JOURNAL OF MICROBIOLOGY & BIOLOGY EDUCATION
DOI: https://doi.org/10.1128/jmbe.v21i1.2019

JMBE Journal of Microbiology & Biology Education

Supplemental Materials

for

Testing the "Grandma Hypothesis": Characterizing Skin Microbiome Diversity as a Project-Based Learning Approach to Genomics

Marcos Pérez-Losada[1,2]*, Kelly M. Crandall[3], and Keith A. Crandall[1]

[1]*Computational Biology Institute, Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health, George Washington University, Washington, DC 20052, USA;*

[2]*CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrario de Vairão, Vairão 4485-661, Portugal;*

[3]*Department of English, Woodgrove High School, Purcellville, VA 20132, USA*

Table of Contents

(Total pages 23)

*Corresponding author. Mailing address: Computation Biology Institute, Department of Biostatistics and Bioinformatics, Milken Institute of Public Health, George Washington University, 800 22nd Street NW, Suite 7810, Washington, DC 20052-0066, USA. Phone: 202-765-8760. E-mail: mlosada@gwu.edu.

Received: 28 October 2019; Accepted: 20 January 2020; Published: 28 February 2020.

©2020 Author(s). Published by the American Society for Microbiology. This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial-NoDerivatives 4.0 International license (https://creativecommons.org/licenses/by-nc-nd/4.0/ and https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode), which grants the public the nonexclusive right to copy, distribute, or display the published work.

Appendix 1

Overview of the grandma hypothesis project-based learning activity

Session 1: Topic: introduction to the Human Microbiome Project, microbiome research and experimental design, research using human subjects and The Institutional Review Board (IRB). This assignment takes 30 minutes.

Session 2: Topic: clinical variables and microbiome sampling. Review clinical information and sampling protocol (Appendix 2). Assignments: collect clinical information, carry out microbial sampling, catalogue and preserve microbiome samples. This assignment takes 60-90 minutes.

Session 3: Topic: microbial DNA extraction. Assignments: extract microbial DNA as indicated in microbiome DNA extraction protocol at https://files.zymoresearch.com/protocols/_d4300t_d4300_d4304_zymobiomics_dna_miniprep_kit.pdf.

Note: other kits can be used instead of the Zymo kit. Frozen swabs were defrosted in ice. This assignment takes 60-90 minutes.

Session 4: Topic: molecular lab - PCR amplification. Assignments: amplify microbial 16S rRNA gene (see Appendix 3). This assignment takes 60 minutes.

Session 5: Topic: molecular lab - PCR quality control. Assignments: quality control of PCR products via gel electrophoresis (see Appendix 4). This assignment takes 60 minutes.

Session 6: Topic: molecular lab - 16S rRNA library preparation and DNA sequencing. Assignments: watch https://www.youtube.com/watch?v=_yC0Bzw3WbQ&t=71s and https://www.youtube.com/watch?v=t0akxx8Dwsk. This assignment takes 10 minutes.

Multiple molecular labs across USA and worldwide offer sequencing facilities. Google is the best tool to find them. Based on our personal experience we recommend the following three labs for metagenomic studies: The GWU Genomics Core (https://www.gwgenomics.org), the University of Michigan

(https://microbe.med.umich.edu/services/microbial-community-analysis) and Argonne National Laboratory (https://www.anl.gov).

Session 7: Topic: inferring sample composition from MiSeq amplicon data. Assignments: review R script dada2 pipeline tutorial at https://benjjneb.github.io/dada2/tutorial.html. This assignment takes 45 minutes.

Sessions 8 to 12: Topic: introduction to microbiome analysis. Assignments: review Microbiome Analyst and Microbiome Analyst tutorial: overview at https://www.microbiomeanalyst.ca/faces/home.xhtml. This assignment takes 30 minutes.

In sessions 10 and 12 students completed problem sets 1 and 2 in Appendix 5. Problem sets take 60 minutes each. Students are given class time to complete them and have two weeks to turn them in.

Session 13: Topic: data analysis and short scientific report (overview). Assignments: introduction to short scientific report writing (Appendix 6). This assignment takes 20 minutes.

Session 14: Topic: data analysis and short scientific report. Assignments: turn in 1st draft of short scientific report. Students have about a month to complete this assignment.

Session 15: Topic: peer-review short scientific report. Assignments: review short scientific report according to comments made by instructors. Once first drafts of short scientific report have been revised, students will have to submit the final version of their short scientific report in two weeks.

Appendix 2

Sampling

Time to completion: 60-90 minutes

Areas to be sampled from each individual in the following order:

1. Both forearms
2. Both calves
3. Behind both ears
4. Between toes
5. Navel

Materials:

Catch-All™ Sample Collection Swabs (Epicentre Biotechnologies, Madison, WI)

SCF-1 solution (50 mM Tris buffer [pH 7.6], 1 mM EDTA [pH 8.0], and 0.5% Tween-20)

Zymo DNA Isolation Kit

Bucket with ice

Gloves

Sampling Procedure

Each area must be sampled twice using two different swabs. Right and left sides are always sampled with the same swab. Both swabs are placed in the same Eppendorf tube.

