Journal of Applied Microbiology

The selection of shotgun metagenomics software introduces biases in microbial profiling and pathogen detection

**Abbreviated running headline**

Metagenomics software selection biases

**Author names**

Ruijie Xu1,2, Sreekumari Rajeev3,†,\*, Liliana C. M. Salvador1,2,4,†,\*

**Affiliation**

1Institute of Bioinformatics, University of Georgia, Athens, GA, 30602, USA

2Center for the Ecology of Infectious Diseases, University of Georgia, Athens, GA, 30602, USA

3Department of Biomedical and Diagnostic Sciences, College of Veterinary Medicine, University of Tennessee, Knoxville, TN, 37996, USA

4Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia, Athens, GA, 30602, USA

† These authors contributed equally to this work

**\*Corresponding authors**

Liliana C. M. Salvador: Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia, 501 D. W. Brooks Drive, Athens, GA 30602, Email: [salvador@uga.edu](mailto:salvador@uga.edu)

Sreekumari Rajeev: Department of Biomedical and Diagnostic Science, College of Veterinary Medicine, University of Tennessee, 2407 River Drive, Knoxville, TN 37996, Email: srajeev@utk.edu

**Abstract**

**Aim**

Shotgun metagenomic sequencing analysis is widely used for microbial profiling of biological specimens and pathogen detection. However, very little is known about the technical biases caused by the choice of analysis software. In this study, we evaluated the most widely used shotgun metagenomics sequence analysis methods to characterize the microbial composition of wild rodent tissue samples.

**Method and Results**

We analyzed shotgun metagenomic sequence data from three sets of wild rodent tissue samples using the metagenomics software, Kraken2, CLARK, and an extended version of CLARK, CLARK-s. We demonstrated the discrepancies in results between the different software, which cause significant differences in microbial identification and biological diversity within and between samples. Our analysis also showed that the three software differed in their ability to identify the presence of *Leptospira,* a major zoonotic pathogen of one health importance in comparison to traditional methods.

**Conclusions**

Significant differences in compositional profiles for the same dataset while using different software and databases may lead to divergent biological conclusions in microbial profiling and zoonotic pathogens detection.

**Significance and Impact of Study**

This study highlights the importance to warrant caution while using shotgun metagenomics for pathogen detection and interpretation of taxonomical profiling analyses.

**Keywords**

shotgun metagenomic sequencing, pathogen detection, *Leptospira*, next-generation sequencing, Kraken2, CLARK, CLARK-s

**Introduction**

Studies analyzing the composition of microbial communities have been utilized in diverse study fields, such as ecology (Galbraith *et al.*, 2018; Grossart *et al.*, 2020), agriculture (Mashiane *et al.*, 2017; Granjou and Phillips, 2019), human and animal health (Chen *et al.*, 2019; Zhong *et al.*, 2019), and pharmacology (Chavira *et al.*, 2019; Wang *et al.*, 2019), as well as in zoonotic agent detection (Tun *et al.*, 2012). Zoonotic origin pathogens are responsible for over 60% of the infectious diseases identified in humans and can cause significant social and economic burdens (Karesh *et al.*, 2012; Cuervo-Soto, López-Pazos and Batista-García, 2018). Traditional methods used to identify the microbial agents within a biological specimen have relied on different laboratory techniques, including culture (Handelsman, 2004), antigen detection (Desmonts and Remington, 1980; Lequin, 2005), and nucleic acid marker detection (Yang and Rothman, 2004; Driscoll, 2009) protocols. However, these laboratory methods are limited to studying a single pathogen of interest and lack the ability to scrutinize the community of microorganisms potentially present in a sample. Next-Generation Sequencing (NGS) technologies have provided researchers with a set of culture-independent tools that identify pathogens directly from DNA sequences (Ghosh, Mehta and Khan, 2019), and characterize the diversity and abundance of microbial populations in biological specimens. These characteristics have led to the emergence of NGS technologies as popular tools for microbial profiling and pathogen detection (Tun *et al.*, 2012; Skarżyńska *et al.*, 2020; Grützke *et al.*, 2021).

Taxonomical profiling analysis in the metagenomics discipline utilizes two popular approaches: the 16S rRNA and the shotgun metagenomic sequencing-based approach (Jovel *et al.*, 2016). The 16S rRNA sequencing-based method uses polymerase chain reaction (PCR) to amplify hypervariable regions of bacterial 16S rRNA gene and compares these regions to a 16S reference database (DB) (Johnson *et al.*, 2019). In contrast, the shotgun metagenomic sequencing-based approach sequences all given DNA present in a sample (Sharpton, 2014). Although lower in cost (Breitwieser, Lu and Salzberg, 2019), 16S rRNA markers are only available in the genomes of most bacteria and archaea (Woese, Kandlert and Wheelis, 1990; Janda and Abbott, 2007). On the other hand, the taxonomical profiling of shotgun metagenomics sequencing data is done by comparison with a reference whole-genome database (DB). Since the data contain all genetic information present in the sample, this approach avoids the amplification biases observed in 16S rRNA sequencing (Ranjan *et al.*, 2016). Most importantly, it has broader applications such as functional profiling and allows for the identification of viruses and other microorganisms with simple genomes (Clark and Pazdernik, 2016).

Out of all currently developed shotgun metagenome sequencing-based taxonomical profiling tools, Kraken2 (Wood, Lu and Langmead, 2019), CLARK (Ounit *et al.*, 2015), and CLARK-s (Ounit and Lonardi, 2016), an extended version of CLARK developed to increase classification sensitivity, are the most frequently used software. These software were designed with the k-mer spectra comparison algorithms (Ye *et al.*, 2019a). Although not as sensitive as NCBI BLAST (Johnson *et al.*, 2008), these k-mers mapping algorithms allow faster classification analysis and require relatively smaller CPU usage. Previous benchmarks on shotgun metagenomic sequencing taxonomical profiling software have evaluated the performances of both Kraken2 and CLARK using either in silico or in vitro datasets (Peabody *et al.*, 2015; Escobar-Zepeda *et al.*, 2018; Ye *et al.*, 2019a). They were found to perform equally well at the species level (Peabody *et al.*, 2015; Escobar-Zepeda *et al.*, 2018; Ye *et al.*, 2019b). The advantage of using these artificial datasets is that their performances can be evaluated by comparing their microbial profiles with the known composition of the artificial datasets. However, the performance of these tools to analyze the microbial profiling and diagnostic applications of real-world datasets has been less studied. For specimens collected from wild animals, the microbiome compositions are unknown and potentially contain taxa that do not have their reference genomes in the used reference genome DB or that have never been identified. These situations can become a potential source of technical errors for accurate detection and profiling a sample's microbiome.

In this study, we compared the microbial profiles of tissue samples from two species of *Rattus* (*Rattus rattus* and *Rattus norvegicus*) using the shotgun metagenome sequencing taxonomic classification software, Kraken2, CLARK, and CLARK-s. We specifically address how the use of different DBs and software influence diagnostic results for a specific pathogen of interest and how different profiling results can affect the downstream analyses. We also focused on the specific detection of the zoonotic pathogen *Leptospira* in rat kidneys. The objectives of the current study are to 1) compare the taxonomical profiles of our dataset classified by Kraken2 using three different DBs; 2) compare the microbial profiles of our dataset classified by Kraken2, CLARK, and CLARK-s; 3) identify the presence of potential zoonotic pathogens such as *Leptospira* from each software’s profiling results; 4) address if different software can bias the indices characterizing within samples microbial diversity and between samples microbial relationships; and 5) compare taxa identified significantly different in abundance between different tissue samples from each software’s microbial profile. We present data demonstrating the significant differences among the characterizations of the microbial communities analyzed from the microbial profiles obtained using different DB and software. We also show that the three software report discrepant results for the presence of *Leptospira* and that their microbial profiling is found to be less sensitive for pathogen detection than traditional laboratory techniques. This study presents the biases introduced by metagenomic profiling software for microbial community characterization and the limit of using shotgun metagenomics as the tool for pathogen detection.

**Materials and Methods**

**Samples.** Tissue samples from the kidney (K), spleen (S), and lung (L) were obtained from four rats from two different species, *Rattus rattus* (R28) and *Rattus* *norvegicus* (R22, R26, and R27). Rats were captured from the island of Saint Kitts (longitude 17.3434° N and latitude – 62.7559°W) following protocols approved by the Ross University School of Veterinary Medicine (RUSVM) IACUC (approval # 17-01-04). DNA was extracted from samples using DNeasy Blood and Tissue Kits (QIAGEN Scientific Inc., MD, USA), following the manufacturer's protocol.

