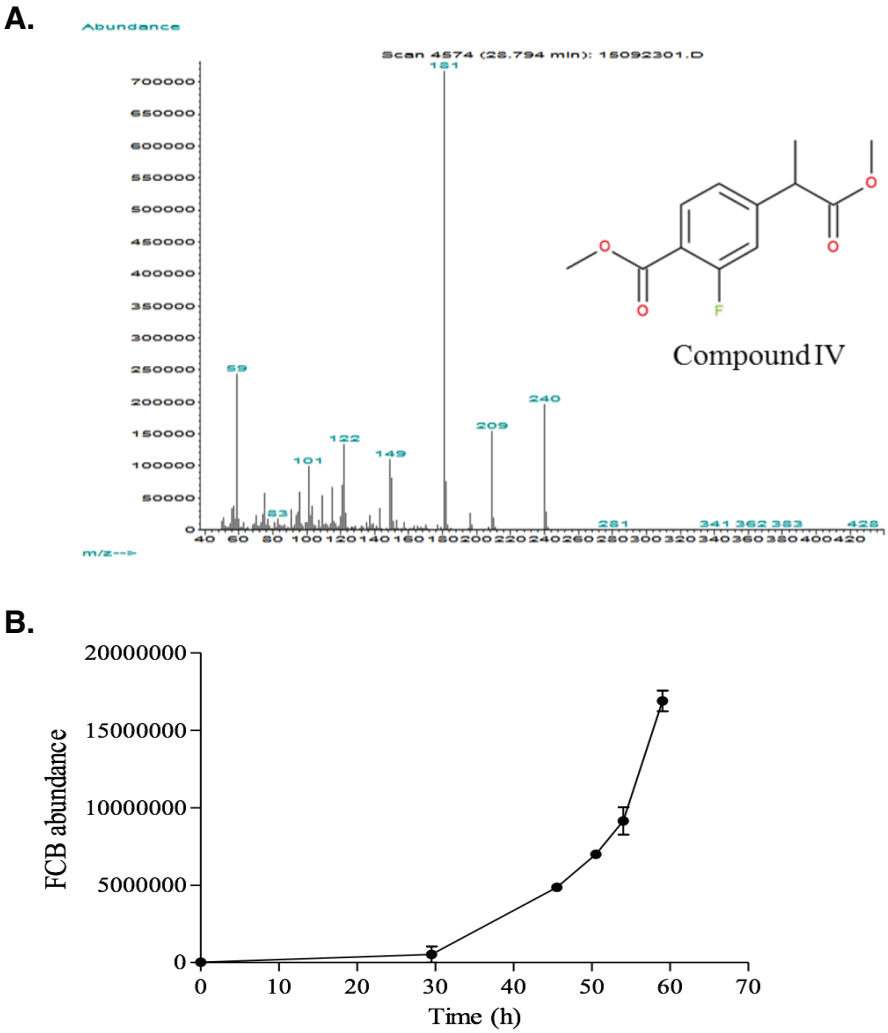


Fig. 4. A. Mass spectrum of the putative 3-fluro-4-(1-carboxyethyl)benzoic acid metabolite (Compound IV: FCB) with two methyl ester groups as result of diazomethane derivatization during GC/MS analysis. **B.** FCB accumulation during Flur-4 growth in flurbiprofen-amended MSM measured via GC/MS

The degradation of flurbiprofen thus appears to be analogous to the degradation of 2- and 4-fluorobiphenyl by *Pseudomonas pseudoalcaligenes* KF707, which also proceeds via a standard biphenyl degradation pathway yielding 2- and 4-fluorobenzoate as end products (Murphy, 2016; Murphy et al., 2008). The literature regarding the fate of single-ring fluoroaromatics is somewhat sparse. While successful degradation of 2- and 4-fluorobenzoates have been reported in several studies, 3-fluorobenzoates have shown to be recalcitrant due to the production of toxic intermediates (Engesser et al., 1988; Hughes et al., 2011; Seeger et al., 2001). It is not clear whether the apparent dead-end nature of FCB is due to the presence of the fluorine on the ring, the presence of the propanoic acid moiety, or both.

Unlike most popular NSAIDs, flurbiprofen is fluorinated. Fluorine substituents introduce many useful properties to pharmaceuticals and are generally incorporated in order to increase the pharmaceuticals biological half-lives (Gaye and Adejare, 2009; Shah and Westwell, 2007). On the other hand, introduction of fluorine into chemicals creates environmental problems due to increased lipophilicity and recalcitrance (Khetan and Collins, 2007; Murphy, 2016). The unfluorinated analog of FCB, 4-(1-carboxyethyl) benzoic acid has been previously reported as an abiotic degradation product of ibuprofen (Caviglioli et al., 2002), though to the best of our knowledge there is no information about 4-(1-carboxyethyl) benzoic acid fate or bioactivity.

Similarly to our findings about flurbiprofen metabolites, a recent biodegradation study using activated sludge reported the transient appearance of an acid-labile yellow color when flurbiprofen was added, followed by the accumulation of FCB which was tentatively identified via LC/MS (Yanaç and Murdoch, 2019). Similar to our own results, they did not detect any further degradation of FCB. An aerobic pathway for the degradation of substituted phenylacetic acids via catecholic intermediates has been previously described for the biotransformation of ibuprofen and related compounds (Murdoch and Hay, 2013, 2005). Yanaç and Murdoch (2019)

hypothesized that FCB might not undergo further biotransformation since organisms doing so could produce a toxic fluorinated-catechol (3-fluoro-4-phenylcatechol). Given the structural similarity of this theoretical fluorinated-catechol to 3-fluorocatechol, a known ring-cleavage poison, an FCB derived fluorinated-catechol would be expected to inhibit flurbiprofen metabolism via meta-cleavage, which may explain the apparent dead-end nature of FCB.

4.3.3. Metagenome Analyses of Flur-4 as a Putative Flurbiprofen-Degrading Bacterial Consortium

Initial BLASTn analysis of the 16S rRNA gene sequence revealed that our Flur-4 culture, which at the time was believed to be a pure isolate, was 97% identical to *Phenylobacterium immobile*. *P. immobile* is well known for its eponymous ability to degrade a narrow range of aromatic compounds including chloridazon, antipyrin and pyramidon (Eberspächer, 2015; Lingens et al., 1985). To the best of our knowledge, there is no report of flurbiprofen degradation by *P. immobile* or by other strains belonging to the *Phenylobacterium* genus.

Given that our biochemical data suggested that Flur-4 metabolized flurbiprofen through a biphenyl-like degradation pathway, we investigated the presence of biphenyl-like degrading genes (*bph*-like genes) in Flur-4 genome assembly. Preliminary analysis of the Flur-4 genome annotations showed scattered putative genes involved in biphenyl degradation, but no obvious biphenyl-like ring-hydroxylating dioxygenase that could be responsible for metabolizing flurbiprofen to dihydroxyflurbiprofen (DHF). The tBLASTn hits to the enzymes of the well-characterized biphenyl-degrading operon of *Pseudomonas furukawaii* KF707 (NCBI accession number: AJMR01000119.1) showed 32.5% - 44.8% pairwise similarity to translated sequences of putative genes in the Flur-4 genome (Supplementary Materials Table S1). None of the hits,

however, aligned on the same contig, thus they were not localized in an operon resembling most previously described biphenyl-degraders like *Burkholderia sp.* LB400 and *P. furukawaii* KF707 (previously named *Pseudomonas pseudoalcaligenes* KF707) (Furukawa et al., 2004; Mondello, 1989; Taira et al., 1992).

Subsequent metagenome analyses performed through MG-RAST 4.0.3 revealed that Flur-4 was not a pure culture as previously believed, but rather was a bacterial consortium composed mainly by *Phenylobacterium* (49.5%), *Caulobacter* (20.2%), *Brevundimonas* (3.3%), and *Sphingomonas* (3.2%) (Supplementary Material Fig S1). Thus, the presence of various bacterial genomes in Flur-4 sequencing data might explain the scattering of the putative *bph*-like genes and the absence of a single biphenyl-like operon that could be responsible for flurbiprofen degradation.

4.3.4. Re-examining Flurbiprofen-Degrading Isolates Using a Metagenomic Approach

After re-examining our initial Flur-4 frozen stocks by plating them on a flurbiprofen minimal medium (MSMFY), we observed colonies with a raised well-defined rounded shape (3-5 mm), surrounded by smaller translucent pinpoint colonies (~1 mm) (Fig. 5A). Despite the multiple attempts to isolate the bigger raised-rounded colonies, the smaller pinpoint colonies always appeared in the agar plates even in MSM plates supplemented solely with flurbiprofen. In fact, the bigger raised-rounded colonies seemed to grow on top of the smaller translucent pinpoint colonies. The bigger raised-rounded colonies were named BC (big colony), and the smaller translucent pinpoint colonies PZ (phantom zone). Though we were able to isolate PZ colonies on MSMFY agar plates, no yellow color appeared on either agar plates or in liquid MSMFY (Fig. 5B). Analysis of the 16S rRNA gene sequences showed that they were identical to

each other, but neither BC or PZ were a *Phenylobacterium* strain; instead, both were closely related to *Ochrobactrum* (97-99% similarity). These findings were surprising, since *Ochrobactrum* constituted a very small component of the initial Flur-4 bacterial consortium at the time it was sequenced (0.22%) (Supplementary Materials Fig S1). The reason for this discrepancy was unclear, though it may have been the consequence of changes in the consortium during prolonged culturing or enrichment after sequencing, but before final storage in the culture collection. To gain more insight to the nature of the only flurbiprofen-degrading isolates we could recover from the stored consortium, genomic analyses were performed on both BC and PZ cultures.

