Undergraduate Journal of Experimental Microbiology and Immunology (UJEMI) Vol. 27:1-14

A comparison of metabarcoding analysis between ASVs and OTUs - using data regarding the effects of chronic radiation on the bank vole gut microbiota

Cameron Inglis, Danesh Rahmani, David Yeung, Daniel Mun

Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada

SUMMARY The gut microbiome is dynamic within its host and susceptible to changes in both diversity and functionality. A variety of environmental factors can disturb the balance of the gut microbiome, including chronic environmental radiation. It has been previously documented that exposure to chronic levels of radiation can induce changes in the gut microbiome in a time-dependent manner. Bank voles from the area in and around the Chernobyl Exclusion Zone provide an ideal model for examining the effects of chronic environmental radiation on the gut microbiome. Past studies on these subjects have been conducted through Operational Taxonomic Unit (OTU) based metabarcoding analysis, which works on the basis of clustering according to sequence similarity. Notably, a newer Amplicon Sequence Variant (ASV) based method of metabarcoding analysis is rapidly increasing in popularity for conducting such analyses. This is in large part due to its capacity for specific single nucleotide differentiation and identification of microbial gene sequences at a species level. Herein, we conduct a direct comparison of these two metabarcoding analysis methods for identifying the effects of chronic radiation on the gut microbiome of bank voles. Our results suggest that while OTU and ASV metabarcoding analyses yield similar results with regard to examining microbial diversity and functionality, ASV-based analysis is able to capture slightly more detailed gut microbial diversity compared to OTU-based analysis.

INTRODUCTION

The Gut Microbiome and its Susceptibility to Change. The gut microbiome is made up of trillions of microorganisms which exist in a symbiotic relationship with their host (1). While the host provides these microorganisms with food, nutrients and a favourable environment in which they can reside, organisms of the microbiome are crucial to their host's digestion, nutrient cycling, susceptibility to disease, and immune regulation (1). Importantly, gut microbiome composition is not static; susceptible to both qualitative and quantitative changes in microbial diversity, thereby affecting the collective functionality of these host organisms (2, 3). Variations in gut microbiota composition have profound effects on their collective functional profiles and the benefits conferred to their host (1, 3). There are numerous factors which determine gut microbiota composition, including but not limited to: host diet, age, antibiotic intake, and environment (4).

Effects of Radiation on the Gut Microbiome. Although the degree to which many of these factors are able to affect gut microbiome composition have been thoroughly documented, the impacts of environmental pollution, specifically high environmental radiation, on these microorganisms is largely unknown. It has been previously reported that not only does the human gut microbiome seem to contribute to radiosensitivity, but exposure to chronic radiation also results in significant alterations of gut microbial abundance and diversity (5). Further research has shown a possible link between chronic exposure to low-dose radiation and changes in gut microbiota composition and metabolism in mice (6). At a genus level, radiation exposure has been shown to cause significant alterations to gut microbial abundance in a time dependent manner, resulting in an elevated abundance of *Clostridium, Helicobacter* and *Oscilibacter*, and a decreased abundance of *Bacteroidetes and Barnesiella* (6). These changes in gut microbiota composition were seen to induce functional

Published Online: September 2022

Citation: Cameron Inglis, Danesh Rahmani, David Yeung, Daniel Mun. 2022. A comparison of metabarcoding analysis between ASVs and OTUs - using data regarding the effects of chronic radiation on the bank vole gut microbiota. UJEMI 27:1-14

Editor: Andy An and Gara Dexter, University of British Columbia

Address correspondence to: https://jemi.microbiology.ubc.ca/

differences on biochemical pathways, especially those affecting metabolism and DNA repair, following prolonged exposure to radiation (4, 6).

Chernobyl. One of the most relevant examples of environmental radiation pollution in recent history is the Chernobyl disaster of 1986, in which ionizing radiation was released from Chernobyl at an alarming rate, resulting in considerable radioactive pollution within the immediate and surrounding regions (7). Although human populations were evacuated from those areas most severely affected by this environmental pollution, identified as the Chernobyl Exclusion Zone (CEZ), wild animals inhabiting the CEZ have been exposed to chronic radiation (7). As such, these wild animals constitute an ideal repository for examining the effects of environmental radiation on the gut microbiome.

The Dataset Used and Previous Studies. To this end, Lavrinienko *et al*. analyzed 16s rRNA sequences from fecal samples of 137 bank voles within and around the CEZ with 63 being exposed to high levels of radiation (CH), 43 being exposed to low levels of radiation (CL), and 31 being exposed to no significant levels of radiation (KL) (8). The dataset from which these analyses were made includes a number of metadata variables pertaining to the bank voles from which the fecal samples were isolated, including location. Lavrinienko *et al.* determined that while radiation exposure did not have a significant effect on the species richness of gut microbiome samples, there were significant differences in species abundance among bank voles which had been chronically exposed to high levels of radiation. Namely, the authors identified a notable negative correlation between exposure to radiation and the ratio of *Firmicutes* to *Bacteroidetes* within the vole microbiome (8). Lavrinienko *et al.* observed that chronic exposure to radionuclides greatly impacts the functional profiles of the bank vole gut microbiome, specifically those involved in the assimilation and transport of carbohydrates, xenobiotics biodegradation and DNA repair (8). Collectively, their results highlight how exposure to environmental pollution is able to cause dramatic changes in the composition and function of the gut microbiome.

OTU- vs. ASV-based Analysis. While their experimental protocol was thorough, Lavrinienko *et al.* conducted their analysis based on Operational Taxonomic Units (OTUs) (8). OTU-based classification of RNA sequences is typically done at 97% sequence similarity, by which similar genetic sequence variants are organized into clusters (9). Although OTU-based methods have traditionally been the gold standard for microbial analysis in molecular biology, it is limited in terms of detecting minute differences, in both diversity and composition, which may be present within samples. This is due in large part to the fact that clustering sequences at 97% similarity allows only for organism differentiation at the genus level (9). Conversely, ASV-based analysis provides an updated alternative for studying microbial diversity; differentiating sequence variants based on single nucleotide changes and enabling the precise identification of microbes at a species level. Herein, we aim to conduct an ASV-based diversity and functionality analysis on the vole dataset compiled by Lavrinienko *et al.*, in an attempt to both reaffirm and elaborate on the established knowledge of how exposure to chronic radiation affects the gut microbiome. We predict that the increased specificity of sequence differentiation identified by ASV-based analysis will offer greater insight into the composition and functional profiles of vole gut microbiome samples when compared to previous OTU-based studies.

METHODS AND MATERIALS

Study system. Lavrinienko *et al.* collected the fecal samples of 137 different bank voles from 64 trapping locations across three sampling sites in Northern Ukraine from May to June of 2016 (8). Each of these sampling sites covered areas of differing environmental radiation levels ranging from a mean of 30.1 µSv/h in the high radiation zone, CH, to a mean of 0.25µSv/h and 0.33µSv/h in the low radiation zone, CL, and control radiation zone, KL, respectively (8). DNA was extracted from the naturally passed fecal pellets following the Earth Microbiome Project Protocol (www.earthmicrobiome.org/protocols-and-standards), followed by the demultiplexing of raw DNA sequences and PCR amplification of 16S ribosomal RNA (16S rRNA) genes to prepare the sequence library used in our study (8, 10). The V4 variable regions of these 16S rRNA genes were amplified using the 515F/806R

primer pair, and paired-end 250 base pair reads were generated from barcoded amplicons sequenced on Illumina MiSeq by Beijing Genomics Institute, Hong Kong, China (8, 11).