**Metagenomic shotgun sequencing.** DNA sample quality was assessed via analysis of the DNA purity and integrity with the agarose gel. DNA purity (OD260/OD280) and concentration were measured using the Nanodrop and Qubit 2.0. The library for metagenomic sequences was constructed with 1 μg DNA per sample. Sequencing libraries were generated using NEBNext® Ultra™ DNA Library Prep Kit for Illumina following manufacturer's instructions. The DNA sample was fragmented (350 bp), end-polished, A-tailed, ligated with Illumina sequencing adaptor and amplified with the PCR technique. The PCR products were then purified for sequencing. Before sequencing, samples were clustered on a cBot Cluster Generation System, then sequenced on an Illumina HiSeq platform for paired-end reads.

**Data pre-processing**. Sequencing adapters, low-quality reads, and host DNA reads within the metagenomic samples were removed using the software KneadData (The Huttenhower Lab, no date) with the default Trimmomatic (Bolger, Lohse and Usadel, 2014) (version 0.33) settings (SLIDINGWINDOW:4:20 MINLEN:50) and the "—very-sensitive" Bowtie (Langmead *et al.*, 2019) (version 2.3) option. The hosts' reference sequences, which were used to separate host reads from the microbial reads, were downloaded from the NCBI's RefSeq DB (Human: GCA\_000001405.28\_GRCh38.p13; *R. norvegicus*: GCF\_015227675.2\_mRatBN7.2; *R. rattus*: GCF\_011064425.1\_Rrattus\_CSIRO\_v1).

**Metagenomic profiling.**

Software. Nine metagenomics profiling software (Blastn, Diamond, Kraken2, Bracken, Centrifuge, CLARK, CLARK-s, Metaphlan3, Kaiju) were used to determine the tissues' metagenomic profiles. Kraken2 (Wood, Lu and Langmead, 2019, p. 2) is the newest version of the most widely used taxonomical classification system. The classifier uses a k-mer assignment algorithm to identify the lowest common ancestor of the query sequence in the DB. CLARK (Ounit *et al.*, 2015), a widely used classifier designed for high speed and accurate system used for taxonomical profiling, builds discriminative k-mers (k-mers found exclusively in one reference genome) from all k-mers in the DB to increase classification speed. CLARK-s (Ounit and Lonardi, 2016), which is an extended framework of CLARK, builds a spaced- k-mer based classifier using CLARK's discriminative DB. The classifier was designed to increase sensitivity by allowing mismatches in a limited number of positions during a query.

Database building. Three different DBs were used to evaluate the performance of Kraken2. A prebuilt DB, MiniKraken2\_v1\_8GB, was downloaded directly from Kraken2's website. The standard Kraken2 DB was built with Kraken2's DB building command "Kraken-build" with the "—standard" option. The customized Kraken2 DB was built with the addition of all standard DB libraries and the addition of the Refseq genome R. norvegicus (GCF\_000001895.5\_Rnor\_6.0) and Refseq genome *R. rattus* (GCF\_011064425.1\_Rrattus\_CSIRO\_v1). CLARK's discriminative 31-mer DB was built with the addition of Human, Bacteria, Archaea, and Viruses' NCBI RefSeq libraries, RefSeq genomes of *R. rattus* and *R. norvegicus*, and UniVec\_Core sequences in the Custom libraries.

Taxonomical profiling. All Kraken2, CLARK, and CLARK-s taxonomical profiling analyses were performed according to the software' manuals. To compare the differences in profiling composition across different DBs in Kraken2, we first looked at the number of reads classified into each of the following taxonomic groups: domains Eukaryota, Bacteria and Archaea, Viruses, 'Other Sequences' (reads were classified but do not belong to any of the other taxonomic groups), and 'Unclassified' (reads that were not classified using the Kraken2 classifier in both domain and genus levels).

**Data visualization**. The taxonomical profiles classified with Kraken2, CLARK, and CLARK-s were shown in histogram format using the R package "ggplot2" (Ginestet, 2011).

**Statistical analysis**. Pairwise significant difference assessments were performed by Wilcoxon signed-rank exact test, which is a non-parametric statistical hypothesis test used for comparing repeated measurements on a single sample. Significant differences across all three software were assessed by the non-parametric statistical Friedman rank-sum test. This test has been used to detect differences in measurements across more than two groups. Both statistical tests mentioned above were performed with the R package "stats" (R Core Team, 2020). Alpha (Shannon, 1948; Simpson, 1949) and beta diversity (Bray and Curtis, 1957) indices (Whittaker, 1960) were used to describe the relationship of the microbes within and between samples, respectively, and were calculated with the R package "vegan" (Oksanen *et al.*, 2013). Plotting of the beta diversity indices was done using Multidimensional scaling (MDS) (Mair, 2018), which reduces the dimensions of the pairwise comparison matrix between samples for the visualization of in-between sample microbial community relationship in a lower dimension, was done with the R package "phyloseq" (McMurdie and Holmes, 2013). Microbial compositions present in each sample were assessed by a differentially abundant taxa analysis using the R package "DeSeq2" (Love, Huber and Anders, 2014).

**Results**

**Profiling compositions with different DBs**. To address the biases introduced from database selection during metagenomics profiling, four different databases (minikraken, standard, customized, and maxikraken) were used to classify the Rattus samples using Kraken2.

Three of the four databases was built previously and provided by the science community without charge (minikrakenV2, standard, maxikraken), while the customized databases were build following the protocols provided in the Kraken2 manual. maxikraken2 DBs, although could be downloaded directly, requires over 150 GB memory the workstation used for analysis. While minikrakenV2, distributed by the developer of Kraken2, only requires less than 8GB. customized database (60 GB) was built with the same composition of the standard database (53 GB), with the addition of the two Rattus genomes, which is the host species of the dataset. With 12 threads of CPU used on UGA’s high memory computing node, the building of the customized database took ~15 hrs (Table I) to complete the building process. Kraken2 will loaded database selected into the workstation for every analysis the software perform, thus the memory resources utilized during Kraken2’s analyses are directly correlated with the choice of the databases. Time of the analyses also changes with the selection of the databases, but the differences is only in the range of seconds.

The average number of total classified reads using these DBs range from 10,755 (SD: 20,651) using the minikraken DB to 21,402 (SD: 27,043) using the maxikraken DB (Table I.1). The numbers of reads classified under the the four highest taxnomy level (Domain), Eukaryota, Bacteria, Viruses and Archaea taxa, by each databases are presented in Figure 1a-d. The statistical significance of the differences in comparisons of the classification results for each Domain taxon classified by the four DBs were validated using the paired Wilicoxon signed rank test. The adjusted p value for all comparisons between DBs are available in Table I.2. Most of these pairwise comparisons between the results of DBs were significant. For example, the differences in the number of reads classified under the Bacteria taxon were significantly different for all pairwise DB comparisons (Figure 1b). For Eukaryota taxon, only the number of reads classified by the standard and customized DBs were found not significantly different in comparion (Figure 1a). Classification results for Viruses are more similar across DBs compare to that of Eukaryota and Bacteria, but still with 4 out of 6 comparison different significantly. In the end, for Archaea classification, only the classification results of minikraken were found significanly different when compared with the results of other DBs, the classification results of other three DBs were not different significantly between each other.

To understand how differences in classification results can directly impact the characterization of the microbial communities in each sample, we calculated two alpha indices (Shannon and Simpson) at the species level, characterizing the species richness (diversity) and the evenness of the microbial communities within each Rattus sample and compared them with the observed unique number of species identified by each DB (Observed). We have found that although the observed unique taxon were significanly different across the classification results of all four DBs (Figure 2a), only the Shannon index, which describes the species richness and evenness within a community, obtained frin minikraken DB were found significantly different when compared with the results of other DBs (Figure 2b, Table I.3). Morever, the Simpson index, which describes the evenness of the microbial communities within each sample, were also found mostly similarly between results of the four DBs. Only the evenness indices obtained from the results of the standard and customized DBs were found significantly different in comparison (Figure 2b, Table I.3).