After processing the raw sequences of BC and PZ with metaSPAdes v3.13.0 and

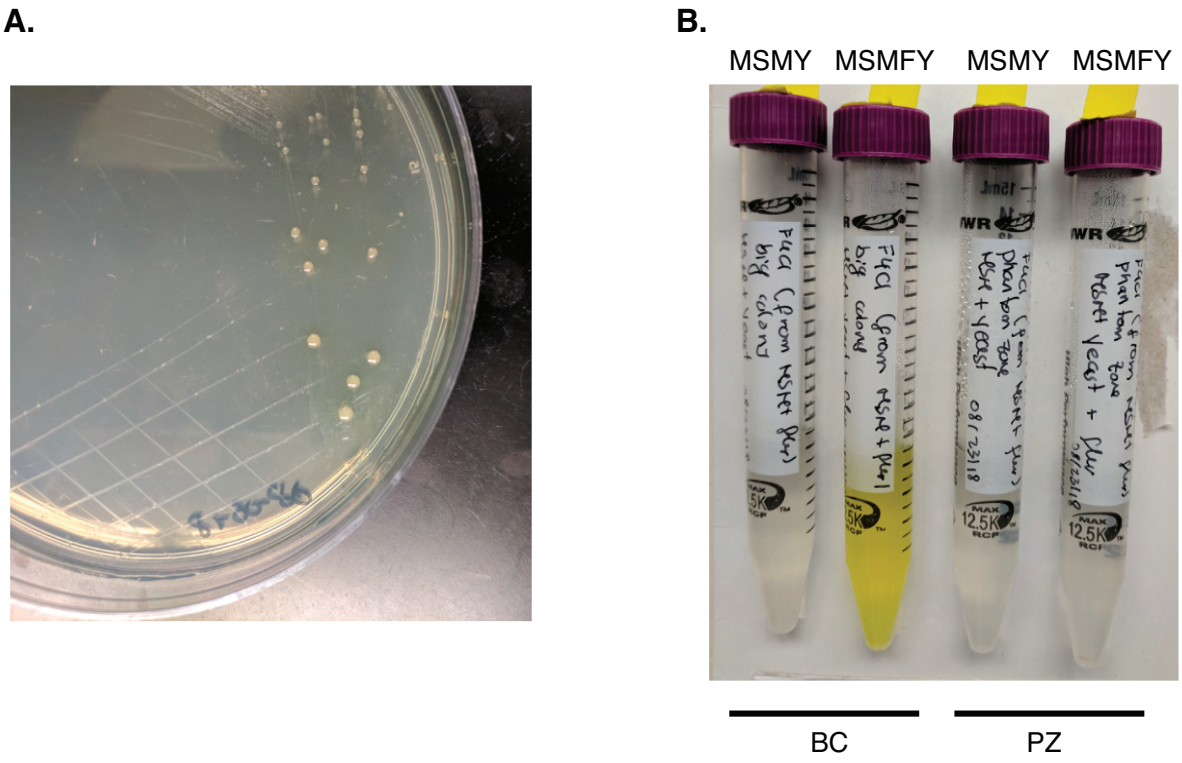


Fig 5. A. Morphology of the two distinctive colonies selected on MSM agar plates supplemented with flurbiprofen (400 mg/L) and yeast extract (0.05 g/L), after re-examining Flur-4 frozen stocks. A bigger raised-rounded colony (BC: big colony) and a smaller translucent pinpoint colony (PZ: phantom zone). **B.** Appearance of putative ring-fission metabolite (yellow color) in the liquid MSMFY inoculated with BC, but not with PZ. MSMFY: liquid MSM supplemented with flurbiprofen (400 mg/L) and yeast extract (0.05 g/L); MSMY: liquid MSM only supplemented with yeast extract (0.05 g/L), but no flurbiprofen.

MaxBin2 v2.2.4, only one metagenome-assembled genome (MAG) was recovered from each culture. The two MAGs, one from BC and one from PZ, were both identified as *Alphaproteobacteria* belonging to the genus *Ochrobactrum* confirming our preliminary analysis of the 16S rRNA genes. The two MAGs were named: *Ochrobactrum* BC and *Ochrobactrum* PZ. CheckM v1.0.8 determined that both *Ochrobactrum* MAGs were "near-complete" (100% completeness), but with a contamination of 15.2% for BC and 12.0% for PZ (Supplementary Material Fig S2A). Moreover, analysis using protein sequence similarity via KAIJU v1.5.0 found that 66.2% and 63.7% of the BC and PZ genome sequences, respectively, belonged to the genus *Ochrobactrum* (Supplementary Materials Fig S2B). Though this contamination suggested the presence of other bacterial genomes, we were not able to reconstruct additional MAGs from the raw sequences likely due to sequence depth limitations.

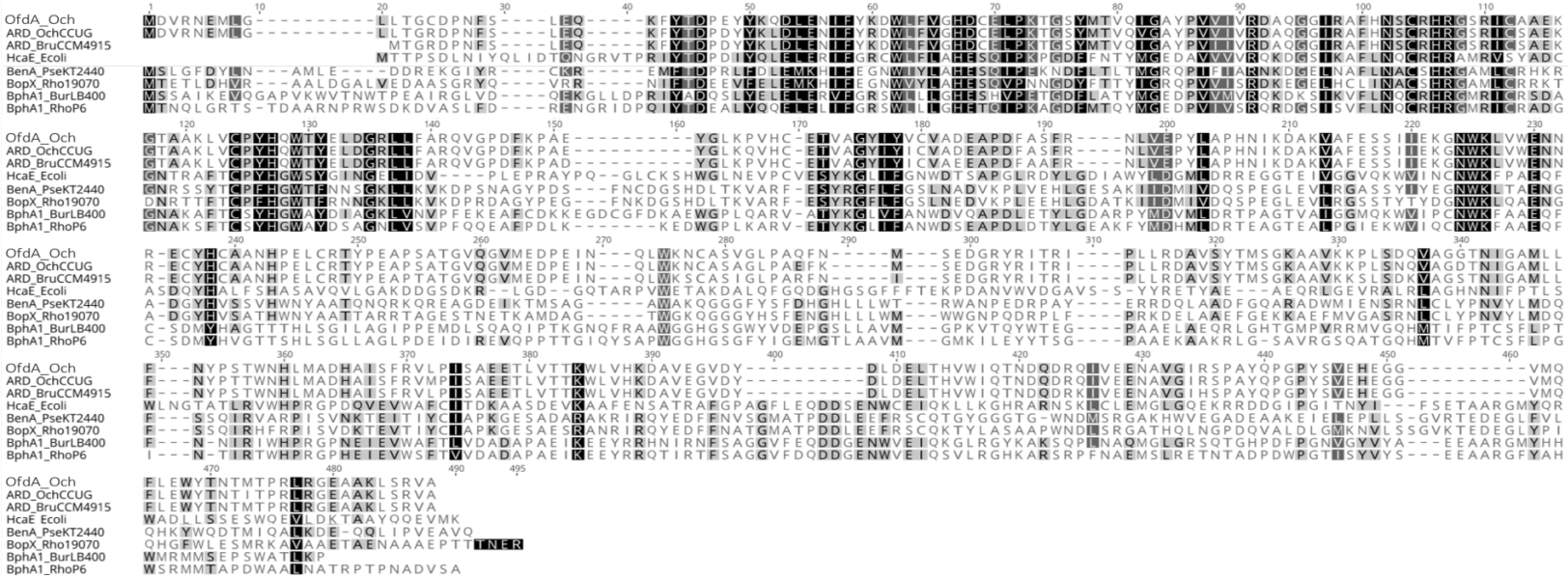
From the tBLASTn analyses performed on the *Ochrobactrum* MAGs with the well-characterized biphenyl-degrading operon from *P. furukawaii* KF707, the BphA1 enzyme (Biphenyl dioxygenase α-subunit) was found to align with 33.1% similarity to the translated sequence of a putative gene annotated as benzoate 1,2-dioxygenase (Table 1). This putative gene was identical in both *Ochrobactrum* MAGs. For the purposes of identification in this paper we named it: *ofdA* (putative flurbiprofen dioxygenase α-subunit from *Ochrobactrum*). Although a putative benzoate 1,2-dioxygenase was also found in our preliminary analyses of the Flur-4 genome assembly (Supplementary Materials Table S1), its alignment to the translated sequence of *ofdA* showed 42.49% similarity.

Table 1. Best tBLASTn hits found in the *Ochrobactrum* BC and PZ metagenome-assembled genomes (MAGs) using the *bph* operon of *P. furukawaii* KF707 as query (NCBI accession number: AJMR01000119.1). BphA2, BphA3, BphC, and BphD did not return any hit when compared against the *Ochrobactrum* MAGs.

Bph enzyme	Pairwise similarity	Query coverage	Annotation in BC and PZ given by RAST
BphA1 biphenyl dioxygenase α-subunit	33.1%	37.0%	benzoate 1,2-dioxygenase
BphA4 ferredoxin reductase	30.2%	98.5%	ferredoxin reductase
BphB dihydrodiol dehydrogenase	47.0%	94.0%	3-oxyl-acyl-carrier reductase

Interestingly, when the translated sequence of *ofdA* was compared against the non-redundant protein database of the NCBI, the tBLASTn hits had more than 90% similarity to putative aromatic ring-dioxygenases (ARDs) annotated in genomes of *Ochrobactrum* and *Brucella* isolates. Yet, no biochemical data has been published to provide clues about substrate specificity or to confirm the function of these putative ARDs. The tBLASTn results in the NCBI, however, identified the translated sequence of *ofdA* as part of the 3-phenylpropionate dioxygenase family (HcaE). When compared against the well-characterized HcaE from *Escherichia coli,* the translated sequence of *ofdA* had 48.39% similarity (Fig 6). By comparing the translated sequence of *ofdA* to putative ARDs from *Ochrobactrum* and *Brucella* genomes, and to functionally-characterized 3-phenylpropionate-, biphenyl- and benzoate- 1,2-dioxygenases, we found that the translated sequence of *ofdA* formed a distinctive clade with the putative ARDs (Fig 7). While the translated sequence of *ofdA* was 45-48% similar to well-characterized dioxygenases, with the putative ARDs the translated sequence of *ofdA* was almost identical in the ClustalW alignment (Fig 6).