1. Sampling

Rub all areas for 30 seconds and in between toes for 10 seconds each using sterile catch-all swabs moistened with SCF-1 solution.

See notes below

2. Insert the two swabs into the same correctly labeled tube. Aseptically cut the head of the swab from the handle using scissors and screw the tube cap back in place.

3. Store samples at -80 ºC

Notes

Forearms and calves: hold the shaft of the swab parallel to the surface of the skin and rub the swab back and forth along the area applying firm pressure for approximately 30 seconds.

Navel: introduce the swab all the way into the navel, hold the shaft of the swab and rub in circles gently but making sure both bottom and sides are swabbed for approximately 30 seconds.

Between toes: rub the skin between toes applying firm pressure for approximately 10 seconds each area.

Ear swabbing: to access the site, fold the ear forward with one hand to expose the crease. With the other hand, hold the shaft of the swab parallel to the surface of the skin and rub the swab back and forth along the crease approximately 50 times, applying firm pressure (this requires approximately 30 seconds).

Risks

The methods of specimen collection used in this project pose only minimal risk to the study subjects. These risks are described here and were also verbally explained to the participants. Skin: rubbing of the skin may cause slight irritation or transient redness at the sampling sites. Navel: there may be slight irritation at the sampling sites. It is possible that swabs include BSL2 microrganisms, but due to the lack of culturing, this should not be a concern.

Appendix 3

PCR protocol

Time to completion: 60 minutes

Reagents

1. Accuprime Pfx Supermix

2. Two IDT 16S V4 primers (forwards and reverse) at 10 uM concentration (see 1):

Forward: AATGATACGGCGACCACCGAGATCTACACCTGCGTGTTATGGTAATTGTGTGCCAGCM GCCGCGGTAA

Reverse: CAAGCAGAAGACGGCATACGAGATTCAGCGTTAGTCAGTCAGCCGGACTACHVGGGT WTCTAAT

3. DNA template

4. Mock community (positive control): D6305 at https://www.zymoresearch.com/collections/zymobiomics-microbial-community-standards

5. PCR grade H_2O

PCR well plate

1. Dispense 16 ul of PCR cocktail into each well of a new 96 well plate.

2. Transfer 1 ul of each barcoded forward and reverse primer to the corresponding well on the PCR plate.

3. Transfer 2 ul of template DNA per well to the corresponding well on the PCR plate.

4. Add 2 ul of mock community to the positive control well.

5. Add 2 ul of PCR grade H_2O to the negative control well.

6. Spin down contents.

7. Place in thermocycler (Applied Biosystems).

Thermocycler program

Use the following program.