In addition to the characterization of the microbial community within each sample, relationships between the microbial communities are also very important in metagenomics studies. The pairwise relationships between every two Rattus samples in the dataset were described with the Bray-Curtis index, and clustered hierarchically (Figure 3). The statistical significance of differences between the Bray-Curtis indices obtained from the results of different DBs were validated pairwisely using the paired Wilicoxon signed rank test (Table I.4). These indicies describing the between-samples relationships were found significantly different when using different DBs. Only the Bray-Curtis indices obtained from the results of maxikraken DB and customized DB were found not different significantly. Furthermore, hierarchical clustering of the samples, which describes the relationships between-samples compare to that of other samples in a dataset, has also impacted by the differences in profiling results using different DBs. We have observed that the three kidney samples (R22.K, R26.K, and R27.K) were found always clustering with one spleen sample (R26.S) in all four DBs’ classification, however, their relationships with spleen sample R27.S changes with the results of different DBs. For example, in the clustering with minikraken DB result, R27.S clustered more closely with spleen sample R22.S before clustering with the three kidney samples, but in the clustering results of other three DBs, R27.S clustered closely with the three kidney samples and R26.S while R22.S always clustered closely with sample R28.K. Despite the changes in the lower hierarchical levels, the two major clusters describing the general relationships between samples has not changed by using different DBs. Three lung samples (R22.L, R26.L, and R27.L) has always clustered closely together away from other samples, while all kidney and spleen samples formed a separate cluster with R28.L.

**Profiling compositions using different metagenomics profiling software**.

The resources required to build database and to classify each sample diverges largely across software (Tabe I). Except for CLARK, CLARK-s, Diamond, and Kaiju, the analysis of the rest of the software could be ran with a pre-built database. With 12 threads of CPU used on UGA’s high memory computing node, building of CLARK’s database took over 42 hours to complete, with over 400 GB memory utilized (Table I). CLARK-s database, required to build on top of the CLARK’s database, took around 40 additional hour to complete, with around 300 GB memory utilized (Table I). Building of Diamond’s database, with the same computational setting, completed in ~2.4 hours utilizing ~ 8 GB, while Kaiju’s database took ~ 5 hours to complete utilizing ~115 GB of memory (Table I). As for analysis time, using 12 CPU on UGA’s high memory computing node, Diamond used ~5 hour on average to classify one sample and Blastn used ~2 hr to classify one sample. Rest of the software could finish classifying one sample within a minute on average (Table I).

To compare the profiling compositions of the different sofware, we calculated the number of total classified reads for each sample and determined these samples’ profiling compositions at the domain, phylum, genus, and species levels. The average number of total reads classified by each software ranges from 10,955 using CLARK-s to 77,499 using Diamond (Table II.1). The number of unique taxa classified by each software also ranges from 18 taxa by Metaphlan3 to 4816 taxa by Kaiju (Table II.1). Furthermore, we have found that Metaphlan3 has not classified any reads in samples of Rattus R26 (R26.K, R26.L. and R26.S) and sample R22.L and R27.K, while other software has classified on average 1252 (SD: 1408), 32748 (SD: 32178), 133 (SD: 112), 111068 (SD: 113203), and 4011 (SD: 4325) reads with these five samples respectively (Table II.2).

The number of classified reads were break down at the domain level taxa, where we closely examined the number of reads classified into Eukaryota, Bacteria, Virsues, and Archaea by each software (Figure 4). The number of reads classified into the Eukaryota taxon has the largest differences across the classification results of different software, where only the number of reads classified by Centrifuge and Diamond were found not significantly different in this taxon. Furthermore, due to the limitation of their DB composition, Metaphlan3, CLARK-s, and Kaiju did not reported reads classified into the Eukaryota taxon. Compare to reads classified into the Eukaryota taxon, the number of reads classified into the Bacteria taxon were very similar across software. Only reads classified by CLARK and CLARK-s were found significantly different in Bacteria classifications with most other software (except for when compared with Metaphlan3 and Kaiju). The classifications of Viruses by different software, on the other hand, were divided into two groups, where each group are not significantly different within each other, but different with the results classified by the software in the other group. The first group includes the Virsues classification results of Blastn, CLARK, CLARK-s, Metaphlan3, and Kaiju, and the second group includes the results of Kraken2, Bracken, and Centrifuge. Diamond classification didn’t identify any reads as Viruses in the Rattus samples. Archaea’s classification using different software are also very similar, only the classification results using Centrifuge were found significantly different with the classification results of most other software (Blastn, Diamond, Kraken2, CLARK, and CLARK-s). In addition, Bracken and Metaphlan3 didn’t classify any reads into the Archaea taxon.

The read distribution at the Phylum and Genus level were also examined to increase the resolution of comparisons between software. At the Phylum level, the number of unique phyla taxa identified by each software ranges from 5 using Metaphlan3 to to 59 using Kaiju. We extracted the top 5 phylum taxa identified from each sample and combined reads classified to other phyla into the “p\_\_Other\_Phyla” (Figure S1). Top 5 Phyla has described a large percentage of read classification for all software’s classifications. However, the distribution of reads classified into different phyla taxa are different across software. For example, Virus taxon, “p\_Pisuviricota”, has contributed to over 85% (569/665) of the reads classified in sample R22.K using Blastn, while this taxon was not identified by any other software’s classification. Nevertheless, Metaphlan3 has classified all of its reads in sample R22.K into “p\_\_Viruses\_unclassified”, and CLARK and CLARK-s has classified 63% (120/190) and 57% (95/166) of sample R22.K’s read to two different Virus taxa, “p\_\_Uroviricota” and “p\_\_Artverviricota”. Kaiju has also classified 21% of sample R22.K’s reads into “p\_\_Artverviricota” (34/157). Similar distributions in reads involving Virus taxa classification were also observed in sample R26.K, R26.S, and R27.K, where Blastn classified 54% (657/1207), 20% (28/140), and 11% (422/3794) of reads into “p\_Pisuviricota”, respectively, CLARK and CLARK-s classified a large percentage of reads into Virus taxon “p\_\_Uroviricota” (CLARK: 71% (636/900), 31/76 (41%), and 18% (201/1099); CLARK-s: 18% (50/271), 18% (7/67), 10% (83/1334), respectively), but other software has only identified a small number or none reads into a Virus taxon (Kraken2 has classified 4 reads into taxon “p\_\_Uroviricota”). Except for differences in Virus taxa identification, the distribution of the Bacteria reads classified by BLASTn, Kraken2, Bracken, Centrifuge, CLARK, CLARK-s, and Kaiju are relatively consistent across samples without Virus taxa identified. The diversity of taxa identified by Metaphlan3 are significantly less than that of other software’s classification, only the most abundant taxa were captureing the majority of the classified reads with Metaphlan3 classification. For example, Metaphlan3 has identified 100% of sample R27.L’s reads as “p\_\_Proteobacteria, while other software has identified 29% (SD: 12%) of R27.L’s reads as “p\_\_Proteobacteria” on average with unique number of Phylum taxa identified range from 2 by Diamond (91% of reads classified as “p\_\_Tenericutes”) to 50 by Kaiju. Diamond’s classification is also showing differences in read classification when comparing with results of other software. The most notiable difference is the relative abudance of taxon “p\_\_Firmicutes” classified by Diamond across samples. In the lung samples, “p\_\_Firmicutes” was classified in 17% of R22.L (SD: 9%), 20% of R26.L (SD: 9%), and 14% of R27.L (SD: 8%), but Diamond has only classified 2% (133/4900) of reads as “p\_\_Firmicutes” in sample R26.L, while “p\_\_Firmicutes” taxon was not identified in R22.L and R27.L by Diamond. On the other hand, Diamond has identified a relative larger proportion of reads as “p\_\_Firmicutes” in samples R27.S (24%) and R28.L (19%) compare to that of most other software (R27.S: mean: 2%, SD: 2%; R28.L: mean: 3%, SD: 3%), except for the Centrifuge classification (R27.S: 24%, R28.L: 9%).

Moving down to the Species level classification, the number of reads classified under taxa (strains) with the same species was aggregated together to obtain the unique number of species classified by each software. Out of all software, metaphlan3 has classified the least number of species taxa with only 18 species. On the other hand, Kaiju has classified the most number of distinct Species taxa 4128 species (Table II.4). From the species level classifications, 9 species taxon were identified by all nine software (*Leptospira interrogans*, *Leptospira borgpetersenii*, *Faecalibacterium prausnitzii*,*Bordetella pseudohinzii*, *Bordetella bronchiseptica*, *Bordetella pertussis*, *Bacteroides uniformis*, *Phocaeicola vulgatus*, and *Bartonella elizabethae*). Centrifuge and Kaiju has the largest overlapping in the species taxa identified (2285 taxa), followed by Kraken2 vs. Centrifuge (1737 taxa) and vs. Kaiju (1723 taxa). The species-level classification of the three software has shared 1,379 species taxa in total. In addition, Blastn has also shared 1253 species level taxa with Centrifuge, 1207 taxa with Kaiju, and 1126 taxa with Kraken2. CLARK and CLARK-s’s classification has also shared 1219 and 1059 species taxa wtith Kaiju specificially. To assess if different software has identified same species taxa as the most abundant taxa, species taxa with at least 10% of the reads from each sample were identified from each software’s classification. Metaphlan3 in this case, has identified most number of unique species taxa (18 taxa), while Blastn and Kaiju has the least (7 taxa). CLARK vs. CLARK-s and Kraken vs. Bracken shared most number of taxa in this category (9 and 8 taxa, respectively). Two species taxa were identified by all software as the top ten percent most abundant species taxa which are *L. interrogans* and *Bartonella elizabethae*.