A.



B.

	OfdA_Och	ARD_OchC...	ARD_BruC...	HcaE_Ecoli	BenA_Pse...	BopX_Rho...	BphA1_Bu...	BphA1_Rh...
OfdA_Och		99.04%	99.26%	48.39%	45.59%	46.72%	45.43%	45.34%
ARD_OchCCUG	99.04%		99.26%	48.60%	46.04%	46.94%	45.65%	45.12%
ARD_BruCCM4915	99.26%	99.26%		47.84%	45.45%	47.39%	46.14%	45.70%
HcaE_Ecoli	48.39%	48.60%	47.84%		48.93%	50.53%	61.79%	61.16%
BenA_PseKT2440	45.59%	46.04%	45.45%	48.93%		86.27%	53.90%	53.52%
BopX_Rho19070	46.72%	46.94%	47.39%	50.53%	86.27%		54.21%	53.83%
BphA1_BurLB400	45.43%	45.65%	46.14%	61.79%	53.90%	54.21%		85.84%
BphA1_RhoP6	45.34%	45.12%	45.70%	61.16%	53.52%	53.83%	85.84%	

Fig 6. ClustalW2 alignment (**A**) and pairwise similarity matrix (**B**) of the putative flurbiprofen-dioxygenase found in the *Ochrobactrum* BC and PZ MAGs (OfdA_Och) versus putative aromatic dioxygenases annotated in *Ochrobactrum* and *Brucella* genomes (ARD_OchCCUG and ARD_BruCCM4915, respectively), and well-characterized dioxygenases: 3-phenylpropionate dioxygenase α-subunit from *E. coli* (Hca_Ecoli), benzoate 1,2-dioxygenase α-subunit from *Pseudomonas sp.* KT2440 and *Rhodococcus sp.* 19070 (BenA_PseKT2440 and BopX_Rho19070, respectively), and biphenyl 1,2-dioxygenase α-subunit from *Burkholderia sp.* LB400 and *Rhodococcus sp.* P6 (BphA1_BurLB400 and BphA1_RhoP6, respectively).

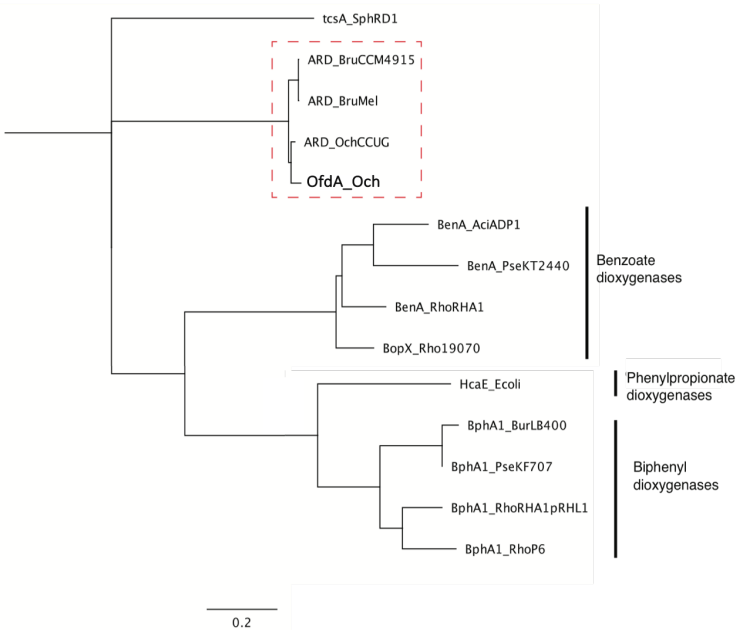


Fig 7. Phylogenetic tree with well-characterized dioxygenases showed that the putative flurbiprofen-dioxygenase from *Ochrobactrum* BC and PZ MAGs (OfdA_Och) grouped with putative aromatic dioxygenases annotated in *Ochrobactrum* and *Brucella* genomes (red dashed box) (ARD_OchCCUG and ARD_BruCCM4915, respectively). 3-phenylpropionate dioxygenase α-subunit from *E. coli* (Hca_Ecoli), benzoate 1,2-dioxygenase α-subunit from *Pseudomonas sp.* KT2440 and *Rhodococcus sp.* 19070 (BenA_PseKT2440 and BopX_Rho19070, respectively), and biphenyl 1,2-dioxygenase α-subunit from *Burkholderia sp.* LB400 and *Rhodococcus sp.* P6 (BphA1_BurLB400 and BphA1_Rho). TscA_SphRD1 (putative triclosan dioxygenase from *Sphingomonas* RD1) was used as outgroup (Furukawa et al., 2004, Kagle et al., 2015)

Remarkably, the genomic context of *ofdA* in both *Ochrobactrum* MAGs was highly similar to that of the putative genes encoding the ARDs in the genomes of *Ochrobactrum* and *Brucella,* with a putative NADH oxidoreductase (named here *ofdB*) next to the putative dioxygenases in all cases (Fig 8). These findings suggest that our *Ochrobactrum* genomes (BC and PZ) encode a novel two-component dioxygenase two-component (hydrolase and electron transferase). Although further cloning and expression of these genes is required to determine their function, we speculate that they might be involved in the dioxygenation of the non-halogenated ring as the first step of flurbiprofen metabolism. Though four-component dioxygenases have generally been found to be responsible for the metabolism of biphenyl to 2,3-dihydroxy-4-phenylhexa-4,6-diene (Taira et al., 1992), two-component ring-hydroxylating dioxygenases have been reported in the metabolism of other aromatic compounds (Furukawa et al., 2004, Kagle et al., 2015).

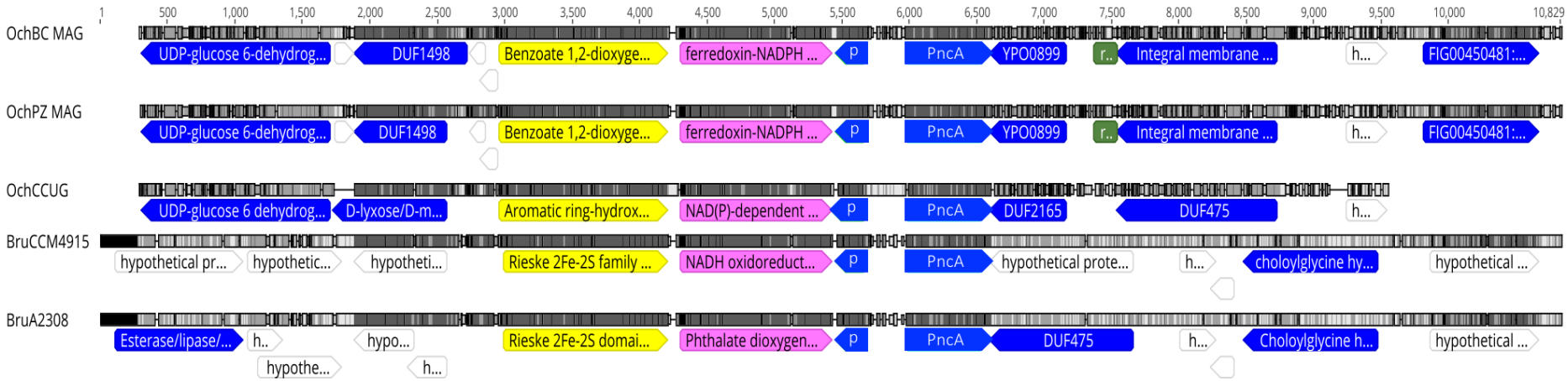


Fig 8. Nucleotide alignment of the putative dioxygenase OfdA (in yellow) and NADH reductase, OfdB (in pink) found in both *Ochrobactrum sp.* BC and PZ MAGs (OchBC and OchPZ, respectively), and their genomic context compared to the putative aromatic ring-dioxygenases and reductases annotated in *Ochrobactrum* and *Brucella* genomes (GenBank assembly accession: OchCCUG: GCA_003049685.2, BruCCM4915: GCA_000022745.1, BruA2308: GCA_000054005.1). P.: PhnA gene, r.: repeat region (dark green), h.: hypothetical protein (white)

Given that *ofdA* and *ofdB,* and their genomic context were identical between both *Ochrobactrum* MAGs, we compared the rest of their genomes in order to explain why PZ could not produce the yellow ring-fission product, but BC could (Fig 5). Whole genome comparison via RAST found that 124 putative genes in BC were absent in PZ, three of which were annotated as encoding proteins involved in aromatic metabolism: a muconate cycloisomerase, a BenR transcriptional activator of a benzoate-degrading operon, and a n-hydroxybenzoate hydroxylase involved in gentisate degradation (Supplementary Materials Table S2). There were no more gentisate-degrading genes annotated in either of the *Ochrobactrum* MAGs, and the putative n-hydroxybenzoate hydroxylase was only 75 amino-acids long, which is approximately 4X smaller than related enzymes with known activity. Given these findings, we hypothesized that the lack of BenR in PZ MAG prevents upregulation of the genes encoding OfdA, resulting in no ring fission product from flurbiprofen. Studies in *Pseudomonas putida* have demonstrated that BenR induces the degradation of benzoate and methylbenzoate, and also functions with the regulator XylS during toluene degradation (Cowles et al. 2000). Future work, however, needs to be done to test BenR's role in the degradation of flurbiprofen. This should include PCR confirmation of its absence in PZ followed by cloning and expression of *benR* from BC ($benR_{BC}$) in PZ to determine if it restores flurbiprofen degradation.

