Assessment of bacterial and viral gut communities in healthy and tumoral colorectal tissue using RNA and DNA deep-sequencing

Ainhoa Garcia-Serrano
Karolinska Institutet
Dhananjay Mukhedkar
Hopsworks AB
Emilie Hultin
Karolinska Institutet
Ulla Rudsander
Karolinska Institutet
Yvonne Wettergren
Sahlgrenska University Hospital, Sahlgrenska Academy at University of Gothenburg
Agustín Enrique Ure
Karolinska Institutet
Laila Sara Arroyo Mühr
Karolinska Institutet
Joakim Dillner (joakim.dillner@ki.se)
Karolinska University Hospital Huddinge

Article

Keywords:

Posted Date: March 17th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2650737/v1

Abstract

Human gut microbiome studies typically focus on 16S RNA analyses and bacterial identification at the genus level. We analyzed bacterial and viral communities in colorectal tissue using both DNA and RNA sequencing and improved taxonomy resolution to species level.

Specimens from 10 colorectal cancer patients and 10 matched control patients were DNA and RNA sequenced using Illumina Novaseq. Following taxonomy classification using Kraken 2, alpha and beta diversities (different metrics) as well as relative and differential abundance were calculated.

There were no viral differences, but *P. nesessarius* had a highly increased presence in tumors (p=0.001). RNA analyses showed that *A. massiliensis* had a highly decreased transcription in tumors (p=0.002) while *F. nucleatum* transcription was highly increased in tumors (p=0.002).

In conclusion, joint assessment of the metagenome (DNA) and the metatranscriptome (RNA) at the species level identifies specific bacterial species as tumor-associated.

Introduction

Human microbiome studies have grown exponentially in the last decades, documenting several key associations to health and disease. Changes in microbiome diversity and dysbiotic states have been related to several diseases such as diabetes, obesity, autoimmune diseases, and cancer[1–5]. The Hallmarks of Cancer now includes polymorphic microbiomes as new enabling characteristics, highlighting its importance in both the carcinogenic processes as well as in prognosis or development of resistance to chemotherapy[6].

Colorectal Cancer (CRC) is the third most common incident cancer worldwide and the second in terms of mortality[7]. CRC is one of the most studied cancers when it comes to the human microbiome[8]. Diet is a major CRC risk factor and the microbiome has a direct effect on the nutrient metabolism in the colorectal epithelium[1,3]. The ecological composition of colonic mucosa can directly influence tissue microenvironment and its functionality. It has been widely demonstrated how gut microbiome shifts can lead to proinflammatory states that favorize tumorigenic processes[6,9]. Many bacterial taxa such as enterotoxigenic *Bacteroidetes fragilis* (ETBF) or *Escherichia coli* pks + have been related with tissue damage and DNA mutations via the production and secretion of bacterial enterotoxins[10]. Even so, the most CRC-associated bacterium is *Fusobacterium nucleatum spp*. Abundance of *F. nucleatum* has been strongly associated not only with the presence of colorectal cancer but also with patient outcomes and even with chemotherapy resistance[11–14].

Although many studies have described the human gut microbiome, also in large cohorts[15,16], microbiome assessment techniques vary greatly both in the laboratory phase and in the bioinformatic analysis phase. As next generation sequencing (NGS) technologies are nowadays widely available at low costs, the

varying strategies need to be further explored for, as microbial profile comparisons can vary depending on the technology used. In previous work with part of the samples used in the present study, we investigated the differences between healthy mucosa and tumor tissue from CRC patients as well as the differences between those colorectal cancer patients and healthy subjects using 16S rRNA sequencing[17]. Those analysis were focused only on bacterial communities and were restricted most of the times to a genus level classification due to the technical limitations of only sequencing the 16S rRNA gene[18]. We now wished to address those issues by studying both bacterial and viral communities at the species level in colorectal human samples using both metagenomics and metatranscriptomics (both DNA and RNA deep-sequencing).

Results

Taxonomic classification

After removal of human reads, bacterial taxonomic resolution at species level was at 89.94% for DNA sequencing reads and 88.36% for RNA sequencing reads. We identified up to 5,918 species corresponding to 6,580 taxa when performing metagenomics and 5,915 species within 6,694 taxa when analyzing metatranscriptomics. For viral species, taxonomic resolution was slightly higher, reaching 90.23% for DNA sequencing reads and 89.68% for RNA sequencing reads (628 species were identified from 696 taxa when performing metagenomics and 391 species from 436 taxa when performing metatranscriptomics).

Filtering by a minimum of 0.1% of relative abundance in at least one sample translated into detection of 55 and 64 bacterial species for DNA and RNA sequencing analysis, respectively. The top 10 bacterial species per group are presented in Figures 1A (DNA) and 3A (RNA), and the relative abundances of all species is found in Supplementary Tables S1 (DNA) and S3 (RNA). Ten viral species were detected by metagenomics and 6 viral species were identified using metatranscriptomics, after filtering. Viral species relative abundances are presented in Figures 2A (DNA), 4A (RNA) and in Supplementary Tables S2 (DNA) and S4 (RNA).