Read data processing in QIIME2. The demultiplexed 16S rRNA reads were imported from the dataset and joined with the manifest file using the QIIME2 tools *import* function, treating the reads as SingleEndFastqManifestPhred33 format. These demultiplexed sequences underwent a quality control step via the QIIME2 Diverse Amplicon Denoising Algorithm 2 (DADA2) plugin (12, 13). Quality control via DADA2 eliminates sequencing errors from amplicon clusters, chimeric sequences, and sorts sequencing reads into ASVs (12, 13). During this process, sequences were truncated at 180nts to maintain a median phred quality score of 37 (12, 13). At this point our analysis diverges into analysis by two different methods of feature identification; ASV, and 97% OTU. QIIME2 steps are outlined in Script #1.

OTU assignment, rarefaction, and phylogeny. In effort to make our data comparable to the OTUs generated by Lavrinienko *et. al*, we imported the same GreenGenes 97% OTU reference database, which provides reference sequences for OTU assignment, into QIIME2 (14). The QIIME2 vsearch cluster-features-open-reference tool was used to assign OTUs to the feature table and representative sequences previously generated by DADA2 (15). This clustering tool generated two artifacts: a feature table artifact, and a representative sequences artifact of open reference frame OTUs with 97% sequence similarity. The feature table was rarefied to a depth consistent with that applied by Lavrinienko *et. al*: 18,000 sequence reads per sample (8). Conversely, the 97% OTU representative sequences artifact was used to generate a rooted phylogenetic tree, produced by applying the QIIME2 phylogeny tool; *align-to-tree-mafft-fasttree* (16, 17, 18). QIIME2 steps are outlined in Script #1.

ASV rarefaction and phylogeny. The feature table artifact generated by DADA2 was rarefied to a depth of 18,000 sequence reads per sample in order to maintain the ability to produce comparative results to Lavrinienko *et. al* and analysis by OTU. Additionally, this depth was chosen because representative sampling across the treatment zones, as well as feature richness, were seen to be maintained to the same degree as rarefaction at 18,000 for OTUs. Using the feature table and representative sequence artifacts generated by DADA2, the QIIME2 phylogeny tool *align-to-tree-mafft-fasttree* was applied again to create a rooted phylogenetic tree used in downstream diversity metric calculations (16). QIIME2 steps are outlined in Script #1.

Taxonomic classification for OTU and ASV features. For taxonomic classification, we opted for a pre-trained Naive Bayes classifier trained on Silva 138 99% OTUs from the 515F/806R region of sequences (19, 20). This classifier was applied to the representative sequence artifacts derived from both ASVs and OTUs via the QIIME2 feature-classifier *classify-sklearn* plugin (21, 22). The taxonomy artifacts generated from this plugin were used in differential abundance analysis and contain information relating the taxonomic classifications to each ASV and OTU cluster.
Lastly, the data were exported for analysis in R (v 4.1.2) (23). The data was exported in BIOM v2.1 format and processed in R. QIIME2 steps are outlined in Script #1.

Data and statistical analysis for ASVs and OTUs. Alpha diversity metrics; Faith's phylogenetic diversity (PD) and Pielou's evenness, were produced for both ASVs and OTUs using the QIIME2 diversity *core-metrics-phylogenetic* plugin (12). Pairwise Kruskal-Wallis tests were conducted to determine which metadata categories played a significant role in the differences of gut microbial diversity between treatment areas CH, CL and KL.

Beta diversity and differential abundance in R. Bray-Curtis dissimilarities and Weighted UniFrac were computed in R and clusters were identified with 95% confidence for both ASV assigned features and OTU clusters. For differential abundance analysis between treatment areas CH, CL, and KL, the relative abundances for each ASV and for each OTU were calculated. Only ASVs or OTUs at an abundance of at least 0.1% were considered in the analysis. To refine our selection to detect significantly different taxa among groups of ASVs

and OTUs, data were analyzed at a p-value threshold of 0.05 (<0.05), corresponding to a false discovery rate of 5%. Differentially abundant taxa were identified for analysis by both ASV and OTU. The results from differential abundance were further refined to identify changes in the abundance of Bacteroidetes, Firmicutes, and Proteobacteria among the abundances computed by ASV and by OTU, respectively. Lastly, the mean relative abundance of bacterial taxa at the phylum level were computed and displayed as two bar charts; results of analysis by ASV and results of analysis by OTU. RStudio steps are outlined in Script #2R and Script #3R.

Determining functionality through PICRUSt2. The functional composition of the vole gut microbiota was assessed via PICRUSt2 (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States 2) to predict functional abundances (24, 25). Using the 16S rRNA gene sequencing data, predictions for enzyme commission (EC) numbers and MetaCyc pathways were generated and subsequently interpreted via the Statistical Analysis of Metagenomic Profiles (STAMP) software (26). The predicted metagenomes were classified using Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologs and summarized using KEGG pathways at levels 2 and 3.

RESULTS

Radiation-associated differences in alpha diversity. Feature tables generated from ASV assignment by DADA2 were able to identify 6,044 unique features. When features were identified using open reference frame 97% OTU clustering, 2,955 unique features were found. Faith's phylogenetic diversity analysis revealed differences in microbial richness by sample site for both ASVs and OTUs. Both modes of analysis were able to detect significant differences in phylogenetic diversity between CH and KL ($q < 0.005$), while only analysis by ASV had sufficient resolution to detect a significant difference in richness between CL and KL ($q < 0.05$) **(Figure 1.**, **Table 1)**. The alpha diversity metric of Pielou's evenness was also

TABLE. 1 Kruskal-Wallis test results for pairwise comparison of Faith's phylogenetic diversity between treatment areas for OTU and ASV.

Group 1	Group 2	q-value by OTU	q-value by ASV
CH (n=63)	CL (n=43)	0.206115	0.248116
	KL (n=31)	0.002291	0.000407
CL (n=43)	KL (n=31)	0.185599	0.046999

calculated for analysis by ASV and by OTU - indicating that the number of microbes between sample sites did not differ significantly (**Supplemental Figure 1., Supplemental Table 1**). Taken together, these results suggested that total species richness, but not evenness, was significantly higher in analysis by ASV compared to analysis by OTU **(Figure 1)**.

Radiation-associated changes in beta diversity metrics. Two Beta diversity metrics; Bray-Curtis dissimilarities and weighted UniFrac, were calculated for grouping by ASV and by OTU. While both of these metrics consider differences in taxonomic abundance, Bray-Curtis additionally considers species richness while weighted UniFrac also considers sample relatedness (27). In both ASV and OTU-based analysis, inter-individual differences in gut microbiota communities for the bank vole samples were exhibited to a high degree, and samples from uncontaminated areas CL and KL displayed larger inter-individual variation compared to those from CH. This not only suggests that exposure to chronic high-level radiation induced predictable and particular changes to the gut microbiome, but also highlights how although ASV and OTU-based methods differ, they are both able to capture minute and important differences in microbial diversity **(Figure 2, Figure 3)**. There was found to be a substantial overlap of clustering between samples from uncontaminated areas