We also interrogated the *Ochrobactrum* MAGs for genes encoding a traditional meta-ring cleavage dioxygenase similar to BphC in biphenyl pathways and hydrolase similar to BphD that would further metabolize the ring fission intermediate of flurbiprofen to the benzoic acid-derivative FCB. To our surprise, there were no other tBLASTn hits or annotations to these other biphenyl-degrading enzymes, which suggest that neither of the *Ochrobactrum* MAGs has a complete biphenyl-like pathway, or that it is so divergent as to escape detection using the default annotation functions in RAST.

Since we had previously detected ring hydroxylating and ring-cleaving phenotypes in fosmid libraries constructed from aromatic-degraders (Kagle et al., 2015; Murdoch and Hay, 2013; Porter and Hay, 2007), we carried out a functional screen of a BC fosmid library since BC accumulated a yellow color indicative of flurbiprofen ring-cleavage (Fig 5). Unfortunately, after the exhaustive screening of our *E. coli* EPI300 fosmid library (approximately 2,300 colonies in total), we were not able to detect any clone exhibiting the expected yellow color. Based on our metagenome analyses, *Ochrobactrum* BC MAG had an estimated size of 5.8Mb. Therefore, our fosmid library had an approximate 16X coverage, which is at least three times more than what is recommended to ensure 99% probability of getting the genome sequence of interest in the library. In fact, we previously used fosmid libraries with 5X-8X coverage to identify novel aromatic-degrading operons (Kagle et al., 2015; Murdoch and Hay, 2013; Porter and Hay, 2007). In addition to screening for the production of yellow color, we also screened for the accumulation of catecholic intermediates via addition of ferric chloride. The latter method has proven to be effective for visualizing the accumulation of dihydroxylated metabolites of aromatic compounds in 96-well plate format in fosmid libraries of other aromatic degraders (Kagle et al., 2015; Murdoch and Hay, 2013). Failure to detect flurbiprofen degradation activity in our fosmid library could be due to limitations in our detection methods, poor expression, or even because another factor encoded in trans was required for expression such as the putative transcriptional activator BenR mentioned above in the BC MAG. Unlike the benzoate operon, where *benR* is located contiguously with the genes encoding benzoate-degrading enzymes (Cowles et al. 2000), $benR_{BC}$ was not contiguous with *ofdA* and was located on a separate contig without any genes encoding putative aromatic-degrading enzymes next to it. Given our past success with the fosmid library approach and the genomic information about *ofdA* in the *Ochrobactrum* MAGs, we hypothesize that failure to detect flurbiprofen dioxygenase and ring cleave activity in our fosmid

library was because *benR* or some other required gene was not on the same fosmid. Alternatively, considering our previous findings with the Flur-4 metagenome (Supplementary Materials Fig S1 and Table S1), and the contamination detected in the *Ochrobactrum* MAGs (Supplementary Materials Fig S2), it is also possible that the putative flurbiprofen-degrading genes would not all be encoded by a single organism.

A BphC-like dihydroxy dioxygenase was previously found in the initial Flur-4 genome assembly with 42.9% similarity to the BphC from *P. furukawaii* KF707 (Supplementary materials Table S1), but it was 88.9% similar to a putative dihydroxyphenylacetate dioxygenase (HpaD) annotated in *Phenylobacterium immobile* genome (NCBI accession number: WP_091743113). This finding made us wonder whether a *Phenylobacterium* strain might be part of the contamination detected in the *Ochrobactrum* MAGs (Supplementary Materials Fig S2) and work in consort with *Ochrobactrum* BC to metabolize flurbiprofen.

Further MG-RAST analyses which included not only the binned sequences of the *Ochrobactrum* MAGs, but considered all sequencing reads, found that *Ochrobactrum* was the most abundant genus in the metagenomes (89.8% and 90.2% for BC and PZ, respectively). As a process control, we compared this with a genome sequenced in the same run from a *Pseudomonas putida* isolate, IPL5. More than 98% of the ORFs in the assembled genome of the IPL5 isolate was annotated as belonging to the genus *Pseudomonas*. While only 0.03% of the ORFs in the BC and PZ metagenomes were annotated as belonging to the genus *Phenylobacterium*, though this was 10X higher than what was detected in the *P. putida* IPL5 sequencing library (0.003%). Thus, despite of its low abundance, we cannot completely exclude the possibility that a *Phenylobacterium* strain could have had a functional role in flurbiprofen metabolism. We were not able, however, to recover *Phenylobacterium* isolates from our frozen stocks to further evaluate this hypothesis.

On the other hand, the MG-RAST analysis of unassembled sequencing reads from both BC and PZ (MG-RAST accession number: 4838500.3 and 4838501.3) allowed us to identify genes encoding putative biphenyl-like degrading enzymes in the BC and PZ metagenomes which were not part of the *Ochrobactrum* MAGs' binned sequences (Supplementary Materials Table S2). A BphC-like dioxygenase and BphD-like hydrolase that would be responsible for the observed ring-cleavage phenotype and benzoic acid-derivative FCB, respectively, were both found in the BC and PZ metagenomes, but at very low abundances (<0.001% of the total annotations). These findings suggest that although the PZ metagenome apparently encodes genes involved in the downstream degradation of flurbiprofen, the lack of *benR*-like in PZ itself, would prevent the production of the DHF that could otherwise be metabolized by a less abundant members of the consortium to the ring fission product.

Furthermore, among the MG-RAST results there were also annotations related to other biphenyl-degraders found in the BC metagenome, but not in PZ such as a BphC1-like dioxygenase (biphenyl-2,3-diol 1,2-dioxygenase), which has been well-characterized in *Rhodococcus globerulus* (McKay et al. 2003). Therefore, future work should focus on the transcriptional profiles of BC and PZ when exposed to flurbiprofen. Through RT-qPCR targeting specifically the transcripts of the putative *bph*-like genes found in this study (e.g. *ofdA, bphC, bphC1*, and *bphD*), it would be possible to investigate their regulation and contribution to flurbiprofen degradation. For instance, transcription of *bphC-, bphC1-*, and *bphD-* like genes should be confirmed in BC when exposed to flurbiprofen, as well as whether they are induced by BenR. Likewise, once *benR* is expressed in PZ, transcripts of *bphC* and *bphD* would be expected. Furthermore, genomic constructs carrying *bphC-,* or *bphC1-* like genes, or both would help to determine whether both or only one of them would be responsible for the production of the ring-fission metabolite observed in BC.

Numerous genes related to plasmid replication and conjugation were also reported in the metagenomic reads by MG-RAST (MG-RAST accession number: 4838500.3 and 4838501.3). Thus, it is possible that the presumptive contamination registered previously by CheckM and KAIJU (Supplementary Materials Fig S2) would correspond to genes acquired recently by *Ochrobactrum* via HGT (horizontal gene transfer). Then, the putative *bph*-like genes could be located in a plasmid recently acquired from a *Phenylobacterium* strain. The relative abundance of genes related to plasmid replication like *repA* was, however, 100X more abundant than *bph* genes like *bphC* (MG-RAST accession number: 4838500.3 and 4838501.3). Hence, our hypothesis about *bph*-like genes transferred via plasmid indeed requires to be tested using bioinformatic tools that allow the reconstruction of plasmids from metagenomic reads such as metaplasmidSPAdes (Antipov et al., 2019). This type of bioinformatic analysis, however, is still in development and the parameters to identify novel traits from poorly characterized metagenomes need to be proven. Recovery and isolation of plasmids from our frozen stocks would be an additional approach to evaluate our hypothesis. Thus, the presence of *bph*-like genes would be checked by PCR, and their expression in other bacterial strains related to *Ochrobactrum* or *Phenylobacterium* could be evaluated via plasmid transformation if some could be recovered intact, though the likelihood of that seems low.

Taken together this work highlights some of the challenges facing bioinformatic approaches to developing hypotheses about novel traits in poorly characterized genomes. The ready availability of easy-to-use genome assembly and annotation pipelines such as K-Base, RAST, and MG-RAST can lull the uninitiated into accepting the outputs as unequivocal statements of facts, when in reality they are merely predictions based on oft-overlooked assumptions including relatively high amino acid similarity cutoffs of 60%. Previous work in the Hay lab demonstrated that enzymes responsible for the degradation of aromatic compounds often

had less than 40% similarity to proteins performing analogous functions on structure related but distinct compound (Kagle et al., 2015; Murdoch and Hay, 2013; Porter and Hay, 2007). These platforms also do a surprisingly poor job of dealing plasmids which has only recently begun to be fully appreciated and necessitated the development of new bioinformatic tools (Antipov et al., 2019).

4.4. CONCLUSIONS

Metabolites detected via GC/MS analysis in our enrichment culture demonstrated that flurbiprofen was metabolized via a biphenyl-like pathway (Fig 3), but isolation of a single flurbiprofen-degrader was challenging. Through the use of metagenomic tools we were able to identify an initial, but unstable, *Phenylobacterium* strain, and an *Ochrobactrum* strain named BC. The later gave rise to a derivative, PZ, that lost the ring-fission phenotype and its ability to grow solely on flurbiprofen.