Tumor vs healthy mucosa (DNA)

Bacteria

Comparing the bacterial species between colorectal tumor specimens and healthy mucosa by analyzing the relative abundance revealed statistically significant differences (p-value <0.01) for 12 bacterial species (Supplementary table S1). Among those, *Polynucleobacter necessarius* (reported to be found mainly in water ponds, but also to be significantly enriched in patients with septic shock and as an important antibiotic resistant gene host[33–35]), was the species with highest F-value, being more abundant in tumors (Figure 1A, Supplementary table S1; F=24.652, p-value < 0.01).

Alpha diversity indexes, which informed about diversity within samples, did not show any statistical differences between groups (Figure 1B). Beta indexes, which informed about differences in bacterial

communities diversity among tumors and healthy tissue, were assessed with 2 different indexes (Figure 1C-D), Jaccard and Bray-Curtis indexes, which showed a clear clustering but not reaching statistically significance.

Differential abundance analysis revealed only one bacterial species being more abundant in tumors when comparing with healthy tissue: *Polynucleobacter necessarius* (Table 1; log2 fold change of 6.5 and adjusted p-value = 0.001).

Virus

Up to 10 viral species were identified when performing metagenomic analysis (Supplementary Table S2). However, there was no statistically significance found when analyzing relative abundance of species, alpha diversity, beta diversity nor differential abundance analysis (Figure 2).

Tumor vs healthy mucosa (RNA)

Bacteria

Comparison of colorectal tumor and healthy mucosa tissue revealed a decreased relative abundance of *Arabia massiliensis* (F=26.021, p-value <0.01) in tumors (Figure 3A, Supplementary table S3). Alpha diversity indexes, which informed about diversity within samples, did not show any statistical differences between groups (Figure 3B). Bacterial communities showed a clear clustering when evaluating Jaccard (p-value 0.001) and Bray-Curtis (p-value 0.001) beta diversity indexes (Figure 3C-D).

Differential abundance analysis reported 2 bacterial species: *Arabia massiliensis* being more abundant in healthy tissue (log2 fold change of -12.68 and adjusted p-value = 0.0001) and *Fusobacterium nucleatum* (log2 fold change of 5.31 and adjusted p-value = 0.002) being more abundant in tumors (Table 1).

Virus

Up to 6 viral species were detected when analyzing metatranscriptomes for both tumors and mucosa (Supplementary Table S2). However, there was no statistically significance found when analyzing species relative abundance, alpha diversity, beta diversity nor differential abundance analysis (Figure 4).

Table 1. Differentially (p<0.01) abundant bacteria (healthy mucosa vs tumor) in DNA and RNA datasets.

Group	Species	Log2 fold change	adj p-value
DNA	*Polynucleobacter necessarius*	6.5	0.001
RNA	*Arabia massiliensis*	-12.68	0.0001
	Fusobacterium nucleatum	5.31	0.002

Discussion

Microbiome studies in relation to human health is an expanding field, that has provided many important insights, but studies may be fraught with variability in sequencing technologies, bioinformatic pipelines and downstream diversity analysis. We compared different analysis (relative abundance, alpha and beta diversities as well as differential abundances) for both DNA and RNA sequencing data, aiming to detect bacteria and viruses up to species level.

Prior to analysis, bacterial and viral species were filtered (0.1% of total bacterial/viral reads presence in at least one sample) to reduce complexity, noise and technical variability while preserving data integrity and representing main communities[36]. Choosing the appropriate cutoff for filtering is a key step in microbiome analysis since it can strongly influence the downstream results and result in false positivity/negativity[36]. We also included multiple metrics for each analysis. Up to 3 different indexes were used to analyze the alpha diversity, offering a wide perspective of within sample diversity, increasing sensitivity for both richness and evenness[37]. Beta diversity analyses included non-phylogenetic indexes accounting for both qualitative (Jaccard) and quantitative (Bray-Curtis) indexes to avoid bias of undersampling meanwhile maintaining sensitivity to rare species[38] and differential abundance analysis to provide more reliable results for specific species enrichment (more than just comparing relative abundances) since this normalized and transformed taxa can account for data sparsity[29].