**Downstream analyses for microbial community characterization**

We have also obtained the Alpha and Beta diversities of the Rattus dataset at the species level to characterize the microbial communities each sample using different software’s classification results.

For Alpha diversities, which focuses on the within-species microbial communities characterization, we obtained three indices comparing each software’s characterization, 1) the observed number of unique species within each sample (Observed), 2) the Shannon index, which characterize the richness of each sample (Shannon), and 3) the Simpson index, which describes the evenness of the microbial abundance within each sample (Simpson) (Figure 5). The numbers of unique taxa observed from each sample were largely different using different sofwares. Out of the 36 pairwise comparison between different software, only 6 comparisons were not significantly different (Table II.4), which are Blastn’s observed taxa with that of Kraken2, CLARK, and CLARK-s, comparison between CLARK and CLARK-s, and comparison between Centrifuge and Kaiju. Nevertheless, the Shannon index, obtained from these softwares are more similar than the observed numbers of taxa. 13 out of 36 comparison were found not significantly different. All the classifications of software found similar in observed taxa was also found not significant in difference for their Shannon indices. Shannon indices obtained with Blastn’s classification was also found not different from the that of Bracken and Diamond. These software similar to the Shannon indices obtained from Blastn was also found similar with each other, ex. Bracken vs. Diamond, Bracken vs. CLARK and CLARK-s, and Diamond vs. CLARK and CLARK-s, and etc. For the evennnes within each sample, the Simpson’s index were least impacted by the differences in classification results across software. Only 7 out of 36 comparison were found significantly different in Simpson indices. Most of these significantly different comparisons were identified between CLARK-s (3/7) and Centrifuge (4/7) with other softwares or between each other.

In addition to within-sample characterization, the pairwise between-sample relationships were measured by the Bray-Curtis indices and clustered using the hierarchical clustering method. Validating comparison using the paired Wilicoxon signed rank test, we identified that the pairwise between-sample relationships evaluated using Blastn was not different from that evaluated with Kraken2, Bracken, and Centrifuge, and between-sample relationships evaluated using CLARK and CLARK-s are not different from most other softwares except for Blastn, Centrifuge, and Metaphlan3, separating these software into two groups. Metaphlan3, with 5 out of 12 samples unclassified completely, was significantly different in the between-sample relationships with that obtained from other software (Table II.5). We further explored the relationships between-samples compare to that of other samples using using hierarchical clustering (Figure 6). Except for the clustering using the Metaphlan3 classification, the classification of rest software has clustered the Rattus samples into two large clusters, first cluster included three lung samples (R22.L, R26.L and R27.L) and second cluster with all the kidney and spleen samples as well as the lung sample of Rattus subject R28 (R28.L). However, the smaller clusters formed inside the second cluster varies among software. For example, Blastn’s classification has clustered three kidney samples (R22.K, R26.K, and R27.K) closely with each other before clustered further with other samples. This cluster was only observed in Centrifuge’s clustering. While in the clustering of other software, these three kidney samples were always clustered closer with other samples before clustered with each other. For example, in Diamond, CLARK, and CLARK-s, one of the three kidney samples has always clustered with R26.S before clustering with each other. In classification of Kraken2, Bracken, and Kaiju, both R26.S and R27.S was clustered with one of the kidney samples before the kidney samples clustered with each other. However, in Blastn’s classification, both R26.S and R27.S was clustered with the rest of the spleen sample first (R22.S and R28.S) before clustered together with the three kidney samples.

**Differentially abundant taxa identification**

We have also identified the differentially abundant taxa between the samples of different tissues. The microbial communities of the lung samples were found most distinctive from that of spleen and kidney samples despite the differences in the classification results reported by different software. Therefore, we have started with identify the differentially abundant taxa between the lung samples and the kidney samples.

Since Metaphlan3 wan not able to classify 2 lung samples and 2 kidney samples, we will exclude the classification analyses of Metaphlan3 from this analysis. At the species level, the number of differentially abundant taxon identified using the classification results of different software ranges from 10 (Diamond) to 596 (Centrifuge), with more taxa significantly higher in abundance in the kidney samples than that of lung samples with all software’s classifications (Figure S2). Five significantly abundant species was found shared by the classification results of all software (*Bordetella pseudohinzii*, *Bordetella bronchiseptica*, *Leptospira interrogans*, *Leptospira borgpeterseni*, and *Mycoplasm pulmonis*). Kaiju, Centrifuge, Blastn have the most number of differentially abundant taxa mostly distinct to themselves (390, 376, and 56 taxa, respectively) (Figure 7a). Furthermore, although Centrifuge identified the largest number of differentially abundant species taxon, Kaiju has identified the most number of unique phylum taxa (42), which means many of Centrifuge’s differentially abundant species has the same phylum taxonomy taxa (Figure 7a). To obtain a more generalized overview for the differentially abundant taxa identified from the classification of each software, we aggregated the species taxa into the phylum level and visualized the presence and absence of each phylum taxon as the differentially abundant taxa across different software in Figure 8. At the Phylum level analysis, taxa “p\_\_Spirochaetes”, “p\_\_Bacterodietes”,“p\_\_Protebacteria”, and “p\_\_Tenericutes” was found present in the results of all software. Diamond was missing four taxa that were identified by rest of the software ("p\_\_Aquificae”, “p\_\_Fusobacteria”, “p\_\_Firmicutes”, and “p\_\_Cyanobacteria”). Kaiju and Centrifuge were th only two software reported virus taxa as differentially abundant. Both software reported virus taxon “p\_\_Negarnaviricota”, and Kaiju reported “p\_\_Nucleocytoviricota” and “p\_\_Uroviricota”, distinctively. Archaea taxa was only reported by Kaiju, Centrifuge, and Blastn’s . All three software have reported "p\_\_Euryarchaeota”, and both Kaiju and Centrifuge reported “p\_\_Candidatus Micrarchaeota” and "p\_\_Candidatus Lokiarchaeota”. Finally, Kaiju uniquely “p\_\_Candidatus Thermoplasmatota”.

The differentially abundant taxa identified between lung samples and spleen samples were similar with those identified in the between lung samples and kidney samples. Kaiju in this case has identified the most number of differentially abundant species (484 taxa), while Diamond has identified least (44 taxa). All of the differentially abundant taxa were more abundant in the lung samples. Six species were overlapping between the differentially abundant taxa identified by the classifications of all software (*Mycoplasm pulmonis*, *Mycoplasma bovoculi*, *Mycoplasma neurolyticum*, *Bordetella pseudohinzii*, *Bordetella bronchiseptica*, and *Bacteroides uniformis*), three of the overlapping species were also identified as differentially abundant species overlapped among all software during lung vs. kidney samples comparison. Kaiju still has the most number of distinct differentially abundant species taxa (335 taxa), followed by centrifuge (268 taxa), and Blastn (46 taxa) (Figure S10). On the Phylum level, “p\_\_Bacterodietes”, “p\_\_Tenericutes”, “p\_\_Cyanobacteria” ,“p\_\_Protebacteria”, and “p\_\_Firmicutes” was identified by all software as differentially abundant. Taxa "p\_\_Aquificae”, "p\_\_Actinobacteria”,and “p\_\_Fusobacteria” were identified in by all software except for Diamond. Archaea phylum, "p\_\_Euryarchaeota”, was still the Archaea taxon identified by Blastn, Centrifuge, and Kaiju, rest of the Archaea taxa were either only identified by Kaiju and Centrifuge, or Kaiju alone. Virus taxon, “p\_\_Negarnaviricota”, was only identified by Centrifuge as differentially abundant, while Kaiju identified virus taxa only reported “p\_\_Nucleocytoviricota” and “p\_\_Uroviricota”. Morever, in this comparison, CLARK has also reported virus taxon, “p\_\_Uroviricota”, as significantly abundant.

Finally, we futher identified the differentially abundant species between kidney and spleen samples. The number of species identified ranges from 6 by Diamond and 57 by Blastn (Table II.8). More taxa was identified significantly abundant in the kidney samples compare to the spleen samples, especially in the genus level (Figure S8). Kaiju, the software identified the second highest number of distinct differentially abundant taxa at the species level, has five out of ten distinct taxa reported as Viruese (Figure S5). In general, Only 1 species (*Leptospira interrogans*) and 4 phylum taxa (“p\_\_Spirochaetes”, “p\_\_Bacteroidetes", "p\_\_Cyanobacteria”, and “p\_\_Proteobacteria”) was found overlapping with all software’s classifications (Figure S5, Figure S6). The Phylum taxon “p\_\_Firmicutes” was identified from the classifications of all software as the differentially abundant taxon except for Diamond. Kaiju has identified the only virus taxon, “p\_\_Negarnaviricota”, as a differentially abundant taxon.