Despite the metabolite data, genomic analysis of BC and PZ did not provide evidence for genes similar to those encoding biphenyl degradation. Instead, both *Ochrobactrum* MAGs (BC and PZ) carried a novel aromatic-dioxygenase potentially involved in the initial oxidation of flurbiprofen (OfdA: putative flurbiprofen-dioxygenase). Whereas genes encoding putative BphC-like and BphD-like enzymes that would be needed for ring-cleavage and further metabolism of flurbiprofen were present in the metagenomic reads but did not assemble with the *Ochrobactrum* MAGs. This evidence along with the contamination in our metagenomic data suggests that flurbiprofen was metabolized by a bacterial consortium instead of a single pure culture as initially thought. Unfortunately, we were not able to reconstruct additional MAGs

besides *Ochrobactrum* BC and PZ, or recover other strains from our frozen stocks though additional sequencing could help on this front.

Given all these uncertainties and the difficulty of generating knockouts in environmental isolates, the simplest way forward would be to first check expression levels of *ofdA* in BC and PZ. It they are lower in PZ as expected, then the next step would be cloning *benR* from BC and expressing it in PZ which we hypothesize should enable PZ to grow on flurbiprofen and restore the yellow ring-cleavage phenotype. BenR could also be expressed in trans with a fosmid containing *ofdAB* if it could be detected using PCR to screen pooled library fractions. Alternatively, it should be possible to heterologously express *ofdAB* under the control of an inducible promoter. The efforts that will be required to confirm or refute these hypotheses serve as a sobering reminder that despite the power of genomic technologies, traditional biochemical and genetic techniques are essential for full characterizing the vast novel metabolic potential encoded by microorganisms.

4.5. REFERENCES

Abdel-Aziz, A.A.-M., Al-Badr, A.A., Hafez, G.A., 2012. Chapter 4 - Flurbiprofen, in: Brittain, H.G. (Ed.), Profiles of Drug Substances, Excipients and Related Methodology. Academic Press, pp. 113–181. https://doi.org/10.1016/B978-0-12-397220-0.00004-0

Amadio, J., Gordon, K., Murphy, C.D., 2010. Biotransformation of Flurbiprofen by Cunninghamella Species. Appl. Environ. Microbiol. 76, 6299–6303. https://doi.org/10.1128/AEM.01027-10

Andreozzi, R., Raffaele, M., Nicklas, P., 2003. Pharmaceuticals in STP effluents and their solar photodegradation in aquatic environment. Chemosphere 50, 1319–1330. https://doi.org/10.1016/S0045-6535(02)00769-5

Antipov, D., Raiko, M., Lapidus, A., Pevzner, P.A., 2019. Plasmid detection and assembly in genomic and metagenomic data sets. Genome Res. 29, 961-968. https://doi.org/10.1101/gr.241299.118

Arkin, A.P., Cottingham, R.W., Henry, C.S., Harris, N.L., Stevens, R.L., Maslov, S., Dehal, P., Ware, D., et al., 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat. Biotechnol. 36, 566–569. https://doi.org/10.1038/nbt.4163

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Edwards, R.A., Gerdes, S., Glass, E.M., Kubal, M., et al., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 9, 75. https://doi.org/ 10.1186/1471-2164-9-75

Bendz, D., Paxéus, N.A., Ginn, T.R., Loge, F.J., 2005. Occurrence and fate of pharmaceutically active compounds in the environment, a case study: Höje River in Sweden. J. Hazard. Mater., Pharmaceuticals in the Environment 122, 195–204. https://doi.org/10.1016/j.jhazmat.2005.03.012

Bickley, L.K., van Aerle, R., Brown, A.R., Hargreaves, A., Huby, R., Cammack, V., Jackson, R., Santos, E.M., Tyler, C.R., 2017. Bioavailability and Kidney Responses to Diclofenac in the Fathead Minnow (Pimephales promelas). Environ. Sci. Technol. 51, 1764–1774. https://doi.org/10.1021/acs.est.6b05079

Boyd, T., 1994. Indentification and quantification of mono-, di- and trihydroxybenzenes (phenols) at trace concentrations in seawater by aqueous acetylation and gas chromatographic-mass spectrometric analysis. J. Chromatogr. A 662, 281–292. https://doi.org/10.1016/0021-9673(94)80516-4

Camacho-Muñoz, D., Martín, J., Santos, J.L., Aparicio, I., Alonso, E., 2012. Effectiveness of Conventional and Low-Cost Wastewater Treatments in the Removal of Pharmaceutically Active Compounds. Water. Air. Soil Pollut. 223, 2611–2621. https://doi.org/10.1007/s11270-011-1053-9

Caviglioli, G., Valeria, P., Brunella, P., Sergio, C., Attilia, A., Gaetano, B., 2002. Identification of degradation products of ibuprofen arising from oxidative and thermal treatments. J. Pharm. Biomed. Anal. 30, 499–509.

Cowles, C.E., Nichols, N.N., Harwood, C.S., 2000. BenR, a XylS homologue, regulates three different pathways of aromatic acid degradation in *Pseudomonas putida*. J. Bacteriol. 182, 6339-6346. https://doi.org/10.1128/jb.182.22.6339-6346.2000

Cycoń, M., Borymski, S., Żołnierczyk, B., Piotrowska-Seget, Z., 2016. Variable Effects of Non-steroidal Anti-inflammatory Drugs (NSAIDs) on Selected Biochemical Processes Mediated by Soil Microorganisms. Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.01969

Davies, N.M., 1995. Clinical Pharmacokinetics of Flurbiprofen and its Enantiomers. Clin. Pharmacokinet. 28, 100–114. https://doi.org/10.2165/00003088-199528020-00002

de Voogt, P., Janex-Habibi, M.-L., Sacher, F., Puijker, L., Mons, M., 2009. Development of a common priority list of pharmaceuticals relevant for the water cycle. Water Sci. Technol. J. Int. Assoc. Water Pollut. Res. 59, 39–46. https://doi.org/10.2166/wst.2009.764

Domaradzka, D., Guzik, U., Wojcieszyńska, D., 2015. Biodegradation and biotransformation of polycyclic non-steroidal anti-inflammatory drugs. Rev. Environ. Sci. Biotechnol. 14, 229–239. https://doi.org/10.1007/s11157-015-9364-8

Dorn, E., Knackmuss, H.J., 1978. Chemical structure and biodegradability of halogenated aromatic compounds. Substituent effects on 1,2-dioxygenation of catechol. Biochem. J. 174, 85–94. https://doi.org/10.1042/bj1740085

Duncan, K., Uwimpuhwe, H., Czibere, A., Sarkar, D., Libermann, T.A., Fisher, P.B., Zerbini, L.F., 2012. NSAIDs induce apoptosis in nonproliferating ovarian cancer cells and inhibit tumor growth in vivo. IUBMB Life 64, 636–643. https://doi.org/10.1002/iub.1035

Eberspächer, J., 2015. Phenylobacterium, in: Bergey's Manual of Systematics of Archaea and Bacteria. American Cancer Society, pp. 1–12. https://doi.org/10.1002/9781118960608.gbm00793

Engesser, K.H., Cain, R.B., Knackmuss, H.J., 1988. Bacterial metabolism of side chain fluorinated aromatics: cometabolism of 3-trifluoromethyl(TFM)-benzoate by

Pseudomonas putida (arvilla) mt-2 and Rhodococcus rubropertinctus N657. Arch. Microbiol. 149, 188–197.
Fritz, J.S., Schenk, G.H., 1959. Acid-Catalyzed Acetylation of Organic Hydroxyl Groups. Anal. Chem. 31, 1808–1812. https://doi.org/10.1021/ac60155a034
Furukawa, K., Suenaga, H., Goto, M., 2004. Biphenyl Dioxygenases: Functional Versatilities and Directed Evolution. J. Bacteriol. 186, 5189–5196. https://doi.org/10.1128/JB.186.16.5189-5196.2004
Gaye, B., Adejare, A., 2009. Fluorinated molecules in the diagnosis and treatment of neurodegenerative diseases. Future Med. Chem. 1, 821–833. https://doi.org/10.4155/fmc.09.85
Gentili, A., 2007. Determination of non-steroidal anti-inflammatory drugs in environmental samples by chromatographic and electrophoretic techniques. Anal. Bioanal. Chem. 387, 1185–1202. https://doi.org/10.1007/s00216-006-0821-7
Gierse, J.K., Koboldt, C.M., Walker, M.C., Seibert, K., Isakson, P.C., 1999. Kinetic basis for selective inhibition of cyclo-oxygenases. Biochem. J. 339 (Pt 3), 607–614.
Gonzalez-Rey, M., Bebianno, M.J., 2014. Effects of non-steroidal anti-inflammatory drug (NSAID) diclofenac exposure in mussel Mytilus galloprovincialis. Aquat. Toxicol. 148, 221–230. https://doi.org/10.1016/j.aquatox.2014.01.011
Harayama, S., Rekik, M., 1989. Bacterial aromatic ring-cleavage enzymes are classified into two different gene families. J. Biol. Chem. 264, 15328–33.
Hay, A.G., Dees, P.M., Sayler, G.S., 2001. Growth of a bacterial consortium on triclosan. FEMS Microbiol. Ecol. 36, 105–112.
Hosoi, T., Suyama, Y., Kayano, T., Ozawa, K., 2016. Flurbiprofen Ameliorates Glucose Deprivation-Induced Leptin Resistance. Front. Pharmacol. 7. https://doi.org/10.3389/fphar.2016.00354
Hughes, D., Clark, B.R., Murphy, C.D., 2011. Biodegradation of polyfluorinated biphenyl in bacteria. Biodegradation 22, 741–9. https://doi.org/10.1007/s10532-010-9411-7
Itoh, N., Tao, H., Ibusuki, T., 2005. Optimization of aqueous acetylation for determination of hydroxy polycyclic aromatic hydrocarbons in water by stir bar sorptive extraction and thermal desorption–gas chromatography–mass spectrometry. Anal. Chim. Acta 535, 243–250. https://doi.org/10.1016/j.aca.2004.12.002
Kagle, J.M., Paxson, C., Johnstone, P., Hay, A.G., 2015. Identification of a gene cluster associated with triclosan catabolism. Biodegradation 26, 235–246. https://doi.org/10.1007/s10532-015-9730-9
Keegan, K.P., Glass, E.M., Meyer, F., 2016. MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function, in: Martin, F., Uroz, S. (Eds.), Microbial Environmental Genomics (MEG), Methods in Molecular Biology. Springer New York, New York, NY, pp. 207–233. https://doi.org/10.1007/978-1-4939-3369-3_13
Khetan, S.K., Collins, T.J., 2007. Human Pharmaceuticals in the Aquatic Environment: A Challenge to Green Chemistry. Chem. Rev. 107, 2319–2364. https://doi.org/10.1021/cr020441w
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinforma. Oxf. Engl. 23, 2947–2948. https://doi.org/10.1093/bioinformatics/btm404
Lingens, F., Blecher, R., Blecher, H., Blobel, F., Eberspächer, J., Fröhner, C., Görisch, Helma, Görisch, Helmut, Layh, G., 1985. Phenylobacterium immobile gen. nov., sp. nov., a