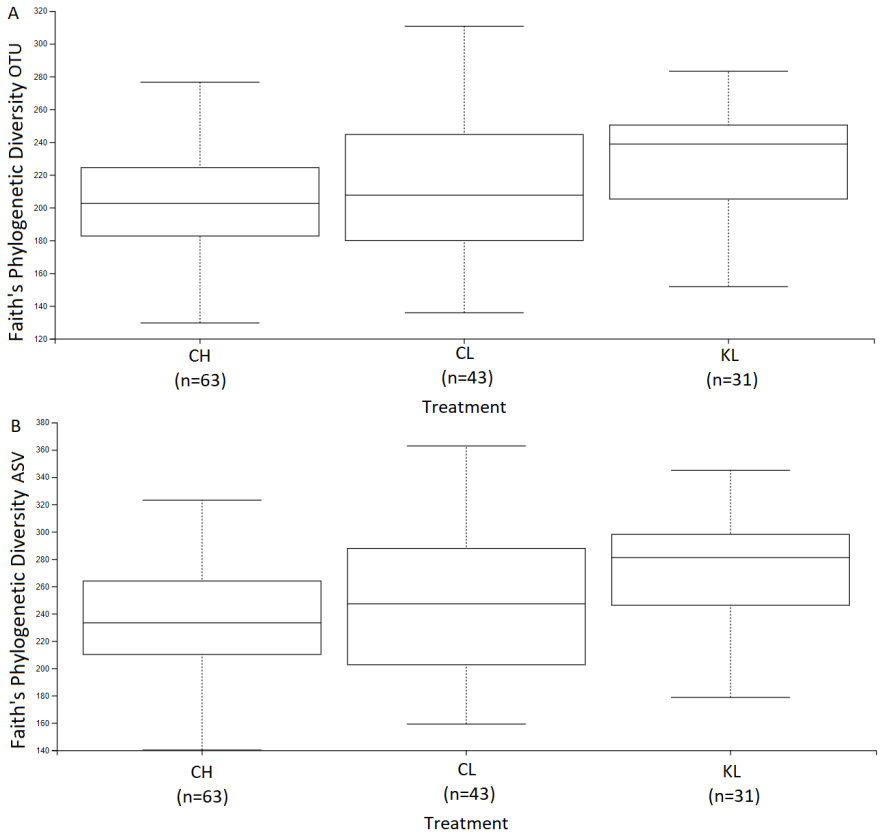

FIG. 1 Faith's phylogenetic diversity for OTUs and ASVs. Quantification of microbial alpha diversity (Faith's PD) across voles and treatment sites with error bars indicating median value ± SE. (A) Faith's PD by OTU quantified by Kruskal-Wallis tests conducted in QIIME2 in a pairwise manner between CH, CL, and CL with significant differences between CH and KL ($q = 0.002$). (B) Faith's PD by ASV quantified by Kruskal-Wallis tests conducted in QIIME2 in a pairwise manner between CH, CL, and KL with significant differences between CH and KL ($q = 0.0004$), and CL and KL ($q = 0.05$). A net decrease in Faith's PD is observed for OTUs relative to ASVs.

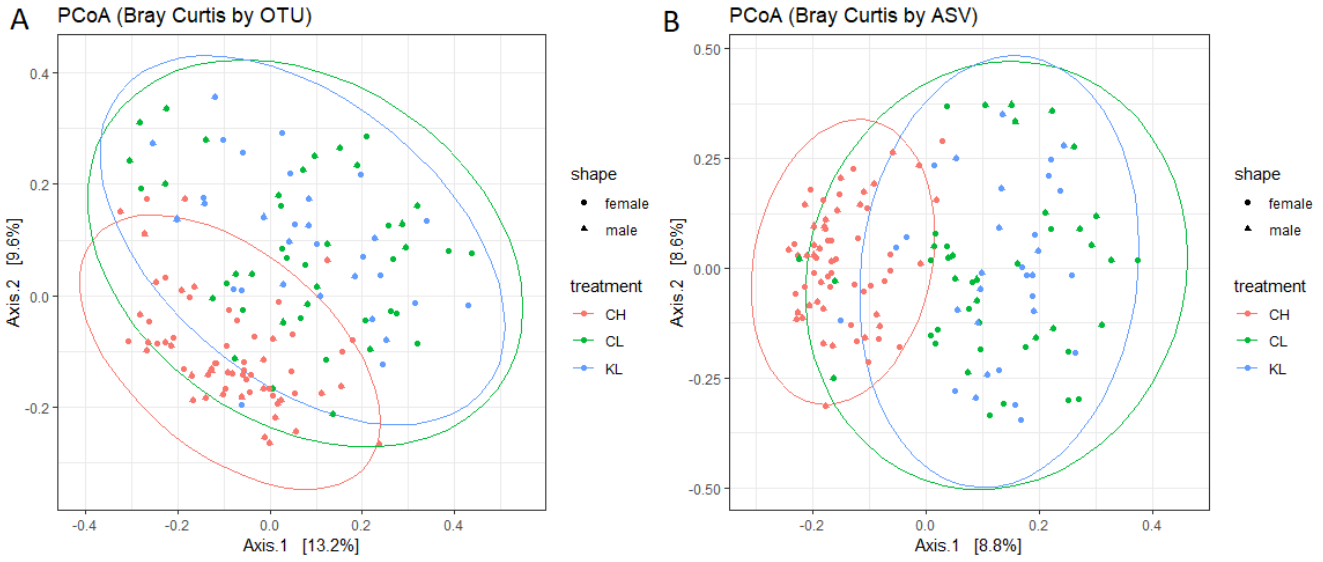

FIG. 2 Bray-Curtis dissimilarities by treatment area for OTUs and ASVs. Bray-Curtis dissimilarity distances between bank vole gut microbiota composition at a rarefaction depth of 18,000 among the three study areas, CH, red (n = 63); CL, blue (n = 43); KL, green (n = 31). Each point represents a single sample, and each shape indicates host sex coloured according to study area. Ellipses represent a 95% confidence interval around the cluster centroid. (A) Bray-Curtis dissimilarity distances from analysis by OTU showing separation between the CH cluster from the CL and KL clusters split along the first and second PCoA axes. (B) Bray-Curtis dissimilarities from analysis by ASV showing separation between the CH cluster from the CL and KL clusters along the first PCoA axis.

CL and KL across the Bray-Curtis and weighted UniFracs **(Figure 2., Figure 3)**. This pattern was seen regardless of the mode of analysis (ASV or OTU). The notable differences between