***Leptospira* detection.**

With the use of the nine software, Leptospira was identified in the three tissues of all four subjects, but each software has reported Leptospira in different samples (Table I). Centrifuge is the only software reported Leptospira in all 12 Rattus sample, where 8 unique Leptospira species has been identified (8 from the pathogenic group, 1 from the saprophytic group). Kaiju has also identified Leptospira from 9 out of 12 samples with 8 unique species (7 from the pathogenic group, 1 from the saprophytic group). Kraken2, following Centrifuge and Kaiju, has classified 6 Leptospira in 6 samples with 3 unique species all from the pathogenic group. Except for Metaphlan3, all software has identified Leptospira from R22.K and R28.K, which has 31 (SD: 3) and 84,344 (SD: 2.2) reads classified under Leptospira on average, respectively. Blastn and CLARK has also identified Leptospira from R22.L, which was also identified by Centrifuge, Kaiju, and Kraken2. Metaphlan3 has only identified Leptospira in R28.K. All samples identified by at least three software has at least 30 reads classified under Leptospira in total (Table SIII.3). Samples that were only identified by Kaiju or Centrifuge has only 2 (R27.K, SD: 1) to 15 (R26.L, SD: 2) reads classified under Leptospira on average. In addition to differences in Leptospira diagnosis caused by the use of different software, diagnosis of Leptospira was different when different databases were used for Kraken2’s classification (Table SIII). Kraken2’s analyses with the maxikraken DB has identified Leptospira in all samples, while standard and customized DB has identified Leptospira in two lung samples (R22.L and R27.L). Standard DB has also identified Leptospira in the three spleen samples (R22.S, R27.S and R28.S). In addition to the metagenomics approaches, the diagnosis of Leptospira in the kidney samples was validated using three traditional methods (PCR/DFA/Culture), Leptopsira was identified in samples R22.K and R28K by all three methods, but only identified by PCR in samples R26.K.

**Discussion**

Profiling the microbial taxonomies from biological specimens allows a better understanding of the microbial communities of samples collected for different fields of studies (Coyte, Schluter and Foster, 2015; Gilbert and Lynch, 2019). The field of metagenomics, developed with the advancement of NGS technologies, allows scientists to build a complete and discriminatory microbial profile with virus, archaea, and bacteria taxa for samples collected from their target of interest (Jovel *et al.*, 2016). These metagenomic profiles can be used to detect relevant pathogens in clinical and epidemiological investigations (Qin *et al.*, 2012; Knights, Lassen and Xavier, 2013) and to observe the interactions between a micro-ecosystem and its changing environment in ecological contexts (Handley, 2019).

In this study, microbial profiles of twelve samples collected from two rat species (*Rattus rattus* and *Rattus norvegicus*) were classified. These two rat species are the major reservoirs of pathogenic *Leptospira* and contribute to its epidemiology and transmissionto humans and animals. Rats harbor this bacteria in their kidneys and are a significant source of environmental contamination (Boey, Shiokawa and Rajeev, 2019). In addition to kidney samples, we also classified the microbial profiles of samples from rat lungs and spleen. We identified the microbial profiles of these samples, which contain many potential rodent pathogen sequences, using the shotgun metagenome sequencing taxonomic classification software, Kraken2, CLARK, and CLARK-s. The differences in the classification outputs were compared and analyzed to address how the use of different taxonomical profiling software on the same dataset could affect the results of the analysis and lead to different biological conclusions.

Previous benchmarking studies (Escobar-Zepeda *et al.*, 2018; Ye *et al.*, 2019a) have performed comprehensive analyses on these software speed and performance (sensitivity, specificity, precision, and accuracy). However, these benchmarks have been generally based on *in silico* datasets or with the support of laboratory synthetic samples. Differences among the microbial taxonomical profiles classified with different tools can be the result of reporting false positive or false negative taxonomies using different algorithms, DBs, or software. These differences sometimes seem negligible in benchmarking studies for tools with similar algorithms but can lead to diverging biological conclusions in the downstream analyses depending on the questions being asked. The biases originated from these analyses have been understudied; therefore, it is crucial to raise awareness for their existence and for the factors that lead to incorrect biological conclusions.

**Kraken2 analysis with three different DBs**.

Incorrect taxonomical profiling of the collected samples’ microbial community can start from choosing an incorrect DB. Current taxonomical profiling software requires a large number of computational resources for DB building and storage. Some software, such as Kraken2, provide an alternative prebuilt DB for users with inefficient computing resources, which allow for a down-sampled DB to be loaded into a machine with RAM as low as 8 GB (full standard Kraken2 DB requires ~30 GB RAM). This DB is built with all libraries included within the standard Kraken2 DBs but down samples both reference sequences within the DBs and query sequences using a hash function. For the benefit of inexperienced users, Kraken2 also provides the option “—standard” for DB building command “kraken2-build”. This option replaces six DB building commands by directly downloading five different RefSeq libraries (bacterial, viral, archaeal, human, and UniVec\_Core)) as well as NCBI’s taxonomic information into the standard DB. However, the discrepancies between microbial profiles classified with these efficient and convenient alternative options and the profiles classified with the customized built DB with the inclusion of the Refseq genomes of the known host (customized DB) are large, which supports research findings of a recent publication (Pereira-Marques *et al.*, 2019).

The comparisons among the microbial profiles classified by different Kraken2 DBs have not only shown the importance of choosing the right DB for taxonomical analysis but also emphasized the importance of including as many genomes of the known taxonomies into the DB as possible; otherwise, the classification results could be greatly altered due to the missing genomes, leading to potentially misleading biological conclusions.

**Kraken2 vs. CLARK vs. CLARK-s**. The performances of Kraken2, CLARK, and CLARK-s have been evaluated in a previous benchmark study (Ye *et al.*, 2019a). The three software are built based on k-mer spectra DBs, while the DB of CLARK-s is built upon the DB for CLARK with the spaced k-mers to increase the accuracy. The time and storage required for building the DBs of the three software are around the same, while CLARK-s requires slightly more in both variables (time: ~10 hrs; storage: ~100 GB). Out of the three software, Kraken2 and CLARK-s require the least and the most amount of memory (RAM) for classification and DB building (~30 GB and ~108 GB, respectively). Thus, none of the three software can be used on a local computer. Kraken2 also takes the shortest classification time compared to CLARK and CLARK-s for the same dataset, and it reports a good performance in the previous benchmark studies (Escobar-Zepeda *et al.*, 2018; Ye *et al.*, 2019a). However, since all three software have reported some degree of false negative and false positive rates in the previous benchmarks (Ye *et al.*, 2019a), we should not assume that one of these software provide more accurate taxonomical profiling results for our *Rattus* metagenomic samples than the others without knowing the true microbial composition for each sample.

To assess the sensitivity of shotgun metagenomics as a tool for pathogen diagnosis, we identified the presence of the zoonotic pathogen *Leptospira* in all of our tissue samples. To increase the sensitivity of pathogen diagnosis, we set the relative abundance cut-off value for *Leptospira* presence as 0.1%, despite the potential false positives diagnosis introduced by the lower cut-off value (Peabody *et al.*, 2015; Couto *et al.*, 2018; Escobar-Zepeda *et al.*, 2018). However, for the shotgun metagenomics analyses, the three software’ detection abilities of *Leptospira* were different among each other: Kraken2 detected *Leptospira* in the kidney samples of rats R22 and R28 and in the lung samples of rats R22 and R27; CLARK identified *Leptospira* in the kidney samples of rat R28; CLARK-s identified *Leptospira* in the kidney samples of rats R22 and R28, but no presence in their lung samples. In a previous study (Rajeev *et al.*, 2020), we identified the presence of *Leptospira* using traditional methodologies in the kidney samples of rats R22, R27, and R28. These results suggest that there are also discrepancies between conventional methods and shotgun metagenomics analyses for pathogen detection, therefore, we suggest that refinements will be needed in the current -omic procedures to improve their reliability.