Gram-Negative Bacterium That Degrades the Herbicide Chloridazon. Int. J. Syst. Evol. Microbiol. 35, 26–39. https://doi.org/10.1099/00207713-35-1-26
Mars, A.E., Kasberg, T., Kaschabek, S.R., van Agteren, M.H., Janssen, D.B., Reineke, W., 1997. Microbial degradation of chloroaromatics: use of the meta-cleavage pathway for mineralization of chlorobenzene. J. Bacteriol. 179, 4530–4537. https://doi.org/10.1128/jb.179.14.4530-4537.1997
McKay, D.B., Prucha, M., Reineke, W., Timmis, K.N., Pieper, D.H., 2003. Substrate specificity and expression of three 2,3-dihydroxybiphenyl 1,2-dioxygenases from *Rhodococcus globerulus* strain P6. J. Bacteriol. 185, 2944-2951. https://doi.org/10.1128/JB.185.9.2944–2951.2003
Mezzelani, M., Gorbi, S., Fattorini, D., d'Errico, G., Consolandi, G., Milan, M., Bargelloni, L., Regoli, F., 2018. Long-term exposure of Mytilus galloprovincialis to diclofenac, Ibuprofen and Ketoprofen: Insights into bioavailability, biomarkers and transcriptomic changes. Chemosphere 198, 238–248. https://doi.org/10.1016/j.chemosphere.2018.01.148
Mondello, F.J., 1989. Cloning and expression in Escherichia coli of Pseudomonas strain LB400 genes encoding polychlorinated biphenyl degradation. J. Bacteriol. 171, 1725–1732. https://doi.org/10.1128/jb.171.3.1725-1732.1989
Moody, J.D., Freeman, J.P., Fu, P.P., Cerniglia, C.E., 2002. Biotransformation of Mirtazapine by Cunninghamella Elegans. Drug Metab. Dispos. 30, 1274–1279. https://doi.org/10.1124/dmd.30.11.1274
Murdoch, R.W., Hay, A.G., 2013. Genetic and chemical characterization of ibuprofen degradation by Sphingomonas Ibu-2. Microbiology 159, 621–632. https://doi.org/10.1099/mic.0.062273-0
Murdoch, R.W., Hay, A.G., 2005. Formation of Catechols via Removal of Acid Side Chains from Ibuprofen and Related Aromatic Acids. Appl. Environ. Microbiol. 71, 6121–6125. https://doi.org/10.1128/AEM.71.10.6121-6125.2005
Murphy, C.D., 2016. Microbial degradation of fluorinated drugs: biochemical pathways, impacts on the environment and potential applications. Appl. Microbiol. Biotechnol. 100, 2617–2627. https://doi.org/10.1007/s00253-016-7304-3
Murphy, C.D., Quirke, S., Balogun, O., 2008. Degradation of fluorobiphenyl by Pseudomonas pseudoalcaligenes KF707. FEMS Microbiol. Lett. 286, 45–49. https://doi.org/10.1111/j.1574-6968.2008.01243.x
Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055. https://doi.org/10.1101/gr.186072.114
Parolini, M., Binelli, A., Provini, A., 2011. Chronic effects induced by ibuprofen on the freshwater bivalve Dreissena polymorpha. Ecotoxicol. Environ. Saf. 74, 1586–1594. https://doi.org/10.1016/j.ecoenv.2011.04.025
Porter, A.W., Hay, A.G., 2007. Identification of opdA, a Gene Involved in Biodegradation of the Endocrine Disrupter Octylphenol. Appl. Environ. Microbiol. 73, 7373–7379. https://doi.org/10.1128/AEM.01478-07
Ricciotti, E., FitzGerald, G.A., 2011. Prostaglandins and Inflammation. Arterioscler. Thromb. Vasc. Biol. 31, 986–1000. https://doi.org/10.1161/ATVBAHA.110.207449
Richardson, S.D., Ternes, T.A., 2011. Water analysis: emerging contaminants and current issues. Anal. Chem. 83, 4614–4648. https://doi.org/10.1021/ac200915r
Richy, F., Rabenda, V., Mawet, A., Reginster, J.-Y., 2007. Flurbiprofen in the symptomatic management of rheumatoid arthritis: a valuable alternative. Int. J. Clin. Pract. 61, 1396–1406. https://doi.org/10.1111/j.1742-1241.2007.01452.x

Riegert, U., Heiss, G., Fischer, P., Stolz, A., 1998. Distal Cleavage of 3-Chlorocatechol by an Extradiol Dioxygenase to 3-Chloro-2-Hydroxymuconic Semialdehyde. J. Bacteriol. 180, 2849–2853.
Seeger, M., Camara, B., Hofer, B., 2001. Dehalogenation, denitration, dehydroxylation, and angular attack on substituted biphenyls and related compounds by a biphenyl dioxygenase. J Bacteriol 183, 3548–55. https://doi.org/10.1128/jb.183.12.3548-3555.2001
Seo, J.-S., Keum, Y.-S., Li, Q.X., 2009. Bacterial Degradation of Aromatic Compounds. Int. J. Environ. Res. Public. Health 6, 278–309. https://doi.org/10.3390/ijerph6010278
Shah, P., Westwell, A.D., 2007. The role of fluorine in medicinal chemistry: Review Article. J. Enzyme Inhib. Med. Chem. 22, 527–540. https://doi.org/10.1080/14756360701425014
Soloway, S., Wilen, S.H., 1952. Improved Ferric Chloride Test for Phenols. Anal. Chem. 24, 979–983. https://doi.org/10.1021/ac60066a017
Suenaga, H., Mizuta, S., Miyazaki, K., Yaoi, K., 2014. Diversity of extradiol dioxygenases in aromatic-degrading microbial community explored using both culture-dependent and culture-independent approaches. FEMS Microbiol. Ecol. 90, 367–379. https://doi.org/10.1111/1574-6941.12390
Taira, K., Hirose, J., Hayashida, S., Furukawa, K., 1992. Analysis of bph operon from the polychlorinated biphenyl-degrading strain of Pseudomonas pseudoalcaligenes KF707. J. Biol. Chem. 267, 4844–4853.
Triscari-Barberi, T., Simone, D., Calabrese, F.M., Attimonelli, M., Hahn, K.R., Amoako, K.K., Turner, R.J., Fedi, S., Zannoni, D., 2012. Genome sequence of the polychlorinated-biphenyl degrader Pseudomonas pseudoalcaligenes KF707. J. Bacteriol. 194, 4426–4427. https://doi.org/10.1128/JB.00722-12
US EPA, O., 2015. Estimation Program Interface (EPI Suite ™) Program Modifications & New Features in v4.11. URL https://www.epa.gov/tsca-screening-tools/estimation-program-interface-epi-suite-tm-program-modifications-new-features (accessed 7.14.19).
Wynne, S., Djakiew, D., 2010. NSAID Inhibition of Prostate Cancer Cell Migration Is Mediated by Nag-1 Induction via the p38 MAPK-p75NTR Pathway. Mol. Cancer Res. MCR 8, 1656–1664. https://doi.org/10.1158/1541-7786.MCR-10-0342
Yanaç, K., Murdoch, R.W., 2019. Biotransformation of the Fluorinated Nonsteroidal Anti-Inflammatory Pharmaceutical Flurbiprofen in Activated Sludge Results in Accumulation of a Recalcitrant Fluorinated Aromatic Metabolite. Glob. Chall. 3, 1800093. https://doi.org/10.1002/gch2.201800093
Yu, Z., Forster, R.J., 2005. Nucleic acid extraction, oligonucleotide probes and PCR methods, in: Makkar, H.P.S., McSweeney, C.S. (Eds.), Methods in Gut Microbial Ecology for Ruminants. Springer Netherlands, Dordrecht, pp. 81–104. https://doi.org/10.1007/1-4020-3791-0_7