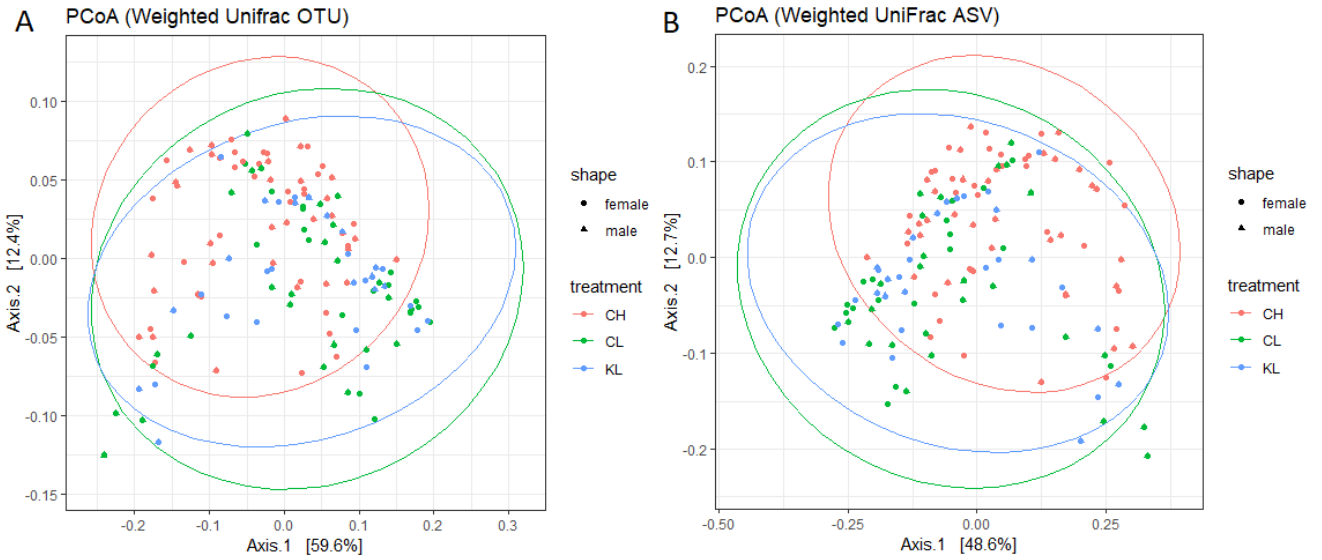

FIG. 3 Weighted UniFrac by treatment area for OTUs and ASVs. Beta diversity by Weighted UniFrac distances between bank vole gut microbiota composition at a rarefaction depth of 18,000 among the three study areas, CH, red (n = 63); CL, blue (n = 43); KL, green (n = 31). Each point represents a single sample, and each shape indicates host sex coloured according to study area. Ellipses represent a 95% confidence interval around the cluster centroid. (A) Weighted UniFrac distances from analysis by OTU showing separation between the CH cluster from the CL and KL clusters split along the first and second PCoA axes. (B) Weighted UniFrac from analysis by ASV showing separation between the CH cluster from the CL and KL clusters split along the first and second PCoA axis.

Figure 2A and **Figure 2B** are that analysis by OTU leads to the separation of CH from CL and CK clusters split between the first and second Principle Coordinate Axes (PCoA), while analysis by ASV leads to separation mainly along the first PCoA axis. With regards to weighted UniFrac, there are commonalities between the two analysis methods in that there is a nearly identical separation split between the first and second PCoA axes of the CH cluster from the CL and KL clusters **(Figure 3**). Considering beta diversity metrics across ASV and OTU analysis, differences in diversity were observed between contaminated and uncontaminated areas in that the CH treatment group shared significantly less inter-sample diversity in terms of species richness compared to CL and KL samples. Our analyses were unable to identify any significant differences through weighted UniFrac, suggesting no variations in sample relatedness across the different treatment groups, and were also unable to identify any detectable differences in taxonomic abundance when comparing the results from ASV and OTU-based analysis.

Relative abundance. Differential relative abundances were deduced for both ASV and OTU analysis, and the phyla determined to have significant ($p < 0.005$) changes in abundance were plotted (**Figure 4)**. While ASV generation in QIIME2 incorporates feature assignment, OTU-based methods require feature assignment according to a reference database. In this analysis, OTUs were assigned features according to the Greengenes 97% OTU reference database (12). As such, taxa names differ slightly in the relative abundance plots **(Figure 4).** For the purposes of this analysis, the Baceroidota and Actinobacteriota phyla are compared directly to Bacteroidetes and Actinobacteria, respectively. Notably, analysis by ASV identified an increase in the relative abundance of Proteobacteria, and a decrease in Spirochaetes - results which were not observed in the OTU differential abundance **(Figure 4)**. OTU differential abundance identified two phyla that increased significantly in abundance in contaminated areas relative to uncontaminated areas: Desulfobacterota and Campilobacterota, which were not present in abundance by ASV analysis **(Figure 4)**. The changes in abundance of key bacteria identified by Lavrinienko *et al.*; Bacteroidetes, Firmicutes, and Proteobacteria, were determined between CH, CL, and KL, and compared using analysis by ASV and by OTU. Both modes of analysis identified Bacteroidetes as having lower relative abundance in CH compared to CL and KL **(Figure 5)**. Although ASV

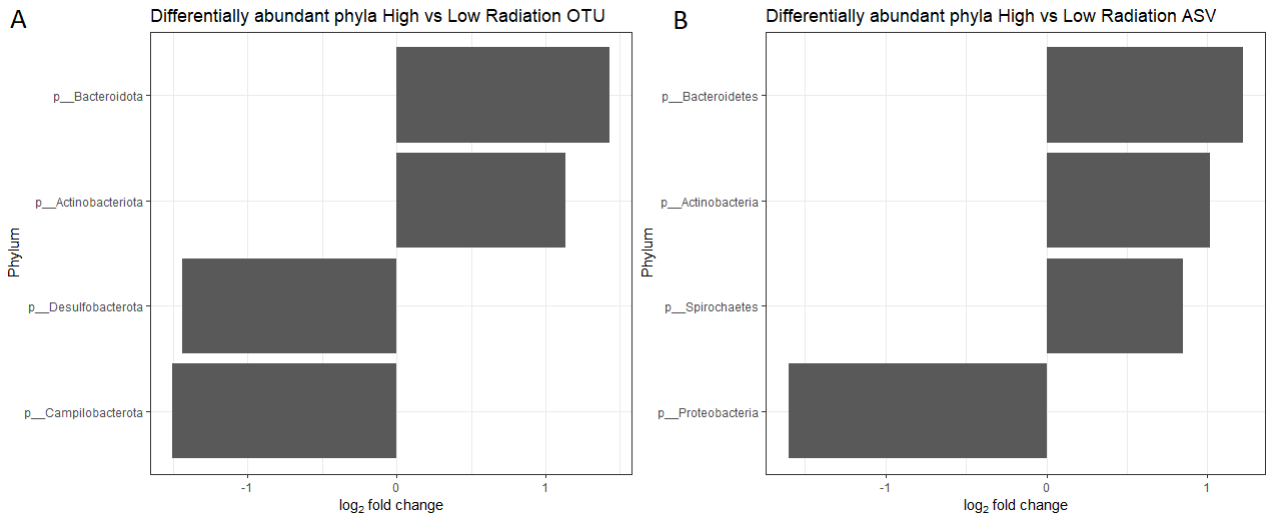

FIG. 4 Differential abundance of microbes between treatment areas for OTUs and ASVs. Differential abundance of microbes by phylum which differed significantly between treatment areas. The bars on the left represent phyla which had increased abundance in contaminated zones relative to uncontaminated zones, and the bars on the right represent phyla which had decreased abundance in contaminated zones relative to uncontaminated zones. Both plots show decreases in abundance of Bacteroidetes and Actinobacteria, whereas (A) shows increases in Desulfobacterota and Campilobacteriota, and (B) shows a decrease in Spirochaetes and an increase in Proteobacteria.

and OTU analysis reveal the same relative abundance proportions of Bacteroidetes, there is an approximate 10% decrease in the relative abundance of Bacteroidetes across all treatment sites in analysis by ASV, compared to OTU **(Figure 5A, B)**. Conversely, both ASV and OTU analysis identify Firmicutes as having higher relative abundance in CH compared to CL and KL **(Figure 5C, D)**. Again, there was an approximate 10% decrease in the overall abundance of Firmicutes across all treatment areas when the analysis was conducted by ASV compared to OTU. The most notable difference in the two methods of analysis was observed in the relative abundance of Proteobacteria. Following OTU-based analysis, the relative abundance of Proteobacteria was found to be close to zero **(Figure 5E)**. This provides a stark contrast to the results from ASV analysis, which determined Proteobacteria to be more abundant in CH than CL and KL, with median relative abundance of approximately 2.5% for the contaminated area, CH, and 1% for uncontaminated areas CL and KL **(Figure 5F)**. The mean relative abundance of the significantly abundant phyla among the treatment areas for both OTU and ASV analysis is displayed in **Figure 6**. Notably, relative abundance by OTU identified three phyla which were absent from phyla assignment by ASV: Compilobacterota, Cyanobacteria, and Desulfobacterota. Given that ASV and OTU analysis pipelines differed feature assignment, it is likely that these three phyla were incorporated into other groups during ASV phyla assignment. Nevertheless, the major difference between phyla assignment by ASV and OTU is that the relative abundance of Proteobacteria is slightly underrepresented in analysis by OTU compared to analysis by ASV (8).