While no viral differences were detected, our study revealed statistical significances for 3 bacterial species. *P. necessarius* and *F. nucleatum* were highly increased in tumor specimens when analyzing metagenomics and metatranscriptomics, respectively. Furthermore, *A. massiliensis* was highly decreased in tumor tissue when analyzing RNA data. *P. necessarius* did not reach the established cut-off when analyzing RNA sequencing data (present in < 0.1% of total bacterial transcripts in at least one sample). This species has been mostly identified in freshwater habitats[33], but there are studies where the corresponding genus has been reported to be significantly enriched in patients with septic shock and as one important antibiotic resistance gene host[34,35]. When encountering environmental microorganisms, it is important to identify if those really come from the tumor/specimen, or if those correspond to deposition, and therefore we highlight the importance of using negative controls to be able to search for microorganisms that may represent deposition/contamination. Also, both DNA and RNA analysis were performed within the exact same samples, and a systematic identification was not detected. Furthermore, observing a lack of RNA transcription casts doubt on possible role of this microorganism in colorectal tumors. Previous studies on skin, have detected several hundreds of human papillomavirus types when analyzing DNA sequences, but most of these viruses were not actively transcribed (apparently representing deposition)[39].

Both *A. massiliensis* (recently identified in stool material)[40] and *F. nucleatum* (the by far most CRC-associated bacterium in previous studies) were highly detected among the RNA sequences of the mucosa and the tumors, respectively. Compared to DNA analysis, *A. massiliensis* did not surpass the established cut-off, and *F. nucleatum* did reach it, but its relative abundance presence was low and not different

between tumor and healthy mucosa specimens. Absence or low prevalence of these species may be explained by the higher abundance of other species.

This study demonstrated how deep sequencing of both DNA and RNA enables a wider perspective of microbiome profiles: individual microbial features can vary depending on whether the metagenome or metatranscriptome is analysed. The role of the recently identified *A. massiliensis*[40] needs to be further investigated to elucidate its implications in human health. While *F. nucleatum* is already known to be a potential colorectal cancer biomarker for screening, diagnosis, and prognosis prediction for patient outcomes[11–14], the association mainly with transcription of *F. nucleatum* may provide further insights on the role of this microorganism in CRC.

Methods

Cohort information

The present study includes biopsies from ten patients who were diagnosed with stage I–III colon cancer after colonoscopy examination and 10 patients that acted as controls where no tumor (neither malignant nor benign) was seen during colonoscopy examination. Cases and controls were matched by age and sex and description of all tumors and 2/10 controls as well as results on their 16S rRNA sequencing analysis have previously been published[17]. Characteristics of age, sex, BMI and tumor stage from these 20 patients can be seen in Table 2. The study was approved by the Regional Ethical Review Board in Gothenburg under study number 233-10 and registered at ClinicalTrials.gov (ID: NCT03072641). Informed consent was obtained from all study subjects. All research was performed in accordance with relevant guidelines and regulations.

Table 2. Clinical information for patients included in the study.

Group	CRC patients	Non-CRC subjects
Sample size	n=10	n=10
Sex	7 Females 3 Males	7 Females 3 Males
Age mean(sd)	71.5 (9.78)	70.8 (9.77)
BMI mean(sd)	24.96 (3.28)	NA
Stage	I (1), II (4), III (5)	-

DNA and RNA extraction

DNA was extracted using the AllPrep DNA/RNA/Protein Kit (Qiagen, Hilden, Germany), followed by PCR inhibitor removal with OneStep-96 PCR Inhibitor Removal Kit (Zymo Research, Irvine, California, USA), and DNA concentration measurement with Qubit 2.0 Fluorometer (Life Technologies, Darmstadt, Germany). Total RNA was extracted with the AllPrep DNA/RNA/Protein Kit (Qiagen, Hilden, Germany) and reverse transcribed to single-stranded cDNA using the High Capacity cDNA Reverse Transcription kit (ThermoFisher Scientific, Massachusetts, USA) following manufacturer's instructions. Double-stranded cDNA was prepared following step 2 of the Maxima H Minus DS cDNA Kit (ThermoFisher Scientific, Massachusetts, USA) and the cDNA was cleaned using GeneJET PCR Purification Kit (ThermoFisher Scientific, Massachusetts, USA). Thereafter, cDNA was subjected to PCR inhibitor removal with OneStep-96 PCR Inhibitor Removal Kit (Zymo Research, Irvine, California, USA) and cDNA concentration measurement with Qubit 2.0 Fluorometer (Life Technologies, Darmstadt, Germany). Both DNA and cDNA were stored at -20 degrees until further analysis.

Whole genome sequencing

Extracted material (DNA and cDNA) was thereafter subjected to library preparation, using Nextera XT DNA library preparation kit (Illumina, San Diego, CA, USA) following the manufacturers' reference guide, starting with 1 ng of DNA/cDNA (as recommended by the manufacturer) and using unique indexed adapters to facilitate pooling of the libraries. Libraries were individually quantified using the QuantiFluor system (Promega, US) and the library sizes were measured using the Bioanalyzer (Agilent, Santa Clara, CA, USA) as quality analysis. All 20 libraries were normalized to 2 nM and pooled prior paired-end sequencing using NovaSeq 6000 system (Illumina, USA) at 2 × 150 bp aiming for 100 M high quality paired end reads/sample.