To further demonstrate that differences between classification profiles produced by different software can lead to divergent biological conclusions, we evaluated alpha and beta diversity indices of the metagenomic samples using microbial profiles classified by the three software. Alpha diversity indices characterize divergence, evenness, and richness within each samples’ microbial community, while beta diversity indices, in contrast with alpha diversities, explore the differences between two samples’ microbial communities. We found that the alpha and beta diversity indices evaluated from the profiles of these three software are significantly different from each other. For example, the alpha diversity indices for CLARK and CLARK-s’s microbial profiles were not statistically different among each other, while samples’ pairwise relationships measured by beta diversity indices were found statistically different between these two software’ microbial profiles. At the same time, alpha diversity indices for Kraken2 profiles were statistically different from those of CLARK-s’s, but the significant differences no longer exist for the profiles of these two software’ pairwise beta diversity indices. These findings in alpha and beta diversities can be confounding for researchers wishing to learn about the composition of their metagenomic samples’ microbial communities.

Identifying differentially abundant taxa across different biological environment or treatment group is another popular analysis for metagenomic samples. Microbial communities for each sample are composed of a large number and varieties of taxa. However, certain taxa could be significantly different in abundance in response to a biological or ecological environment. These taxa can be used as the research target for follow-up analyses to identify potential stimuli for pathology or response to changing ecological environments. In this case, we identified the differentially abundant taxa between samples collected from rat tissues to propose potential biological questions that can be asked with our *Rattus* dataset: “What taxa are found significantly different in abundance between samples collected from different rat tissues?” and “Can we detect potential zoontic pathogens such as *Leptospira?”*. We found that different taxa were reported as differentially abundant by the three software at both the genus and phylum levels (with some overlapping). The virus taxon, *Muromegalovirus*, which was reported significantly different in abundance between kidney and lung samples with ~24–25 - log fold changes for both CLARK and CLARK-s classified profiles are only found less than 1% in relative abundance within a kidney’s sample in Kraken2 classified profile. Taxa reported as differentially abundant by microbial profiles classified with different software can produce misleading biological conclusions, which may seriously influence the interpretations and the directions of further investigations.

The inconsistencies found between the results of different metagenomic classifiers show that significant biological conclusions from metagenomic profiling analyses have the potential to be only the artifacts of the software’ algorithms. Shotgun metagenomics sequences might be too short for current taxonomical profiling software to differentiate microbial taxonomies between similar genomes (Tran and Phan, 2020). The use of real-world datasets has the advantage of addressing this challenge in metagenomic studies from the users’ perspective, reminding the investigators to stay skeptical with the classification results obtained from the profiling software. On the other hand, benchmarking the software’ performances with the real-world dataset, in contrast to using *in silic*o datasets, has the limitation of lacking knowledge about the true microbial compositions within each sample, which means we could not evaluate the performance of software based on their degrees of accuracy and sensitivity nor giving direct suggestions on software’ selection. In addition, metagenomics profiling has been broadly utilized in many fields of studies, including clinical, pharmaceutical, as well as ecological studies. Each field utilizes microbial profiles differently based on the biological question proposed. Our choice of the real-world dataset could only address a limited number of software selection biases. We suggest researchers from different study fields to be aware of the possible error-prone conclusions made from metagenomics profiling analysis, and evaluate it objectively comparing it to other traditional methods (e.g. PCR, culture, or serotyping).

Advancement in sequencing as well as computational technologies allows modern-day biological research to move to a brand-new era. However, while benefiting from the powerfulness and convenience of technologies, we should always critically analyze and validate software outputs based on our prior knowledge.

**Acknowledgements**

This work was supported by the National Science Foundation under Grant No. DGE-1545433 to R.X. and startup funds to L.C.M.S. from the University of Georgia Office of Research. The sample collection, sequencing and analysis was done during S.R.’s tenure at the Ross University School of Veterinary Medicine, Saint Kitts and it was supported by internal grants from the Center for One Health and Tropical Medicine. We also would like to thank Dr. Kanae Shiokawa for her help with collection and processing of rat specimens.

**Conflicts of interest**

No conflict of interest declared.

**Repositories**

The raw sequence files (FASTQ) were submitted to the NCBI Sequence Read Archive under the Bioproject accession number: PRJNA717669. The individual isolates can be accessed under the following Biosample accession numbers: SAMN18507082 - SAMN18507091. All scripts for this publication are freely available on the following Github link: <https://github.com/salvadorlab/MetagenomicsToolsEvaluation>.

**Data summary**

The raw sequence files (FASTQ) were submitted to the NCBI Sequence Read Archive under the Bioproject accession number: PRJNA717669. The individual isolates can be accessed under the following Biosample accession numbers: SAMN18507082 - SAMN18507091. The short-read archive accession numbers are listed in Table S1.

**Ethical Approval**

Rats were captured following protocols approved by the Ross University School of Veterinary Medicine (RUSVM) IACUC (approval # 17-01-04).

**Supporting Information**

Supporting\_document.docx

TableS1\_sample\_data\_information.xlsx

TableS2\_Kraken2\_db\_domain.xlsx

TableS3\_Kraken2\_db\_comparison.xlsx

TableS4\_Kraken2\_std\_vs\_cus\_genus.xlsx

TableS5\_software\_domain\_phylum\_readsSummary.xlsx

TableS6\_software\_genus\_readsSummary.xlsx

TableS7\_Alpha\_index\_values.xlsx

TableS8\_AlphaIndexComparison.xlsx

TableS9\_Beta\_index\_values.xlsx

TableS10\_betaIndexComparison.xlsx

TableS11\_sigTaxa\_foldChange\_pvalue.xlsx

**References**

Boey, K., Shiokawa, K. and Rajeev, S. (2019) ‘Leptospira infection in rats: A literature review of global prevalence and distribution’, *PLOS Neglected Tropical Diseases*. Edited by N. P. Day, 13(8), p. e0007499. doi: 10.1371/journal.pntd.0007499.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) ‘Trimmomatic: A flexible trimmer for Illumina sequence data’, *Bioinformatics*, 30(15). doi: 10.1093/bioinformatics/btu170.

Bray, J. R. and Curtis, J. T. (1957) ‘An Ordination of the Upland Forest Communities of Southern Wisconsin’, *Ecological Monographs*, 27(4), pp. 325–349. doi: https://doi.org/10.2307/1942268.

Breitwieser, F. P., Lu, J. and Salzberg, S. L. (2019) ‘A review of methods and databases for metagenomic classification and assembly’, *Briefings in Bioinformatics*, 20(4), pp. 1125–1136. doi: 10.1093/bib/bbx120.

Chavira, A. *et al.* (2019) ‘The Microbiome and Its Potential for Pharmacology’, *Concepts and Principles of Pharmacology: 100 Years of the Handbook of Experimental Pharmacology*. Edited by J. E. Barrett, C. P. Page, and M. C. Michel, pp. 301–326. doi: 10.1007/164\_2019\_317.

Chen, Y.-Y. *et al.* (2019) ‘Microbiome–metabolome reveals the contribution of gut–kidney axis on kidney disease’, *Journal of Translational Medicine*, 17(1), p. 5. doi: 10.1186/s12967-018-1756-4.

Clark, D. P. and Pazdernik, N. J. (2016) ‘Environmental Biotechnology’, in *Biotechnology*. Elsevier, pp. 393–418. doi: 10.1016/B978-0-12-385015-7.00012-0.

Couto, N. *et al.* (2018) ‘Critical steps in clinical shotgun metagenomics for the concomitant detection and typing of microbial pathogens’, *Scientific Reports*, 8(1), p. 13767. doi: 10.1038/s41598-018-31873-w.

Coyte, K. Z., Schluter, J. and Foster, K. R. (2015) ‘The ecology of the microbiome: Networks, competition, and stability’, *Science*, 350(6261), pp. 663–666. doi: 10.1126/science.aad2602.

Cuervo-Soto, L. I., López-Pazos, S. A. and Batista-García, R. A. (2018) *Metagenomics and Diagnosis of Zoonotic Diseases*, *Farm Animals Diseases, Recent Omic Trends and New Strategies of Treatment*. IntechOpen. doi: 10.5772/intechopen.72634.

Desmonts, G. and Remington, J. S. (1980) ‘Direct agglutination test for diagnosis of Toxoplasma infection: method for increasing sensitivity and specificity’, *Journal of Clinical Microbiology*, 11(6), pp. 562–568. doi: 10.1128/jcm.11.6.562-568.1980.

Driscoll, J. R. (2009) ‘Spoligotyping for molecular epidemiology of the Mycobacterium tuberculosis complex’, *Methods in Molecular Biology (Clifton, N.J.)*, 551, pp. 117–128. doi: 10.1007/978-1-60327-999-4\_10.

Escobar-Zepeda, A. *et al.* (2018) ‘Analysis of sequencing strategies and tools for taxonomic annotation: Defining standards for progressive metagenomics’, *Scientific Reports*, 8(1), p. 12034. doi: 10.1038/s41598-018-30515-5.