Functionality of the vole gut microbiome. Predicted functionality pathways were determined for both OTU and ASV analysis. PICRUSt2 analysis determined no significant differences in the majority of functional pathways between the three treatment groups when comparing OTU- and ASV-based modes. However, multiple group analysis via ANOVA identified changes in functional pathways when comparing OTU and ASV derived feature tables. Most notably, the proportion of sequences of distinct pathways in the CH treatment were comparable between the two different PICRUSt2 modes, in which the ASV feature table analysis indicated a decrease in proportion (**Figure 7**). Three pathways, the biotin biosynthesis, acetyl CoA, and glycol metabolism and degradation pathways, exhibited distinct differences in their sequences; uncontaminated treatment groups (CL, KL) showed a similar proportion of sequences in both OTU and ASV analysis while the contaminated treatment group (CH) showed a significant decrease in proportions of sequences in ASV

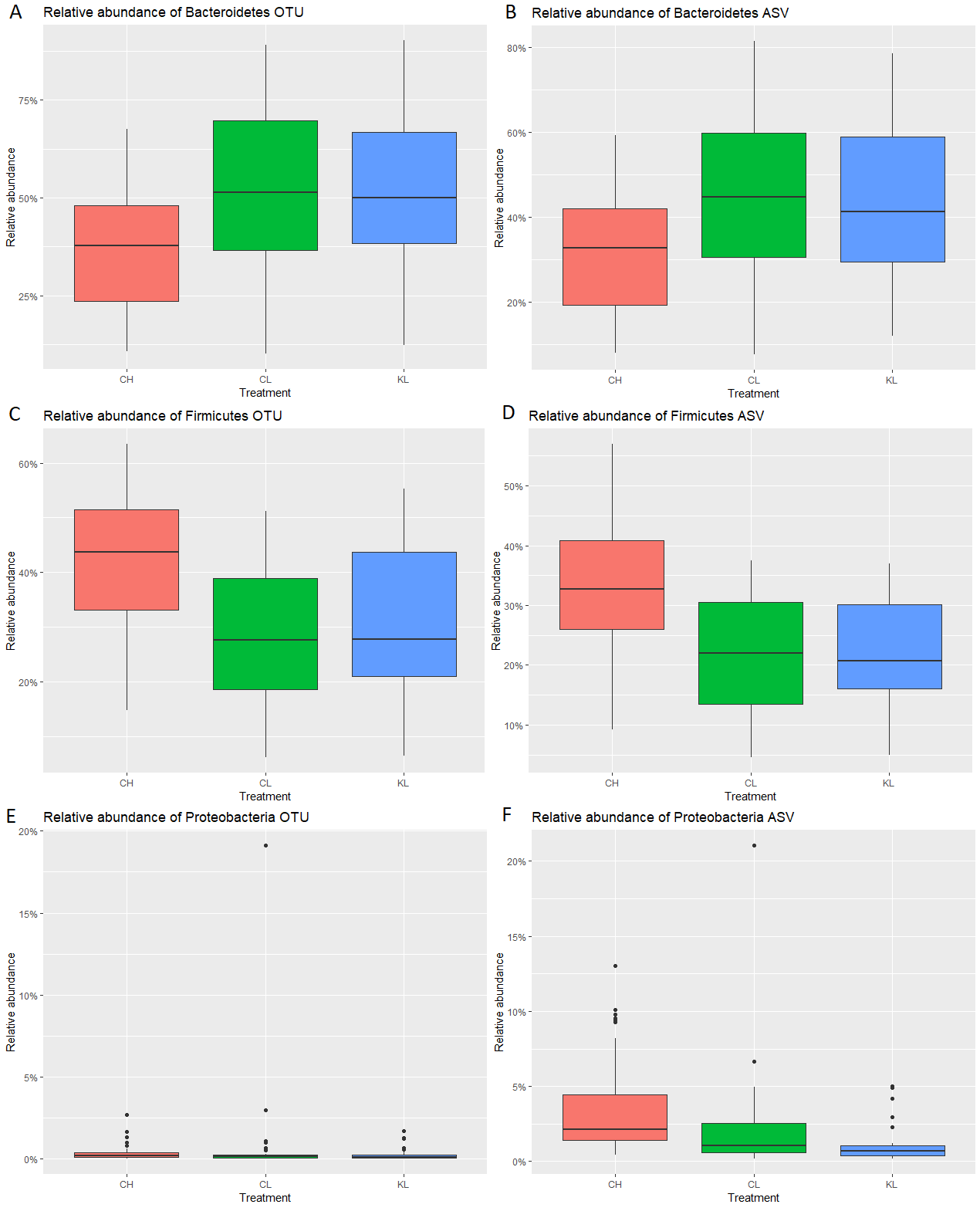

FIG. 5 Relative abundance of Bacteroidetes, Firmicutes, and Proteobacteria among treatment areas for OTUs and ASVs. Box plots representing relative abundance of Bacteroidetes, Firmicutes, and Proteobaceria, among treatment areas with error bars indicating median value ± SE. For both (A) and (B), Bacteroidetes has lower abundance in CH than CL and KL. (A) differs from (B) in that the relative abundance across all treatment areas is ~10% higher. For both (C) and (D), Firmicutes is observed in higher abundance in CH than CL and KL. (C) differs from (D) in that the relative abundance across all treatment areas is ~10-15% higher. (E) shows a lack of identification of Proteobacteria with average abundances approximating zero among all treatment areas. (F) Shows an increased relative abundance of Proteobacteria in CH.