Bioinformatic preprocessing

Indexes, included in the Illumina adapters, were used to assign raw sequence reads obtained from the NextSeq500 (Illumina) platform to the originating samples. Reads were quality checked and adapter trimmed with Trimmomatic using 36 bp as minimal length for the reads[19]. High quality reads were screened against the human reference genome hg19 using NextGenMap[20] and only reads that did not map to the human genome, with >95% identity over 75% of their length, were considered as non-human and further analyzed for microbiome. Once human reads were filtered from the data set, high-quality non-human reads were classified using Kraken2 v. 2.1.1[21], which was run against a reference database containing all RefSeq bacterial and viral genomes (built in December 2020) with a 0.1 confidence threshold.

Diversity analysis and statistics

All diversity analyses were performed at species level using R (v.4.2.2). Packages used in this analysis were biomformat[22], phyloseq[23], ggvenn[24], tidyr[25], ggpubr[26], vegan[27], vtable[28], metagenomeSeq[29], funrar[30], and superheat[31]. The biom files generated with kraken-biom[32] were used, together with sample

metadata, to construct phyloseq objects. Results reported all species which comprised more than 0.1% (in at least one sample) of total bacterial or viral reads, separately. Bacterial and viral species relative abundance comparison using group F-tests was done for healthy mucosa and tumors in both DNA and RNA datasets (Supplementary tables S1-S4). Relative abundance plots for top 10 species were plotted for tumor and healthy mucosa in both DNA and RNA datasets. Observed species, Shannon and Inverse Simpson alpha diversity indexes were calculated after rarefaction (between 68,535 and 5,034,325 reads, depending on the subset analyzed) to standardize species representation regardless of sample depth. Differences between groups for alpha diversity were calculated using unpaired Wilcoxon tests. Bray-Curtis and Jaccard beta diversity indexes were calculated to analyze differences between bacterial and viral communities and ANOSIM tests were used to stablish if differences between groups were greater than differences within groups. Finally, differential abundance (DA) analysis comparing tumor and healthy tissue was performed for DNA and RNA separately. In this last step, species counts were transformed using cumulative sum scaling (CSS), log2 transformation as well as a pseudocount addition to handle data sparsity. P-values obtained from DA analysis were corrected using FDR. Statistical significance was obtained when p-value <0.01.

Declarations

Supplementary files

Supplementary Tables S1 to S4.

Acknowledgements

The authors acknowledge support from the National Genomics Infrastructure funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure.

The authors thank associate Professor Peter Rolny at the Department of Medicine, Division of Gastroenterology, Sahlgrenska University hospital/Östra, Gothenburg,Sweden for being responsible for the inclusion of patients and sampling of biopsies.

Author contributions

J.D conceived and designed the study. J.D, U.R, E.H and Y.W performed the study coordination and sample management. Y.W was responsible for inclusion of patients and sampling biopsies, extraction protocols and reverse transcription of RNA material. E.H did the laboratory analysis for NGS including library preparation and sequencing management. L.S.A.M, A.E.U, D.M and A.G.S analyzed the data. J.D, L.S.A.M and A.G.S assessed the statistical analyses and results interpretation. A.G.S and L.S.A.M drafted the manuscript. All authors read and approved the final manuscript.

Data availability statement

All the aligned, non-human sequences are available at the Sequence Read Archive (SRA) within the bio-project ID PRJNA943491(https://www.ncbi.nlm.nih.gov/bioproject/943491).

Funding information

This Project was funded by the Nordic Information for Action eScience Center (NIASC), a Nordic Center of Excellence in eScience funded by NordForsk (Project no. 62721), the Swedish state under the LUA-ALF agreement (grant number ALFGBG-542821) and by the Human Exposome Assessment Platform (Project No. 874662) granted by Horizon 2020.

Competing interests

The authors declare no competing interests.