Galbraith, D. A. *et al.* (2018) ‘Investigating the viral ecology of global bee communities with high-throughput metagenomics’, *Scientific Reports*, 8(1), p. 8879. doi: 10.1038/s41598-018-27164-z.

Ghosh, A., Mehta, A. and Khan, A. M. (2019) ‘Metagenomic Analysis and its Applications’, in Ranganathan, S. et al. (eds). Oxford: Academic Press, pp. 184–193. doi: https://doi.org/10.1016/B978-0-12-809633-8.20178-7.

Gilbert, J. A. and Lynch, S. V. (2019) ‘Community ecology as a framework for human microbiome research’, *Nature Medicine*, 25(6), pp. 884–889. doi: 10.1038/s41591-019-0464-9.

Ginestet, C. (2011) ‘ggplot2: Elegant Graphics for Data Analysis’, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. doi: 10.1111/j.1467-985x.2010.00676\_9.x.

Granjou, C. and Phillips, C. (2019) ‘Living and labouring soils: Metagenomic ecology and a new agricultural revolution?’, *BioSocieties*, 14(3). doi: 10.1057/s41292-018-0133-0.

Grossart, H.-P. *et al.* (2020) ‘Linking metagenomics to aquatic microbial ecology and biogeochemical cycles’, *Limnology and Oceanography*, 65(S1). doi: 10.1002/lno.11382.

Grützke, J. *et al.* (2021) ‘Direct identification and molecular characterization of zoonotic hazards in raw milk by metagenomics using Brucella as a model pathogen’, *Microbial Genomics*, 7(5), p. 000552. doi: 10.1099/mgen.0.000552.

Handelsman, J. (2004) ‘Metagenomics: Application of Genomics to Uncultured Microorganisms’, *Microbiology and Molecular Biology Reviews*, 68(4), pp. 669–685. doi: 10.1128/MMBR.68.4.669-685.2004.

Handley, K. M. (2019) ‘Determining Microbial Roles in Ecosystem Function: Redefining Microbial Food Webs and Transcending Kingdom Barriers’, *mSystems*, 4(3). doi: 10.1128/mSystems.00153-19.

Janda, J. M. and Abbott, S. L. (2007) ‘16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls’, *Journal of Clinical Microbiology*. American Society for Microbiology Journals, pp. 2761–2764. doi: 10.1128/JCM.01228-07.

Johnson, J. S. *et al.* (2019) ‘Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis’, *Nature Communications*, 10(1), p. 5029. doi: 10.1038/s41467-019-13036-1.

Johnson, M. *et al.* (2008) ‘NCBI BLAST: a better web interface.’, *Nucleic acids research*. doi: 10.1093/nar/gkn201.

Jovel, J. *et al.* (2016) ‘Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics’, *Frontiers in Microbiology*, 7. doi: 10.3389/fmicb.2016.00459.

Karesh, W. B. *et al.* (2012) ‘Ecology of zoonoses: natural and unnatural histories’, *Lancet (London, England)*, 380(9857), pp. 1936–1945. doi: 10.1016/S0140-6736(12)61678-X.

Knights, D., Lassen, K. G. and Xavier, R. J. (2013) ‘Advances in inflammatory bowel disease pathogenesis: linking host genetics and the microbiome’, *Gut*, 62(10), pp. 1505–1510. doi: 10.1136/gutjnl-2012-303954.

Langmead, B. *et al.* (2019) ‘Scaling read aligners to hundreds of threads on general-purpose processors’, *Bioinformatics*, 35(3), pp. 421–432. doi: 10.1093/bioinformatics/bty648.

Lefkowitz, E. J. *et al.* (2018) ‘Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV)’, *Nucleic Acids Research*, 46(Database issue), pp. D708–D717. doi: 10.1093/nar/gkx932.

Lequin, R. M. (2005) ‘Enzyme Immunoassay (EIA)/Enzyme-Linked Immunosorbent Assay (ELISA)’, *Clinical Chemistry*, 51(12), pp. 2415–2418. doi: 10.1373/clinchem.2005.051532.

Love, M. I., Huber, W. and Anders, S. (2014) ‘Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2’, *Genome Biology*, 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.

Mair, P. (2018) ‘Multidimensional Scaling’, *Modern Psychometrics with R*. Edited by P. Mair, pp. 257–287. doi: 10.1007/978-3-319-93177-7\_9.

Mashiane, R. A. *et al.* (2017) ‘Metagenomic analyses of bacterial endophytes associated with the phyllosphere of a Bt maize cultivar and its isogenic parental line from South Africa’, *World Journal of Microbiology and Biotechnology*, 33(4). doi: 10.1007/s11274-017-2249-y.

McMurdie, P. J. and Holmes, S. (2013) ‘Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data’, *PLoS ONE*. doi: 10.1371/journal.pone.0061217.

Oksanen, J. *et al.* (2013) ‘Package vegan’, *R Packag ver*.

Ounit, R. *et al.* (2015) ‘CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers’, *BMC Genomics*. doi: 10.1186/s12864-015-1419-2.

Ounit, R. and Lonardi, S. (2016) ‘Higher classification sensitivity of short metagenomic reads with CLARK-s’, *Bioinformatics*. doi: 10.1093/bioinformatics/btw542.

Peabody, M. A. *et al.* (2015) ‘Evaluation of shotgun metagenomics sequence classification methods using in silico and in vitro simulated communities’, *BMC Bioinformatics*, 16(1), p. 362. doi: 10.1186/s12859-015-0788-5.

Pereira-Marques, J. *et al.* (2019) ‘Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis’, *Frontiers in Microbiology*, 10. doi: 10.3389/fmicb.2019.01277.

Qin, J. *et al.* (2012) ‘A metagenome-wide association study of gut microbiota in type 2 diabetes’, *Nature*, 490(7418), pp. 55–60. doi: 10.1038/nature11450.

R Core Team (2020) ‘R: A Language and Environment for Statistical Computing’, *R Foundation for Statistical Computing*. Available at: https://www.r-project.org/ (Accessed: 25 March 2021).

Rajeev, S. *et al.* (2020) ‘Detection and Characterization of Leptospira Infection and Exposure in Rats on the Caribbean Island of Saint Kitts’, *Animals*, 10(2), p. 350. doi: 10.3390/ani10020350.

Ranjan, R. *et al.* (2016) ‘Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing’, *Biochemical and Biophysical Research Communications*, 469(4), pp. 967–977. doi: 10.1016/J.BBRC.2015.12.083.

Shannon, C. E. (1948) ‘A Mathematical Theory of Communication’, *m The Bell System Technical Journal*, 27, pp. 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x.

Sharpton, T. J. (2014) ‘An introduction to the analysis of shotgun metagenomic data’, *Frontiers in Plant Science*, 5. doi: 10.3389/fpls.2014.00209.

Simpson, E. H. (1949) ‘Measurement of Diversity’, *Nature*, 163(4148), pp. 688–688. doi: 10.1038/163688a0.

Skarżyńska, M. *et al.* (2020) ‘A metagenomic glimpse into the gut of wild and domestic animals: Quantification of antimicrobial resistance and more’, *PLOS ONE*, 15(12), p. e0242987. doi: 10.1371/journal.pone.0242987.

The Huttenhower Lab (no date) *KneadData*. Available at: https://huttenhower.sph.harvard.edu/kneaddata/ (Accessed: 25 March 2021).

Tran, Q. and Phan, V. (2020) ‘Assembling Reads Improves Taxonomic Classification of Species’, *Genes*, 11(8), p. 946. doi: 10.3390/genes11080946.

Tun, H. M. *et al.* (2012) ‘Gene-centric metagenomics analysis of feline intestinal microbiome using 454 junior pyrosequencing’, *Journal of Microbiological Methods*, 88(3), pp. 369–376. doi: 10.1016/j.mimet.2012.01.001.

Wang, J.-J. *et al.* (2019) ‘Metagenomic analysis of gut microbiota alteration in a mouse model exposed to mycotoxin deoxynivalenol’, *Toxicology and Applied Pharmacology*, 372, pp. 47–56. doi: 10.1016/j.taap.2019.04.009.

Whittaker, R. H. (1960) ‘Vegetation of the Siskiyou Mountains, Oregon and California’, *Ecological Monographs*, 30(3), pp. 279–338. doi: https://doi.org/10.2307/1943563.

Woese, C. R., Kandlert, O. and Wheelis, M. L. (1990) ‘Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya’, *Proc. Nati. Acad. Sci. USA*, 87, pp. 4576–4579. doi: 10.1073/pnas.87.12.4576.