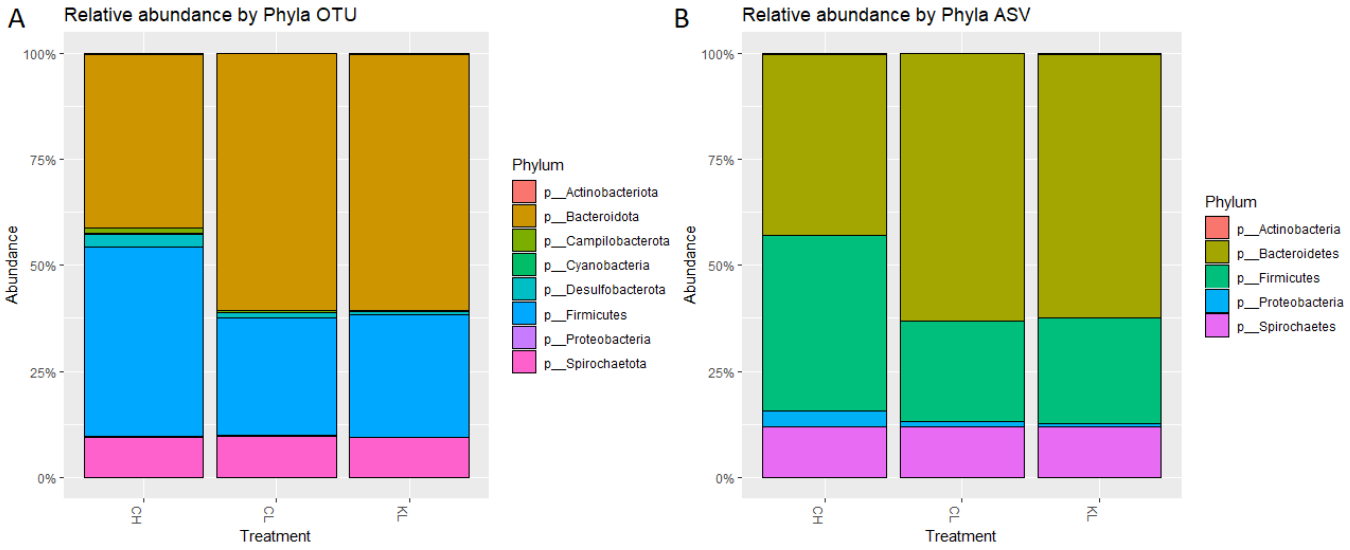

FIG. 6 Mean relative abundance of microbial taxa at the phylum level for OTUs and ASVs. (A) mean relative abundance at phyla level for analysis by OTU showing the identification and abundance of 8 different phyla. (B) mean relative abundance at the phyla level for analysis by ASV showing the identification and abundance of 5 different phyla. (A) differs from (B) in that analysis by OTU resulted in the identification of three additional phyla, Campilobactera, Cyanobacteria, and Desulfobacteria.

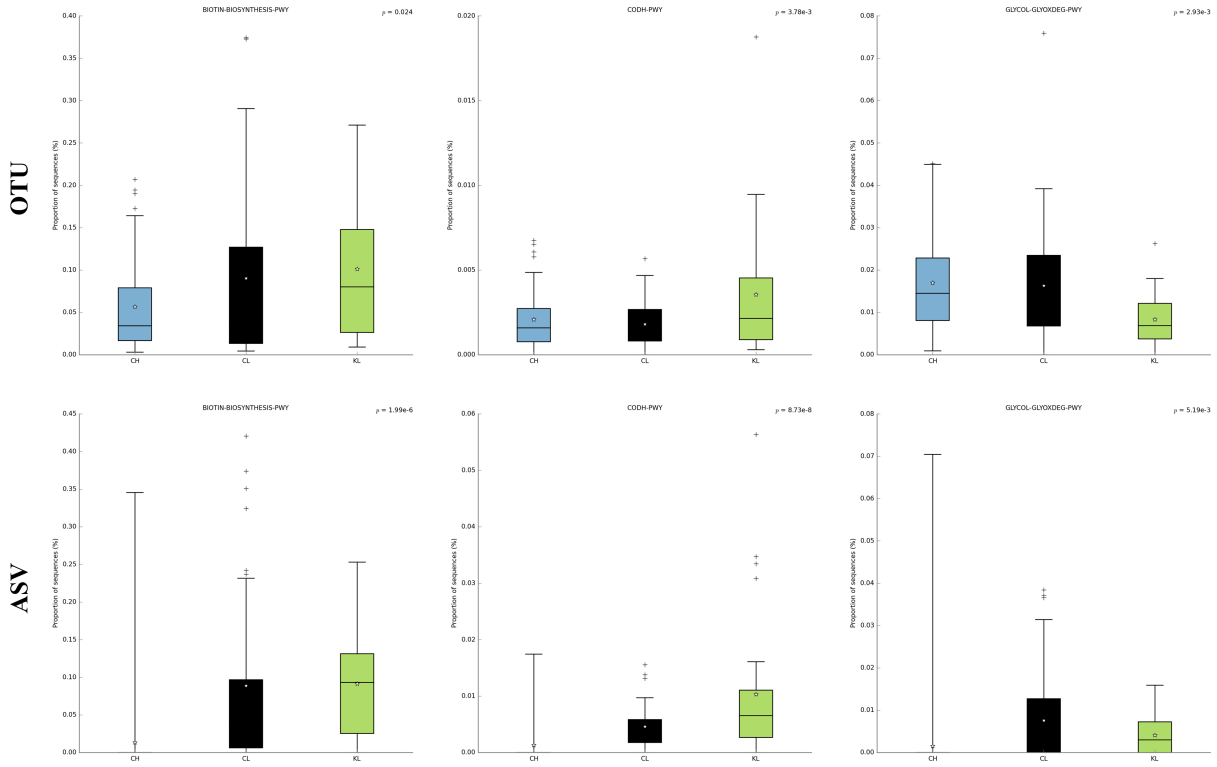

FIG. 7 ASV functional profiling exhibits differences when compared to OTU profiling. Pathways shown are 'BIOTIN-BIOSYNTHESIS-PWY', 'CODH-PWY', 'GLYCOL-GLYOXDEG-PWY' from left to right in both (A) OTU and (B) ASV datasets. The CH treatment group exhibited a significant decrease in proportion of sequences, while CL and KL treatment groups exhibited minimal change when comparing the two datasets. Statistical significance was assessed by ANOVA.

analysis when compared to the OTU counterpart. When CH and CL treatment groups were compared using the Welch's t-test, no particular pathways exhibited significant differences

between the two modes, and both showed similar trends (**Figure 8A**). Interestingly, when CH and KL treatment groups were compared using the Welch's t-test, three pathways exhibited altered trends: the superpathway of glycolysis and Enter-Doudoroff pathway, the colanic acid building block biosynthesis pathway, and the superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glyoxylate bypass. ASV-based analysis identified the CH group as exhibiting a lower mean proportion of these pathways compared to the KL group (**Figure 8B**). In both analyses, the uncontaminated areas displayed notable similarities in vole gut microbiome functional profiles, regardless of the geographical split. The contrast between functional profiles of the contaminated and uncontaminated treatment groups suggest that gut microbiome functionality is influenced by radioactivity.