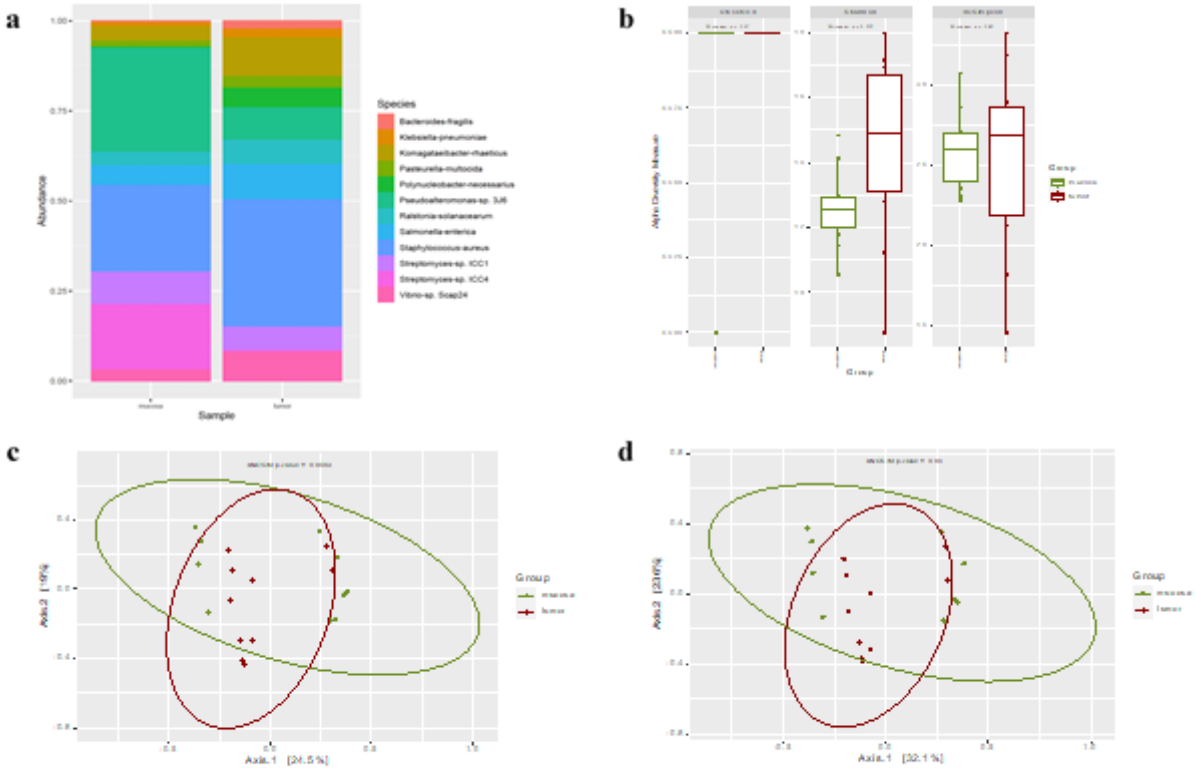
References

1. Valdes, A. M., Walter, J., Segal, E. & Spector, T. D. Role of the gut microbiota in nutrition and health. BMJ (Online) 361, 36–44 (2018).
2. Gomes, A. C., Hoffmann, C. & Mota, J. F. The human gut microbiota: Metabolism and perspective in obesity. Gut Microbes 1–18 (2018) doi:10.1080/19490976.2018.1465157.
3. Dabke, K., Hendrick, G. & Devkota, S. The gut microbiome and metabolic syndrome. Journal of Clinical Investigation 129, 4050–4057 (2019).
4. Miyauchi, E., Shimokawa, C., Steimle, A., Desai, M. S. & Ohno, H. The impact of the gut microbiome on extra-intestinal autoimmune diseases. Nat Rev Immunol 23, 9–23 (2023).
5. Knippel, R. J., Drewes, J. L. & Sears, C. L. The Cancer Microbiome: Recent Highlights and Knowledge Gaps. Cancer Discov 11, 2378–2395 (2021).
6. Hanahan, D. Hallmarks of Cancer: New Dimensions. Cancer Discov 12, 31–46 (2022).
7. Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 0, 1–41 (2021).
8. Lloyd-Price, J., Abu-Ali, G. & Huttenhower, C. The healthy human microbiome. Genome Med 8, 51 (2016).
9. Hale, V. L. *et al.* Shifts in the fecal microbiota associated with adenomatous polyps. Cancer Epidemiology Biomarkers and Prevention 26, 85–94 (2017).
10. Tjalsma, H., Boleij, A., Marchesi, J. R. & Dutilh, B. E. A bacterial driver-passenger model for colorectal cancer: Beyond the usual suspects. Nat Rev Microbiol 10, 575–582 (2012).
11. Gethings-Behncke, C. *et al.* Fusobacterium nucleatum in the colorectum, and its association with cancer risk and survival: a systematic review and meta-analysis. *Cancer Epidemiology Biomarkers & Prevention* cebp.1295.2018 (2020) doi:10.1158/1055-9965.EPI-18-1295.
12. Amitay, E. L. *et al.* Fusobacterium and colorectal cancer: Causal factor or passenger? Results from a large colorectal cancer screening study. Carcinogenesis 38, 781–788 (2017).