Wood, D. E., Lu, J. and Langmead, B. (2019) ‘Improved metagenomic analysis with Kraken 2’, *Genome Biology*. doi: 10.1186/s13059-019-1891-0.

Yang, S. and Rothman, R. E. (2004) ‘PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings’, *The Lancet. Infectious Diseases*, 4(6), pp. 337–348. doi: 10.1016/S1473-3099(04)01044-8.

Ye, S. H. *et al.* (2019a) ‘Benchmarking Metagenomics Tools for Taxonomic Classification.’, *Cell*, 178(4), pp. 779–794. doi: 10.1016/j.cell.2019.07.010.

Ye, S. H. *et al.* (2019b) ‘Benchmarking Metagenomics Tools for Taxonomic Classification.’, *Cell*, 178(4), pp. 779–794. doi: 10.1016/j.cell.2019.07.010.

Zhong, H. *et al.* (2019) ‘Distinct gut metagenomics and metaproteomics signatures in prediabetics and treatment-naïve type 2 diabetics’, *EBioMedicine*, 47, pp. 373–383. doi: 10.1016/j.ebiom.2019.08.048.

**Tables**

**Table 1.** *Leptospira* detection from kidney samples using Kraken2, CLARK, CLARK-s metagenomic profiling and traditional laboratory techniques: Polymerase Chain Reaction (PCR), Direct Fluorence Antibody (DFA) test, and culture. The relative cut-off for *Leptospira* diagnostic in metagenomic samples are 0.1%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sample ID | Kraken2 | CLARK | CLARK-s | PCR/DFA/Culture |
|  | Absolute (Relative %) No. Reads | | |  |
| 22K | 53 (5.60%)\* | 31 (0.03%) | 31 (0.15%)\* | +/+/+ \* |
| 26K | 0 (0%) | 0 (0%) | 0 (0%) | -/-/- |
| 27K | 0 (0%) | 0 (0%) | 0 (0%) | +/-/- \* |
| 28K | 117,315 (96.89%)\* | 72,255 (90.53%)\* | 72,211(96.27%)\* | +/+/+ \* |

\**Leptospira* presence in the sample

**Figures Legends**

Figure 1. Domain level taxonomy profiles in samples from *Rattus rattus* and *Rattus norvegicus* tissues using different Kraken2 DBs. Each row of panels represents the absolute and correspondent percentage of compositional profiles for ‘Eukaryota’(), ‘Bacteria’ (), ‘Viruses’ ()\*, ‘Archaea’(), ‘Unclassified’ (), and ‘Other Sequences’() for each sample tissue (K: kidney, L: lung, S: spleen) from *Rattus norvegicus* (R22, R26, R27) and *Rattus rattus* (R28) using prebuilt MiniKraken2\_v1\_8GB (A, B), standard Kraken2 (C, D), and customized Kraken2 (E, F) databases. Each column in every sub-figure from left to right represents the domain level microbial compositions for samples R22\_K, R26\_K, R27\_K, R28\_K, R22\_L, R26\_L, R27\_L, R28\_L, R22\_S, R26\_S, R27\_S, R28\_S. The sum of all read classified in each sample’s compositional profile is corresponding to 100%. \*Even though viruses were classified by an independent taxonomic system, in this context, we will treat them as if they were an independent domain.

**Figure 2. Kraken2 DBs statistical comparison.** The number of reads of each tissue sample (K: kidney, L: lung, S: spleen) from *Rattus norvegicus* (R22, R26, and R27) and *Rattus rattus* (R28), which were classified into “Unclassified” (A), “Other Sequences” (B), and into four domains (Eukaryota (C), Bacteria (D), Viruses (E), and Archaea (F)) by three different Kraken2 databases (MiniKraken2\_v1\_8GB, Standard, and customized). The number of reads classified into each category for each sample are presented in Table S2. All pairwise statistical comparisons within this figure were performed with a Wilcoxon signed-rank test. Samples: R22\_K (), R22\_L (), R22\_S (), R26\_K (), R26\_L (), R26\_S (), R27\_K (), R27\_L (), R27\_S (), R28\_K (), R28\_L (), R28\_S (). \*\* p-value < 0.01, \*\*\* p-value < 0.001

**Figure 3. Phylum level microbial composition profiling in samples from *Rattus rattus* and *Rattus norvegicus* using Kraken2, Clark, and CLARK-s.** Each row of panels represents the absolute and correspondent percentage of microbial compositional profiles at the phylum taxa under ‘Bacteria’, ‘Viruses’\*, and ‘Archaea’ classified by the three different software Kraken2 (A, B), Clark (C, D), and CLARK-s (E, F)). The sum of all read classified in each sample’s microbial compositional profile is corresponding to 100%. Each column in every sub-figure from left to right represents the domain level microbial compositions for sample R22\_K, R26\_K, R27\_K, R28\_K, R22\_L, R26\_L, R27\_L, R28\_L, R22\_S, R26\_S, R27\_S, and R28\_S. Phylums present in the figure: Proteobacteria (), Actinobacteria (), Cyanobacteria (), Firmicutes (), Bacteroidetes (), Uroviricota (), Tenericutes (), Spirochaetes (), Fusobacteria (), Chlamydiae (), Aquificae (), Chloroflexi (), Thermotogae (), Artverviricota (), Peploviricota (), Deinococcus- Thermus (), Planctomycetes (), Verrucomicrobia (), Nitrospirae (), Thaumarchaeota (), Crenarchaeota (), Euryarchaeota (), Nucleocytoviricota (),Cossaviricota (), Gemmatimonadetes (), Acidobacteria (), Candidatus Gracilibacteria (), Dictyoglomi ().\* For visualization purpose, only phylum taxa with percentage >=0.1% are included.

**Figure 4. Alpha Diversityacross *Rattus* tissues.** The alpha diversity of each sample’s microbial composition is described by three indices (Shannon (A), Simpson (B), Inverse Simpson (B)), which characterize each sample’s microbial composition based on richness, evenness, and species dominance, respectively. All three indices were calculated based on the absolute number of microbial reads (including all reads classified under the genus taxa of ‘Bacteria’, ‘Viruses’, and ‘Archaea’) classified by Kraken2, CLARK, and CLARK-s. All pairwise statistical comparisons within this figure were performed with a Wilcoxon signed-rank test. Samples: R22\_K (), R22\_L (), R22\_S (), R26\_K (), R26\_L (), R26\_S (), R27\_K (), R27\_L (), R27\_S (), R28\_K (), R28\_L (), R28\_S (). \*\* p-value < 0.01, \*\*\* p-value < 0.001.

**Figure 5. Between Sample Microbial Composition Dissimilarity.** Bray-Curtis indices quantify the dissimilarity between two samples’ microbial compositions. Higher Bray-Curtis values indicate a high level of dissimilarity between the two samples’ microbial composition. They were calculated based on the absolute number of reads classified under the genus taxa of Bacteria, Viruses, and Archaea from microbial compositions of Kraken2 (A), CLARK (B), and CLARK-s (C). The red and yellow colors show low and high levels of dissimilarity, respectively. Hierarchical clustering was used to cluster together samples that have similar microbial compositions (dendrograms on the left and top of the heatmaps). Pairwise statistical comparisons between Bray-Curtis indices were calculated from each software’s microbial compositions using Wilcoxon signed-rank tests. \*\* p-value < 0.01, \*\*\*\* p-value < 0.0001.

Figure 6. Sample Clustering with Multidimensional Scaling (MDS) analyses. Bray-Curtis indices were calculated from the microbial composition classified by Kraken2 (A), CLARK (B), and CLARK-s (C) and visualized using a dimension reduction method, MDS. MDS clusters samples with similar microbial composition based on Bray-Curtis indices’ pairwise distance matrix. Samples: R22\_K (), R22\_L (), R22\_S (), R26\_K (), R26\_L (), R26\_S (), R27\_K (), R27\_L (), R27\_S (), R28\_K (), R28\_L (), R28\_S ( ).

Figure 7. Differentially abundant phylum and genus taxa. Genus taxa found differentially abundant in *Rattus* tissue, kidney (R22\_K, R26\_K, and R27\_K, R28\_K), and lung (R22\_L, R26\_L, R27\_L, and R28\_L), by Kraken2 (A), CLARK (B), and CLARK-s (C) were found above along with their log-fold change in abundance across these two tissues. Each circle corresponds to a genus taxa that were found significantly different in abundance across kidney and lung samples’ microbial compositions, and the circle color represents the corresponding phylum. Phylum: Spirochaetes(), Proteobacteria(), Tenericutes(), Firmicutes(), Bacteroidetes(), Cyanobacteria(), Peploviricota().