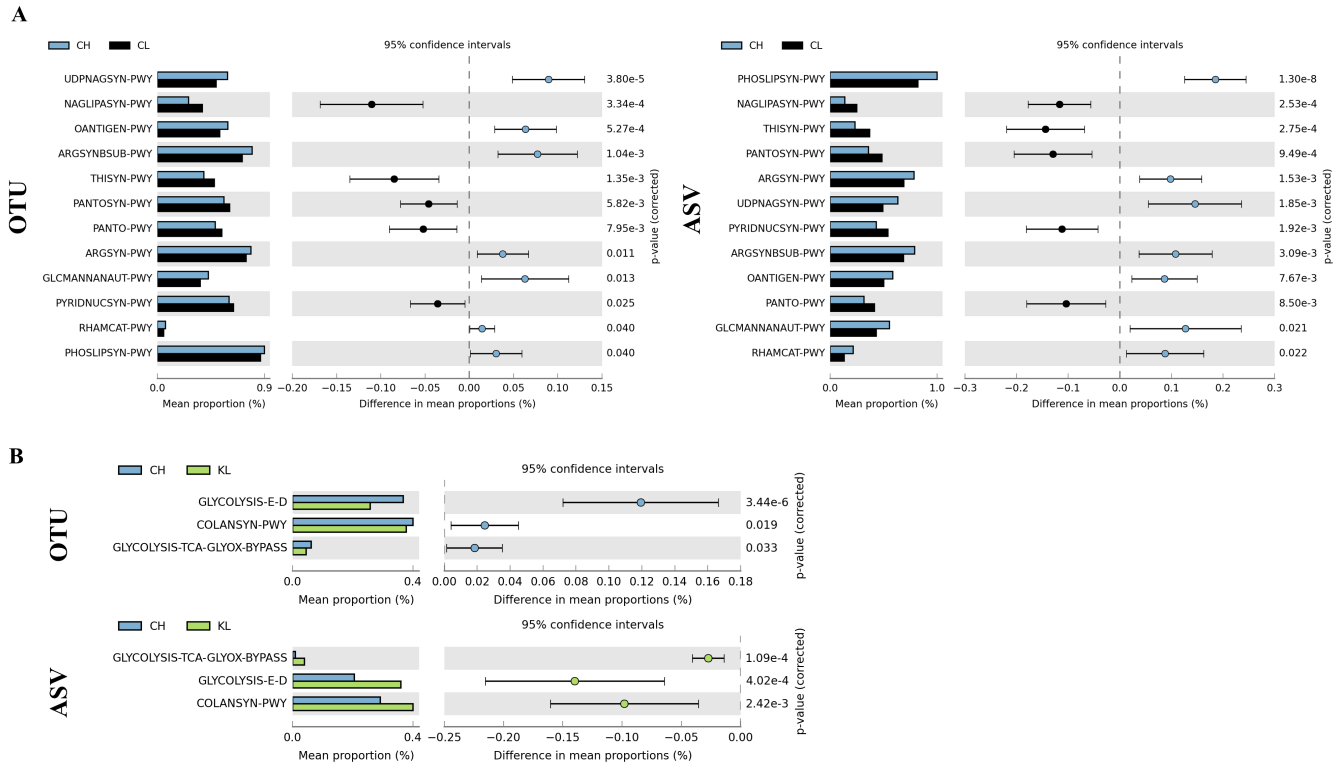

FIG. 8 Two group testing in OTU and ASV. (A) OTU and ASV functional profiling between CH and CL treatment groups show no significant difference, exhibiting similar trends in the mean proportion. (B) OTU and ASV functional profiling between CH and KL treatment groups show significant differences in three pathways - 'GLYCOLYSIS-E-D', 'COLANSYN-PWY', 'GLYCOLYSIS-TCA-GLYOX-BYPASS' - with the CH treatment group exhibiting lower mean proportion compared to the KL treatment group. Statistical significance was assessed by Welch's t-test.

The results from PICRUSt2 analysis provide a weighted Nearest Sequences Taxon Index (NSTI) number for the direct comparison of the accuracy of predicted metabolic pathways to predicted pathways in model organisms (24). The NTSI value generated from analysis by OTU had a mean of 0.161 ± 0.029 while the mean NSTI value from ASV-based analysis was 0.175 ± 0.059. A lower NSTI number indicates higher accuracy in predicted pathways relative to a model organism, suggesting that the functional pathways predicted from OTUs were more accurate than the pathways predicted from ASVs. Comparing results from OTU- and ASV-based analysis, functional profiling exhibited minute differences in most pathways, although there were still several pathways exhibiting significant decreases in functional sequence proportions in the CH treatment group following ASV-based PICRUSt2 analysis.

DISCUSSION

This study sought to compare the results of ASV- and OTU-based metabarcoding analysis using a dataset from Lavrinienko *et al.* to examine the relationship between radiation and

effects on the bank vole gut microbiota (8). Our study aimed to replicate the analyses performed by Lavrinienko *et al.*, including alpha and beta diversity through the QIIME2 pipeline, as well as determination of functional profiles using PICRUSt2. PICRUSt2 is particularly appropriate for our study given it has been updated to allow for the use of ASVs, while the analysis conducted by Lavrinienko *et al*. utilized PICRUSt1, which was limited to OTUs at the time.

Similar results for ASV and OTU analysis based on alpha diversity metrics, with a potential novel finding using ASVs. When looking at evenness measures for each treatment area, no significant differences were found between OTU and ASV analysis. However, an examination of overall evenness indicated that there was a net decrease when comparing OTU analysis to its ASV counterpart. Both analyses were also able to identify a significant difference between the CH and KL treatment groups when measuring Faith's PD. Notably, ASV analysis identified a significant difference for Faith's PD in the two low-radiation areas, CL and KL. This was not detected in the OTU-based analysis. The results of this study very closely mirror findings from general literature, as Nearing *et al.* found the results of ASV analysis to be broadly identical to OTU analysis, with minor differences in certain metrics of alpha diversity and unweighted UniFrac (33). This supports the notion that ASV analysis yields a slightly richer interpretation than OTU analysis (34, 35). As mentioned, we believe that the potentially increased acuity can be attributed to the finer thresholds of ASV analysis, which can be separated by single base changes (34).

No major differences in beta diversity when comparing ASV and OTU analysis. PCoA plots of Bray-Curtis Dissimilarity and weighted UniFrac for both ASV and OTU analyses exhibited near-identical clustering, though there was greater inter-individual variation in the CL and KL groups than in the CH group when conducting the OTU analysis (Fig 2A, 2B). With regard to the degree of separation between groups, it is possible that ASV analysis produces a clearer representation. A plausible mechanism for the minor discrepancy could be attributed to the 97% clustering threshold of OTU analysis, thus leading to the wider gap observed. In contrast, the finer resolution of an ASV-based analysis could potentially narrow this gap.

Slight variations in differential abundance, potentially due to database differences. The differential abundance plots show similar decreases in the abundance of Bacteroidetes, and the ASV-based plot shows an increased abundance of Proteobacteria in CH, which is consistent with the analysis performed by Lavrinienko *et al*. These differential abundance plots differ from Lavrinienko *et al.* in the identification of significantly different abundances of Desulfobacterota, Campilobacerota, Spirochaetes, and Actinobacteria, which are not recognized as having significantly different abundance by Lavrinienko *et al*. We suggest that the difference in detected phyla is due to the newer taxonomic database, SILVA, as well as the use of the outdated Greengenes 97% OTU reference sequence database used to assign the OTUs in our study.

Functional metabolic profile assignment by PICRUSt2 results in similar pathways between ASV analysis and OTU analysis. Overall, there are no major differences in the pathways analyzed. BIOTIN-BIOSYNTHESIS-PWY, CODH-PWY, and GLYCOL-GLYOXDEG-PWY are the only pathways where the CH treatment group decreases in sequence proportion in the ASV analysis, as compared to the OTU analysis. We believe that these findings are of increased accuracy. Further, the creators of PICRUSt2 note that compatibility with ASVs is one of the key contributors to its superiority over PICRUSt1, due to the finer resolution in denoising (25).

Limitations The primary limitation of our study is the lack of vole gut microbiome data to compare our results to aside from Lavrinienko *et al*. A useful analysis to undertake would be the comparison of our results by OTU and ASV analysis to several other papers to determine the validity of the observed differences between modes of analysis. Performing the analysis on multiple standardized datasets would also help average out variations in sampling. Another limitation of our study is that we used an out-of-date 97% OTU reference sequence database for assigning our OTUs. This was done in an effort to have comparable OTU data to Lavrinienko *et al.,* but resulted in outdated taxonomic assignment.

Conclusions The current study investigates the impact of using ASV analysis versus an OTU-based approach. In comparing various measures of alpha and beta diversity and functional metabolic profiles, it appears that an ASV-based approach provides a more detailed inference of the microbial communities present, filling in some of the gaps that can be missed via an OTU-based analysis. These findings lend support to our hypothesis and are in agreement with the current body of literature, which suggests that an ASV-based analysis should take precedence over OTU-based methods (25, 34). This is largely driven by the capability of ASV analysis to capture higher resolution of differences between gene sequences, thereby enabling a higher level of detail with regard to microbial diversity analysis.