13. Yu, T. C. *et al.* Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. Cell 170, 548–563.e16 (2017).
14. Zhang, S. *et al.* Fusobacterium nucleatum promotes chemoresistance to 5-fluorouracil by upregulation of BIRC3 expression in colorectal cancer. Journal of Experimental & Clinical Cancer Research 38, 14 (2019).
15. The Integrative Human Microbiome Project. Nature 569, 641–648 (2019).
16. Ehrlich, S. D. MetaHIT: The European Union Project on Metagenomics of the Human Intestinal Tract. in Metagenomics of the Human Body 307–316 (Springer New York, 2011). doi:10.1007/978-1-4419-7089-3_15.
17. Hibberd, A. A. *et al.* Intestinal microbiota is altered in patients with colon cancer and modified by probiotic intervention. BMJ Open Gastroenterol 4, e000145 (2017).
18. Brumfield, K. D., Huq, A., Colwell, R. R., Olds, J. L. & Leddy, M. B. Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. PLoS One 15, 1–21 (2020).
19. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014).
20. Sedlazeck, F. J., Rescheneder, P. & von Haeseler, A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. Bioinformatics 29, 2790–2791 (2013).
21. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. Genome Biol 20, 257 (2019).
22. McMurdie PJ, P. J. biomformat: An interface package for the BIOM file format. Preprint at (2022).
23. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One 8, e61217 (2013).
24. Linlin Yan. ggvenn: Draw Venn Diagram by 'ggplot2'. Preprint at (2021).
25. Wickham H, Vaughan D, G. M. tidyr: Tidy Messy Data. Preprint at (2023).
26. Alboukadel Kassambara. ggpubr: 'ggplot2' Based Publication Ready Plots. Preprint at https://cran.r-project.org/web/packages/ggpubr/index.html (2022).
27. Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, R. K., Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, G., L. Simpson, Peter Solymos, M. Henry H. Stevens, E. S. and & Package, H. W. vegan: Community Ecology Package. Preprint at (2020).
28. Nick Huntington-Klein. vtable: Variable Table for Variable Documentation. Preprint at (2022).
29. Paulson, J. N., Stine, O. C., Bravo, H. C. & Pop, M. Differential abundance analysis for microbial marker-gene surveys. Nat Methods 10, 1200–1202 (2013).
30. Grenié, M., Denelle, P., Tucker, C. M., Munoz, F. & Violle, C. funrar: An R package to characterize functional rarity. Divers Distrib 23, 1365–1371 (2017).
31. Barter, R. L. & Yu, B. Superheat: An R Package for Creating Beautiful and Extendable Heatmaps for Visualizing Complex Data. Journal of Computational and Graphical Statistics 27, 910–922 (2018).

32. Dabdoub, S. kraken-biom: Enabling interoperative format conversion for Kraken results. Preprint at (2016).
33. Hahn, M. W. Isolation of Strains Belonging to the Cosmopolitan Polynucleobacter necessarius Cluster from Freshwater Habitats Located in Three Climatic Zones. Appl Environ Microbiol 69, 5248–5254 (2003).
34. Wang, C. *et al.* Characterization of the blood and neutrophil-specific microbiomes and exploration of potential bacterial biomarkers for sepsis in surgical patients. Immun Inflamm Dis 9, 1343–1357 (2021).
35. Bai, Y., Ruan, X., Xie, X. & Yan, Z. Antibiotic resistome profile based on metagenomics in raw surface drinking water source and the influence of environmental factor: A case study in Huaihe River Basin, China. Environmental Pollution 248, 438–447 (2019).
36. Cao, Q. *et al.* Effects of Rare Microbiome Taxa Filtering on Statistical Analysis. Front Microbiol 11, (2021).
37. DeJong, T. M. A Comparison of Three Diversity Indices Based on Their Components of Richness and Evenness. Oikos 26, 222 (1975).
38. Beck, J., Holloway, J. D. & Schwanghart, W. Undersampling and the measurement of beta diversity. Methods Ecol Evol 4, 370–382 (2013).
39. Hultin, E., Arroyo Mühr, L. S., Lagheden, C. & Dillner, J. HPV transcription in skin tumors. PLoS One 14, e0217942 (2019).
40. Traore, S. I. *et al.* Description of 'Arabia massiliensis' gen. nov., sp. nov., 'Gordonibacter massiliensis' sp. nov., and 'Bacilliculturomica massiliensis' gen. nov., sp. nov., isolated from a faecal specimen of a 50-year-old Saudi Bedouin woman. New Microbes New Infect 19, 87–90 (2017).