SUPPLEMENTARY MATERIALS

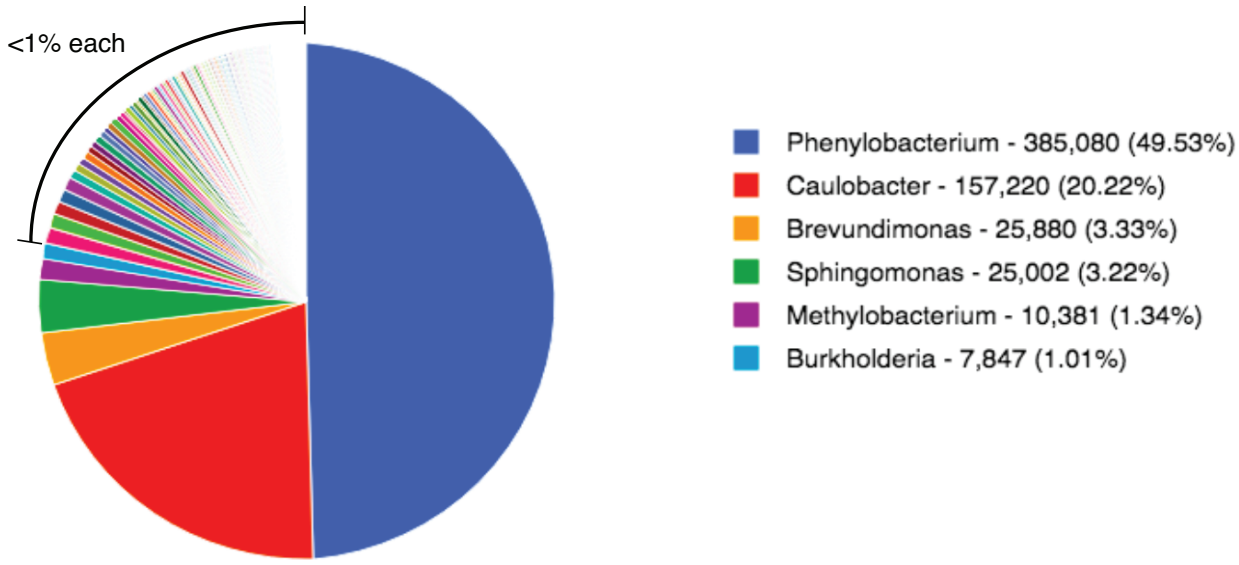
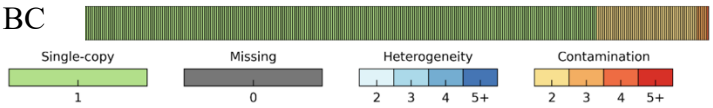
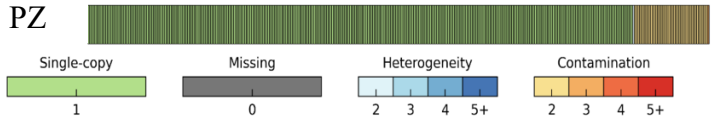


Fig S1. Bacterial consortium composition of Flur-4 sequencing data provided by MG-RAST 4.03. analysis showed that Flur-4 was not a pure culture as initially believed. To note, the genus *Ochrobactrum* represented 0.22% of the total relative abundance.

A.

	Completeness	Contamination
BC 	100%	15.223%
PZ 	100%	12.011%

B.

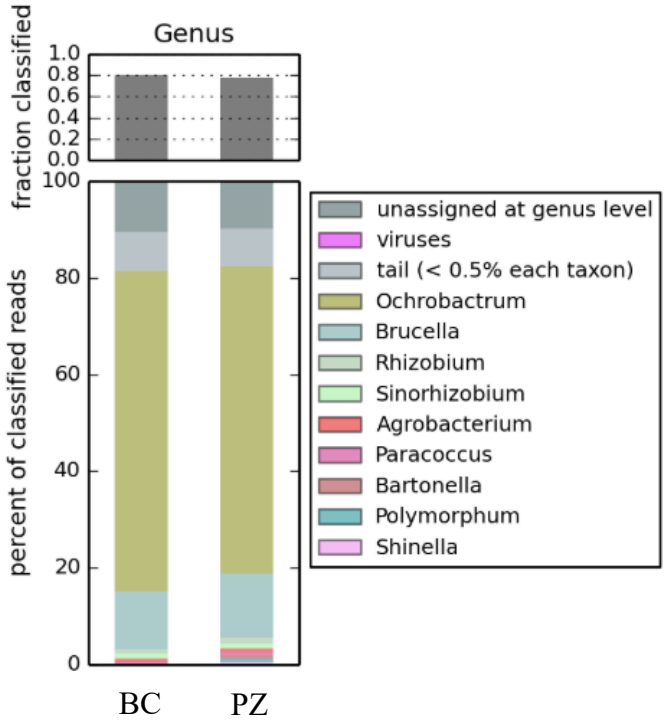


Fig S2. A. CheckM v1.0.8 analysis showed that the *Ochrobactrum* MAGs: BC and PZ were "near complete" with a contamination of 15.2 and 12.0%, respectively. **B.** Taxonomic classification based on protein sequence similarities using KAIJU v1.5.9 found that 66.2% and 63.7% of the sequences corresponded to the genus *Ochrobactrum* in BC and PZ, respectively.

Table S1. Best tBLASTn hits found in the Flur-4 genome assembly using the *bph* operon of *P. furukawaii* KF707 as query (NCBI accession number: AJMR01000119.1)

Bph enzyme	Pairwise similarity	Query coverage	Annotation in Flur-4 genome given by RAST
BphA1 biphenyl dioxygenase α-subunit	36.7%	99.5%	benzoate 1,2-dioxygenase α-subunit
BphA2 biphenyl dioxygenase β-subunit	44.8%	88.8%	3-phenylpropionate dioxygenase β-subunit
BphA3 ferredoxin	35.1%	75.5%	ferredoxin 2Fe-2S
BphA4 ferredoxin reductase	36.7%	97.5%	ferredoxin reductase
BphB dihydrodiol dehydrogenase	32.5%	90.1%	3-oxyl-acyl-carrier reductase
BphC 2,3 -dihydroxybiphenyl dioxygenase	42.9%	95.9%	2,3-dihydroxybiphenyl 1,2-dioxygenase
BphD 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid hydrolase	43.6%	99.0%	2-hydroxy-6-oxo-phenylhexa-2,4-dienoate hydrolase

Table S2. Functional annotation differences found in the metabolism of aromatic compounds between *Ochrobactrum* MAGs BC and PZ (MAG BC vs MAG PZ) resulted from whole-genome comparison by RAST 2.0. * shows the differences in biphenyl degradation found by MG-RAST 4.0.3, which not only included the binned sequences in the MAGs, but considered all the metagenome sequences from the BC and PZ cultures. X indicates the presence of the putative gene.

Subsystem	Functional Annotation	MAG BC	BC Metagenome*	MAG PZ	PZ Metagenome*
Catechol branch of beta-ketoadipate pathway	Muconate cycloisomerase (EC 5.5.1.1)	X	X		X
	Succinyl-CoA: 3-ketoacid-coenzyme A transferase subunit A (EC 2.8.3.5)		X	X	X
	Succinyl-CoA: 3-ketoacid-coenzyme A transferase subunit B (EC 2.8.3.5)		X	X	X
Benzoate degradation	benABC operon transcriptional activator BenR	X	X		X
Gentisate degradation	Putative n-hydroxybenzoate hydroxylase	X	X		X
	Maleylacetoacetate isomerase (EC 5.2.1.2)		X	X	X
Biphenyl degradation	2,3-dihydroxybiphenyl 1,2-dioxygenase		X		X
	2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase (EC 3.7.1.-)		X		X
	2-keto-4-pentenoate hydratase (EC 4.2.1.-)				X
	Acetaldehyde dehydrogenase (EC 1.2.1.10)		X		
	Biphenyl dioxygenase alpha subunit (EC 1.14.12.18)				X
	Biphenyl-2,3-diol 1,2-dioxygenase (EC 1.13.11.39)		X		
	biphenyl-2,3-diol 1,2-dioxygenase III-related protein		X		X

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

5.1. BIOCHAR DOES NOT ATTENUATE TRICLOSAN'S IMPACT ON SOIL BACTERIAL COMMUNITIES

In our study examining the combined effects of biochar and triclosan additions to soil, we found that biochar did not did not significantly impact the soil bacterial communities even though it did affect the mineralization of triclosan. The addition of triclosan, at 10 and 100 mg Kg^{-1} resulted in changes to the microbial community independent of biochar addition, though 1 mg Kg^{-1} had no discernable impact on the microbial community. More specifically, we demonstrated qualitatively and using Non-Parametric Metastats analyses, that richness did not differ between the triclosan treatments, instead, triclosan disproportionally impacted the abundance of a few shared OTUs. Importantly, our 16S rRNA gene sequencing results along with a predictive functional gene profiling allowed us to infer a concentration-dependent bacterial response with enrichment of genes encoding generalized antimicrobial resistance being the primary effect at 10 mg Kg^{-1}, and enrichment of genes encoding degradation dominating at 100 mg Kg^{-1}. Additional work using metatranscriptomics and RT-qPCR is required to evaluate and confirm our hypotheses, particularly targeting the triclosan resistance and degradation determinants found in our study like the genes related to multidrug efflux-pumps (e.g. *soxS* and *marA*), and aromatic dioxygenases with broad substrate specificity (e.g. *clcA, benA,* and *nahA*) encoding degradation. The resulting findings would help to determine the actual difference in resistance and degradation genes found in our study and provide insights into triclosan metabolism in the soil.

One of the biggest surprises from our study was the apparent sensitivity of soil *Sphingomonads* to triclosan addition, since this genus harbors well-characterized triclosan degraders. Further work indicated that the triclosan resistance determinants were rare in that bacterial group. To investigate the role of triclosan-degraders in soil, and whether there is an effect on the soil bacterial structure, future work could involve soil inoculation of *Sphingomonas* RD1, a well-characterized triclosan degrader isolated and studied in the Hay lab (Hay *et al.* 2001, Kagle et al., 2015).