Future Directions Based on the findings of this study, a few prominent avenues may be explored. A cross-examination among various datasets may pinpoint whether certain underlying trends exist. Consequently, this may help to predict when one might expect to observe a large discrepancy between ASV and OTU analysis. Armed with this knowledge, a retroactive application of ASV analysis could lead to more nuanced findings without the need for further data collection, which can often be resource-intensive.

Further, the question of whether ASVs should overtake OTUs and become the sole choice for microbial analysis remains unanswered. As observed in our study, diversity metrics were slightly enhanced with the use of ASV analysis. Proponents of ASV analyses also indicate advantages with technical aspects such as required computational power (34). Yet others such as Schloss suggest that the differences may be due to artificial separation (36). Certainly, much consideration is needed to balance the potential drawbacks and technical advantages that might be conferred when choosing between analyses, which may ultimately vary on a case-by-case basis.

ACKNOWLEDGEMENTS

We would like to thank the Department of Microbiology and Immunology at the University of British Columbia for giving us the knowledge and tools for investigation of this project. Notably, we would like to express our gratitude to Dr. Evelyn Sun, Dr. Stephan Koenig, Dr. Steven Hallam, Zakhar Krekhno and Stefanie Sternagel for their continued guidance, support and feedback throughout this project. Lastly, we would like to thank Lavrinienko *et al.* for providing us with the dataset used in this study (8).

CONTRIBUTIONS

Cameron Inglis: Identified research question, methods, and tools for analysis, wrote script for and interpreted results from QIIME2 and RStudio experiments, wrote QIIME2 and R sections of Methods and Results and made corresponding figures and tables.
Daniel Mun: Identified research objective, wrote script and interpreted results from PICRUSt2, wrote Abstract, Methods, Results and made corresponding figures.
Danesh Rahmani: Abstract, Introduction, Refinement of methods and results. Summary of Feedback Received
David Yeung: Discussion, partial least squares regression (not included)

REFERENCES

1. **Kinross JM, Darzi AW, Nicholson JK**. 2011. Gut microbiome-host interactions in health and disease. Genome Medicine **3**:14.
2. **Ji BW, Sheth RU, Dixit PD, Tchourine K, Vitkup D**. 2020. Macroecological dynamics of gut microbiota. Nature Microbiology **5**:768–775.
3. **Holmes E, Li JV, Marchesi JR, Nicholson JK**. 2012. Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. Cell Metabolism **16**:559–564.
4. **Hasan N, Yang H**. 2019. Factors affecting the composition of the gut microbiota, and its modulation. PeerJ **7**.
5. **Liu J, Liu C, Yue J**. 2021. Radiotherapy and the gut microbiome: Facts and fiction. Radiation Oncology **16**.
6. **Liu X, Zhou Y, Wang S, Guan H, Hu S, Huang R, Zhou P**. 2019. Impact of low-dose ionizing radiation on the composition of the gut microbiota of mice. Toxicological Sciences **171**:258–268.
7. **Moller A, Mousseau T**. 2006. Biological consequences of chernobyl: 20 Years on. Trends in Ecology & Evolution **21**:200–207.

8. **Lavrinienko A, Mappes T, Tukalenko E, Mousseau TA, Møller AP, Knight R, Morton JT, Thompson LR, Watts PC**. 2018. Environmental radiation alters the gut microbiome of the bank vole Myodes glareolus. The ISME Journal **12**:2801–2806.
9. **Hao X, Chen T**. 2012. OTU analysis using metagenomic shotgun sequencing data. PLoS ONE **7**.
10. Protocols and standards : Earthmicrobiome. earthmicrobiome RSS. Retrieved 08 December 2021.
11. 16S/18s/its sequencing. BGI. Retrieved 08 December 2021.
12. **Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MG, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CH, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG.** 2019. Reproducible, interactive, Scalable and Extensible Microbiome Data Science using QIIME 2. Nature Biotechnology **37**:852–857.
13. **Callahan B, McMurdie P, Rosen M, Han A, Johnson A, Holmes S.** 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, **13(7)**, pp.581-583.
14. **DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al.** 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microb 72(7): 5069-5072
15. **Rognes T, Flouri T, Nichols B, Quince C, Mahé F.** 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584.
16. **Katoh K, Standley DM.** 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-780.
17. **Lane DJ.** 1991.16S/23S rRNA Sequencing, p 115-175. In Stackebrandt E, Goodfellow M (ed), Nucleic Acid Techniques in Bacterial Systematic. John Wiley and Sons, Chichester, NY.
18. **Price MN, Dehal PS, Arkin AP.** 2010. FastTree 2 – Approximately maximum-likelihood trees for large alignments. PloS One 5:e9490.
19. **Michael S Robeson II, Devon R O'Rourke, Benjamin D Kaehler, Michal Ziemski, Matthew R Dillon, Jeffrey T Foster, Nicholas A Bokulich.** 2020. RESCRIPt: Reproducible sequence taxonomy reference database management for the masses. bioRxiv
20. **Bokulich, N.A., Kaehler, B.D., Rideout, J.R. et al.** 2018 Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6, 90.
21. **Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E.** 2011. Scikit-learn: machine learning in Python. J Mach Learn Res. 12:2825–30.
22. **Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, Niculae V, Prettenhofer P, Gramfort A, Grobler J, Layton R, VanderPlas J, Joly A, Holt B, Varoquaux G.** 2013. API design for machine learning software: experiences from the scikit-learn project. ECML PKDD workshop: languages for data mining and machine learning; p. 108–22.
23. **R Core Team.** 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
24. **Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C.** 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31**:814-821.
25. **Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, Huttenhower C, Langille MGI.** 2020. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* **38**:685-688.
26. **Parks DH, Tyson GW, Hugenholtz P, Beiko RG.** 2014. STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**:3123-3124.
27. **Rai S, Qian C, Pan J, Rai J, Song M, Bagaitkar J, Merchant M, Cave M, Egilmez N, McClain C.** 2021. Microbiome data analysis with applications to pre-clinical studies using QIIME2: Statistical considerations. *Genes & Diseases*, 8(2), pp.215-223.
28. **Barbera P, Kozlov AM, Czech L, Morel B, Darriba D, Flouri T, Stamatakis A.** 2019. EPA-ng: Massively parallel evolutionary placement of genetic sequences. *Systematic Biology* **68**:365-369.
29. **Czech L, Barbera P, Stamatakis A.** 2020. Genesis and Gappa: processing ,analyzing and visualizing phylogenetic (placement) data. *Bioinformatics* **36**:3263-3265.
30. **Mirarab S, Nguyen N, Warnow T.** 2011. SEPP: SATé-enabled phylogenetic placement, p. 247-258. *In* Biocomputing 2012. *WORLD SCIENTIFIC.*

31. **Louca S, Doebeli M.** 2018. Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**:1053-1055.
32. **Ye Y, Doak TG.** 2009. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLOS Computational Biology* **5:**e1000465.
33. **Nearing JT, Douglas GM, Comeau AM, Langille MGI.** 2018. Denoising the Denoisers: an independent evaluation of microbiome sequence error-correction approaches. *PeerJ* 6:e5364
34. **an BJ, McMurdie PJ, Holmes SP.** 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME* 11:2639.
35. **Needham DM, Sachdeva R, Fuhrman JA.** 2017. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J* 11:1614–1629.
36. **Schloss PD.** 2021. Amplicon Sequence Variants Artificially Split Bacterial Genomes into Separate Clusters. *mSphere* 6:4