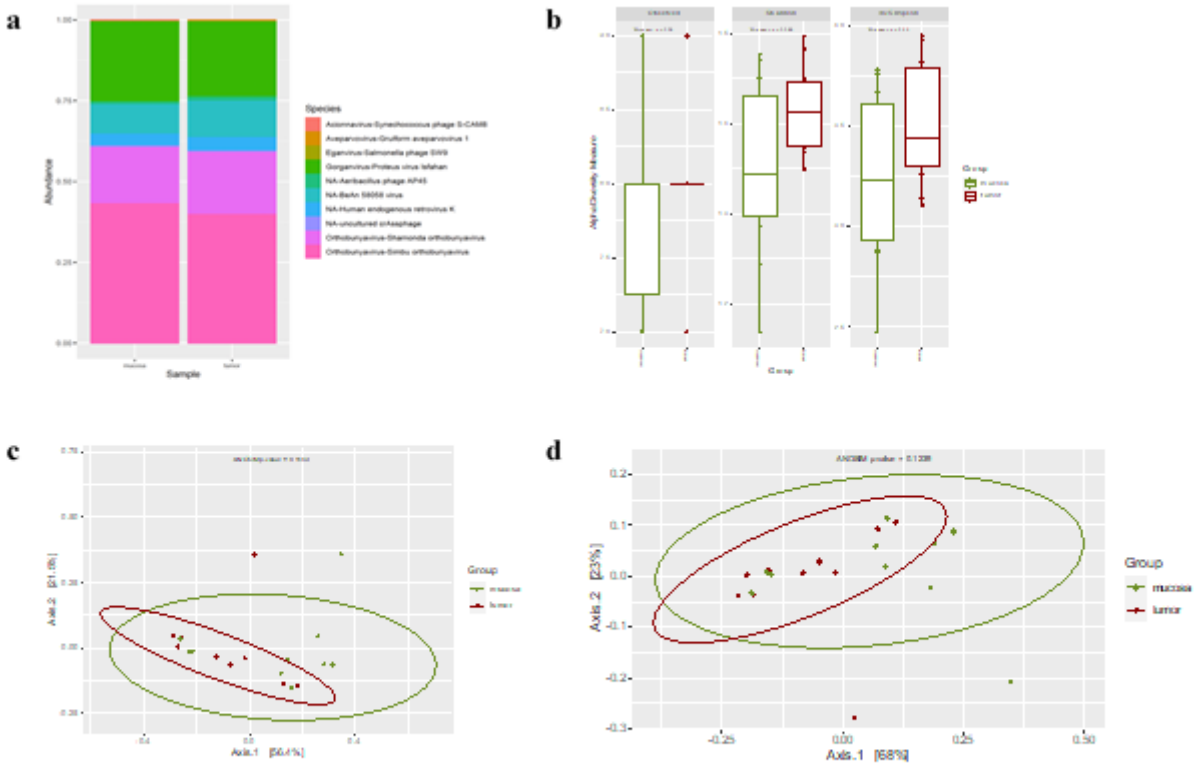
Figures



a) Relative abundance grouped for top 10 bacterial species per group. **b)** Observed, Shannon and Inverse Simpson alpha diversity indexes rarefacted at 5M reads. **c)** Jaccard beta diversity PCoA. **d)** Bray-Curtis beta diversity PCoA.

Figure 1

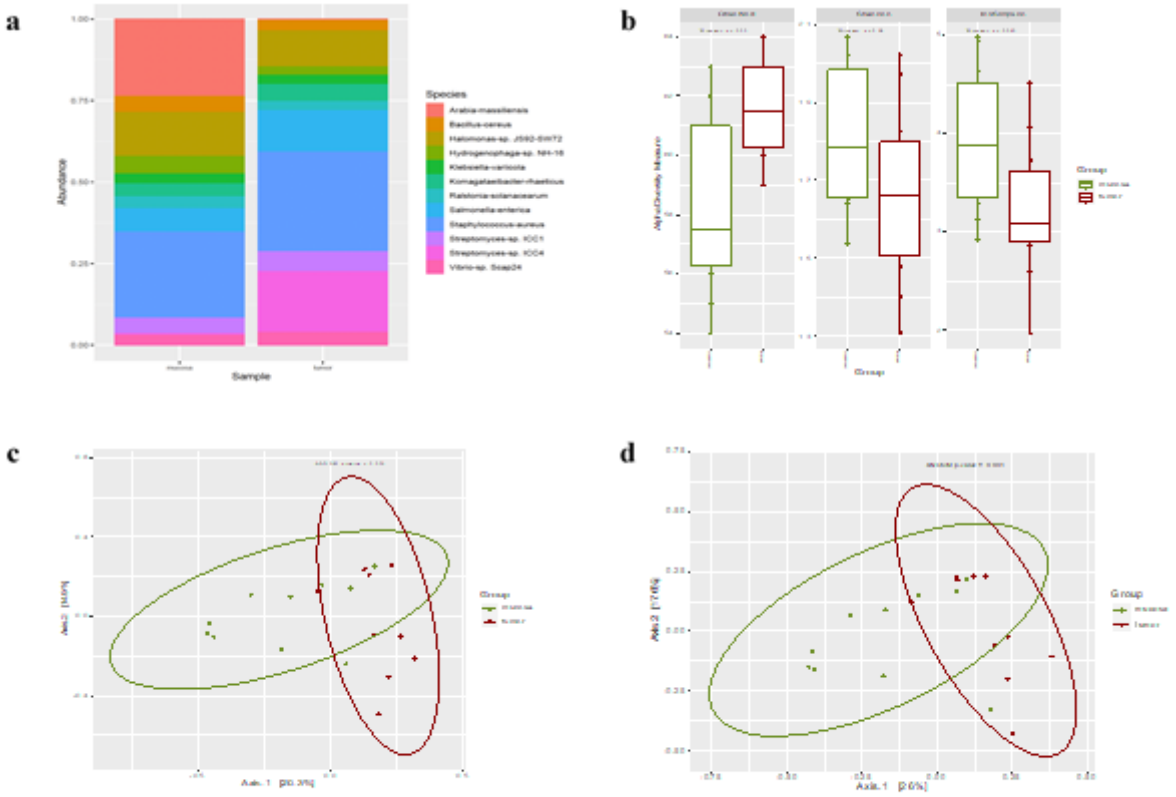
DNA Bacterial diversity of filtered reads.