Our work also reported the presence of *Stenotrophomonas* among the genera enriched by triclosan addition. Previous studies have shown that *Stenotrophomonas maltophilia*, an opportunistic human pathogen, overexpressed a multidrug-efflux pump after triclosan exposure (Sanchez et al., 2005). Moreover, stable isotope probing has shown that *Stenotrophomonas* can use triclosan as carbon source (Lee et al., 2014; Lolas et al., 2012), but a triclosan-degraders from this genus have not been identified yet. Thus, future work should address the characterization of triclosan resistance and/or degradation by *Stenotrophomonas* isolates from soil using approaches such as direct isolation, metagenomics, metatranscriptomics, as well as, biochemical techniques like stable isotope probing and mineralization assays.

5.2. SEX-DEPENDENT DISTURBANCE EFFECT OF GLYPHOSATE ON GUT MICROBIOME

Analysis of mouse gut microbiota via 16S rRNA gene sequencing reveal that the glyphosate-based herbicide (GBH) used in our study (GlyStar® Plus) only affected male mice. The low GBH-dose used (~21 ug/Kg of glyphosate) significantly impacted the bacterial diversity and richness at the end of our study (8 weeks of GBH exposure). Surprisingly, the high GBH-dose

(~210 ug/Kg of glyphosate) did not show any impact on gut bacterial communities. To further elucidate the impact of GBH, we analyzed shifts in the microbiome of individual mice to control for intrinsic interindividual variability of the mouse gut microbiota. By doing so we were able to evaluate the bacterial community shifts within individual mice over time. Our analysis showed expected microbiome changes as control mice matured (no GBH addition), but not in the low-GBH dose. These findings suggest that low GBH exposures might have an impact on gut microbiome maturation, which has not been previously considered. The concentrations we used in this study were based on drinking water standards. At the currently established maximum contaminated level (MCL) of glyphosate which is 0.7 mg/L (US EPA, 2015), a 70 Kg human consuming 2L/day of water would have ingested the same glyphosate dose (~20 ug/Kg) as the mice in our low GBH exposure. Given our results, which admittedly included a small sample size and which ran for a short duration, further work seems warranted. We would recommend additional animal studies with larger sample sizes of male mice exposed to both glyphosate and GBH doses around or below the MCL. With a larger sample size and additional treatments, it will be possible to better control for intrinsic individual variation, and also determine if the effect we observed was due to glyphosate or the other compounds in the GBH formulation we tested.

Interestingly, most of the differentially affected bacterial taxa belonged to the families *Lachnospiraceae* and *Ruminococcaceae.* Considering that glyphosate's effect has been reported to be species- and even strain-specific (Aristilde et al., 2017), and that *Lachnospiraceae* and *Ruminococcaceae* harbor commensal bacteria with important roles in gut health (Koenig et al., 2011; Penders et al., 2007), additional work is required to identify specifically the bacterial taxa impacted by the low-GBH used in our study. Future work using PCR with specific primers for these taxonomic families coupled with single nucleotide polymorphism (SNP) would allow this

specific identification. Future glyphosate-susceptibility assays with several members of these families should also be considered.

5.3. CHARACTERIZATION OF A FLURBIPROFEN-DEGRADING BACTERIAL CONSORTIUM

Though metabolite data obtained via GC/MS analysis, we demonstrated that a flurbiprofen-degrading enrichment could grow on Flurbiprofen as a sole source of carbon using a biphenyl-like degradation pathway. Finding and characterizing biphenyl-like degrading genes (*bph*-like genes) in the metagenomic reads, however, proved to be challenging. The *Ochrobactrum* MAGs (BC and PZ) that were reconstructed from genomes of the only isolates (*Ochrobactrum* BC and PZ) we were able to recover from our frozen stocks, did not encode an intact biphenyl-like degrading operon. Both MAGs, however, carried genes for a putative flurbiprofen dioxygenase (*ofdAB*), which might be involved in the initial oxidation of the non-halogenated ring of flurbiprofen. The metagenomic tools used in this study allowed for whole-genome comparisons of the *Ochrobactrum* MAGs BC and PZ. PZ was a colony variant that did not show the ring-fission phenotype and did not grow on sole flurbiprofen. A putative BenR transcriptional activator was found to be encoded in BC, but not in PZ, which might explain why PZ lost the ability to degrade flurbiprofen. Future work should address the putative function of OfdAB in flurbiprofen metabolism and the potential role of BenR in expression of *ofdAB.* RT-qPCR should be used to determine if expression levels of *ofdA* are different in BC and PZ exposed to flurbiprofen. If so, then *benR* could be cloned from BC ($benR_{BC}$) and expressed into PZ. Restoration of the ring-fission phenotype in PZ would be expected.

Genes encoding putative BphC-like and BphD-like enzymes that would be responsible for further flurbiprofen metabolism including ring-cleavage, were found in metagenomic reads that did not assemble with the *Ochrobactrum* MAGs. Thus, along with the contamination detected by CheckM and KAIJU in the metagenomic reads, our findings suggests that flurbiprofen is likely degraded by bacteria working in consort instead of by a single organism. Though no other strains besides *Ochrobactrum* BC and PZ were recovered from our frozen stocks that could grow on fluribuprofen on plate, future additional sequencing with deep coverage would help to reconstruct additional MAGs and assess our hypothesis. Additional RT-qPCR assays with genomic constructs assessing the putative *bph*-like genes found in this study (e.g. *bphC* and *bphD*) will also contribute to understanding flurbiprofen metabolism by BC, as well as whether they are induced by BenR. A similar approach could also be used to screen for flurbipofen activity in the fosmid libary of BC since, based on genomic analysis, it is likely that no fosmid encoded both *benR* and *ofdAB*.

While our analyses also showed the presence of plasmid-related genes in the metagenomic reads that did not assemble with the *Ochrobactrum* MAGs, their disproportional abundance in comparison to the *bph*-like genes suggested that it is unlikely that *Ochrobactrum* would have been acquired *bph*-like genes via horizontal gene transfer (HGT). Reconstruction of plasmids from metagenomic data using tools as metaplasmidSPAdes or plasmid isolation and sequencing from our frozen stocks would be, however, additional approaches to test this hypothesis.

5.4. FINAL CONSIDERATIONS

Taken together, my work on these biologically active molecules (triclosan, glyphosate, flurbiprofen) in different environments (soil, gut, sewage) has illustrated some of the challenges involved in investigating the roles of environmental bacteria that are highly diverse and involved in complex inter- and intra- species interactions. Next generation sequencing (NGS) is an excellent tool for helping to characterize such bacterial communities mainly because of its capacity to gathering tremendous amount of information about bacterial taxonomy, diversity, and even functionality in a single step: a sequencing run. Since traditional culturing and isolation is a tedious and often unsuccessful approach, NGS has rapidly replaced the traditional bench work. Increasingly, the result of such environmental studies is merely in a list of bacterial species and putative functions. The challenge is not only to process and interpret the sequencing data, but to move beyond correlations among bacterial species abundance and putative functions in the environment, but to prove causation, especially in the characterization of novel traits.

Through the analysis of sequence data from three cases studies, my work provides criteria and guidelines to use some of the bioinformatic tools currently available to investigate and dissect the complex relationships of bacteria with the environment. A dose-dependent effect, but with different responses, such as in the case of triclosan's impact leading to resistance or degradation depending on the dose. A gender-dependent effect on selected bacteria taxa, such as glyphosate treatment with higher unique OTUs. And, a cooperative function as in the presumptive flurbiprofen-degrading consortium. While many questions yet remain, the results of these sequence-informed studies lead, however, to hypotheses about causality and inferences that need to be confirmed and tested by the traditional biochemical and genetic approaches, which are techniques that remain essential to fully characterize and understand bacterial communities.

5.5. REFERENCES

Aristilde, L., Reed, M.L., Wilkes, R.A., Youngster, T., Kukurugya, M.A., Katz, V., Sasaki, C.R.S., 2017. Glyphosate-Induced Specific and Widespread Perturbations in the Metabolome of Soil Pseudomonas Species. Front. Environ. Sci. 5. https://doi.org/10.3389/fenvs.2017.00034

Hay, A.G., Dees, P.M., Sayler, G.S., 2001. Growth of a bacterial consortium on triclosan. FEMS Microbiol. Ecol. 36, 105–112.

Kagle, J.M., Paxson, C., Johnstone, P., Hay, A.G., 2015. Identification of a gene cluster associated with triclosan catabolism. Biodegradation 26, 235–246. https://doi.org/10.1007/s10532-015-9730-9

Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J., Knight, R., Angenent, L.T., Ley, R.E., 2011. Succession of microbial consortia in the developing infant gut microbiome. Proc. Natl. Acad. Sci. 108, 4578–4585. https://doi.org/10.1073/pnas.1000081107

Lee, D.G., Cho, K.-C., Chu, K.-H., 2014. Identification of triclosan-degrading bacteria in a triclosan enrichment culture using stable isotope probing. Biodegradation 25, 55–65.

Lolas, I.B., Chen, X., Bester, K., Nielsen, J.L., 2012. Identification of triclosan-degrading bacteria using stable isotope probing, fluorescence in situ hybridization and microautoradiography. Microbiology 158, 2796–2804.

Penders, J., Thijs, C., van den Brandt, P.A., Kummeling, I., Snijders, B., Stelma, F., Adams, H., van Ree, R., Stobberingh, E.E., 2007. Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study. Gut 56, 661–667. https://doi.org/10.1136/gut.2006.100164

Sanchez, P., Moreno, E., Martinez, J.L., 2005. The Biocide Triclosan Selects Stenotrophomonas maltophilia Mutants That Overproduce the SmeDEF Multidrug Efflux Pump. Antimicrob. Agents Chemother. 49, 781–782. https://doi.org/10.1128/AAC.49.2.781-782.2005