DNA Virus diversity of filtered reads. **a)** Relative abundance grouped for top 10 virus species per group. **b)** Observed, Shannon and Inverse Simpson alpha diversity indexes rarefacted at 68K. **c)** Jaccard beta diversity PCoA. **d)** Bray-Curtis beta diversity PCoA.

Figure 2

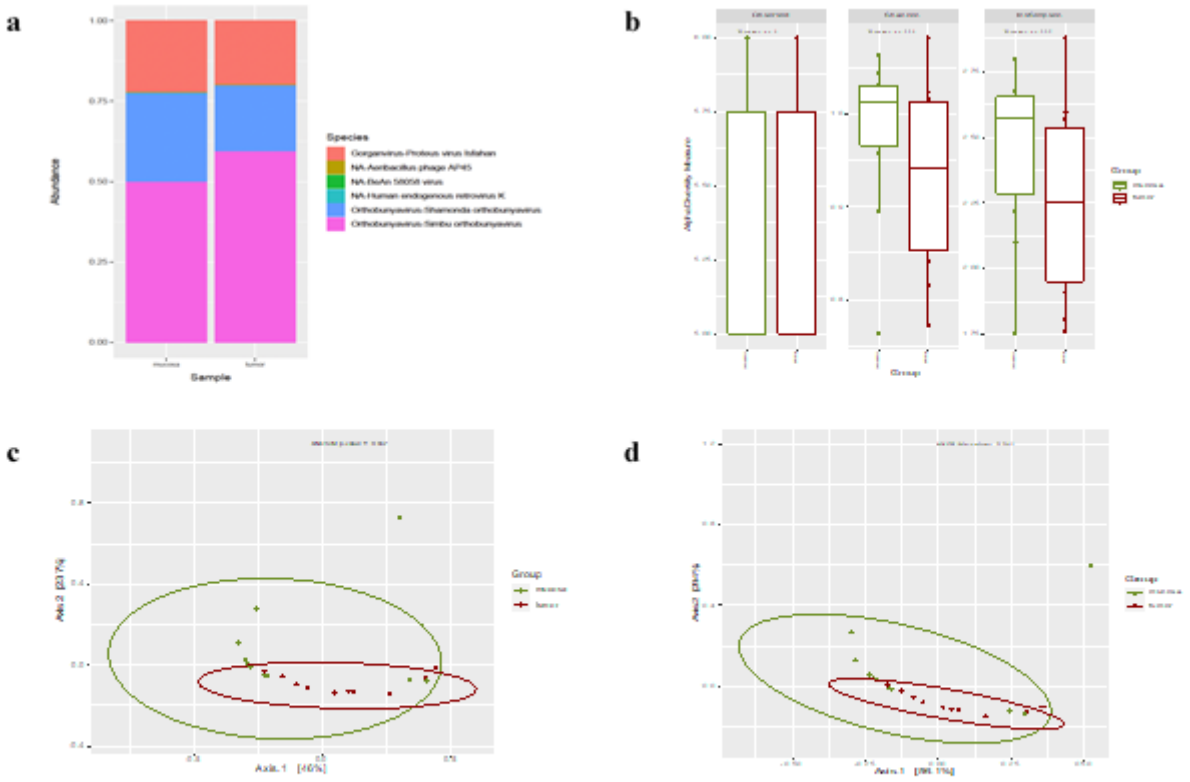
DNA Viral diversity of filtered reads.



RNA Bacterial diversity of filtered reads. **a)** Relative abundance grouped for top 10 bacterial species per group. **b)** Observed, Shannon and Inverse Simpson alpha diversity indexes rarefacted at 771K reads. **c)** Jaccard beta diversity PCoA. **d)** Bray-Curtis beta diversity PCoA.

Figure 3

RNA Bacterial diversity of filtered reads.



RNA Virus diversity of filtered reads. **a)** Relative abundance grouped for top virus species per group. **b)** Observed, Shannon and Inverse Simpson alpha diversity indexes rarefacted at 416K. **c)** Jaccard beta diversity PCoA. **d)** Bray-Curtis beta diversity PCoA.

Figure 4

RNA Viral diversity of filtered reads.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryinformationamendmentscirep.xlsx
- Supplementaryinformation.xlsx