Initial denaturation: 95°C for 2:00 minutes

Amplification: 30 cycles

Denaturation - 95°C for 20 seconds

Annealing - 55°C for 15 seconds

Extension - 72°C for 1 minute

Final extension: 72°C for 10:00 minutes

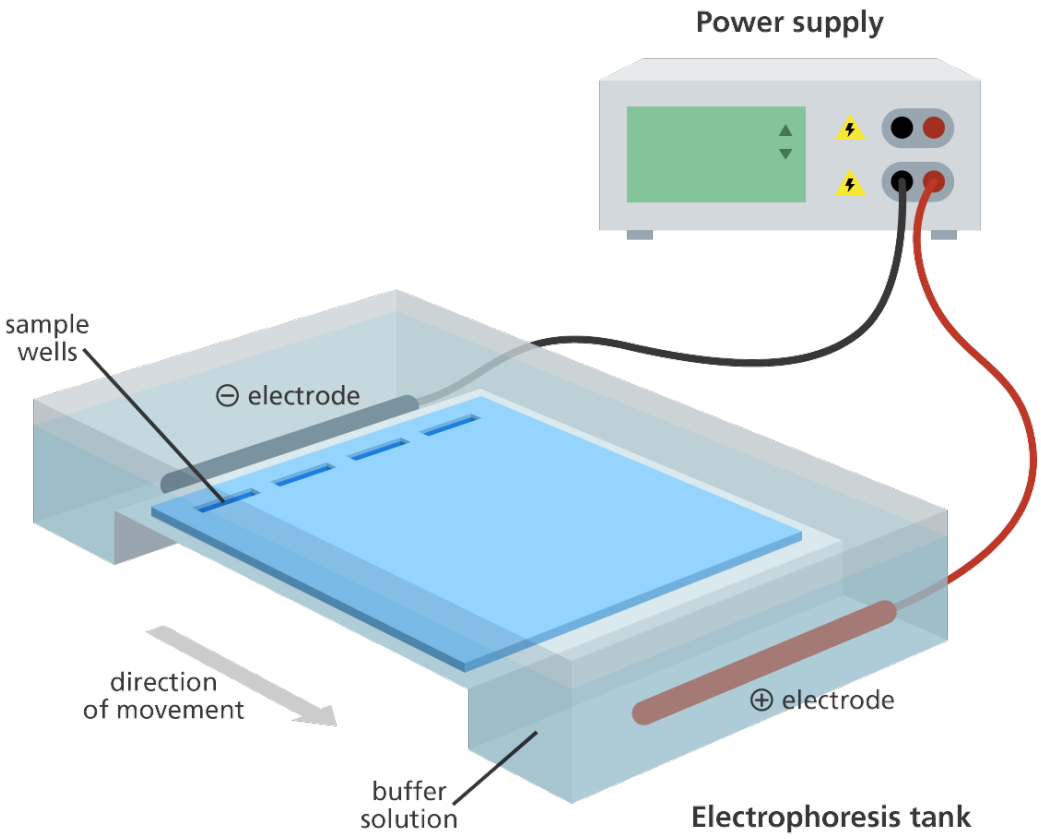
Hold: 4°C for ever

Appendix 4

Gel Electrophoresis

Time to completion: 60 minutes

Gel electrophoresis is a laboratory method used to separate mixtures of DNA, RNA, or proteins according to molecular size. In gel electrophoresis, the molecules to be separated are pushed by an electrical field through a gel that contains small pores. The molecules travel through the pores in the gel at a speed that is inversely related to their lengths. This means that a small DNA molecule will travel a greater distance through the gel than will a larger DNA molecule.



As previously mentioned, gel electrophoresis involves an electrical field; in particular, this field is applied such that one end of the gel has a positive charge and the other end has a negative charge. Because

DNA and RNA are negatively charged molecules, they will be pulled toward the positively charged end of the gel. Proteins, however, are not negatively charged; thus, when researchers want to separate proteins using gel electrophoresis, they must first mix the proteins with a detergent called sodium dodecyl sulfate. This treatment makes the proteins unfold into a linear shape and coats them with a negative charge, which allows them to migrate toward the positive end of the gel and be separated. Finally, after the DNA, RNA, or protein molecules have been separated using gel electrophoresis, bands representing molecules of different sizes can be detected.

Reagents:

1. Buffer: 1x TBE
2. Bio-Rad agarose: 1.5%
3. Bio-Rad gel red stain
4. Bio-Rad premixed loading buffer
5. Bio-Rad DNA 1 kb Ladder

Protocol:

Agarose gel preparation

Mix 50 ml of 1x TBE and 0.5 gr of agarose, heat for 2 minutes and stir every minute

Cool off under water in the sink

Add 3 ul of gel red

Agarose gel running

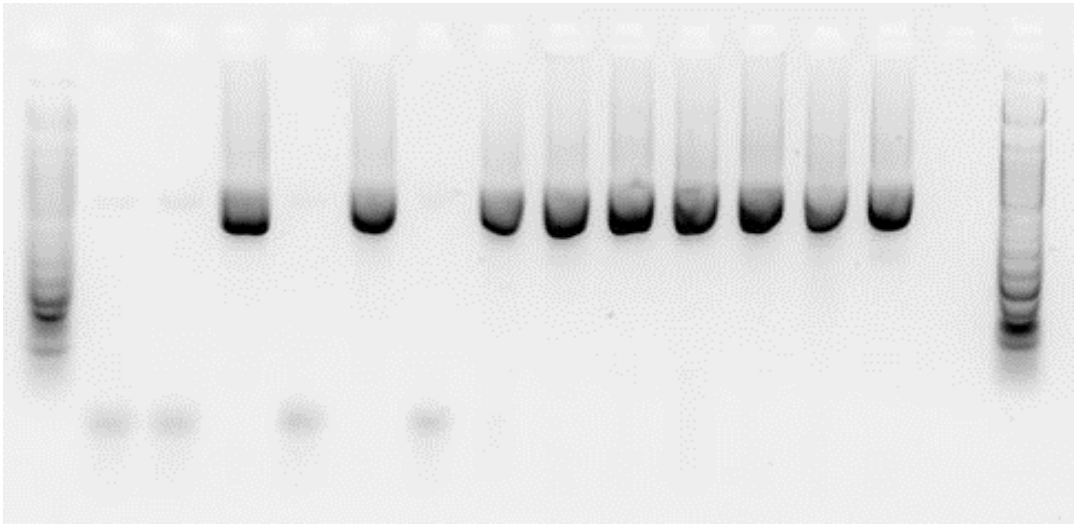
1. Pour in gel rig once warm, not hot, with comb in

2. Mix 3 ul of PCR product and 3 ul of loading buffer per sample
3. Load PCR product + loading buffer in each well
4. Load 3.5 ul of ladder
5. Run at 100v for 30 minutes
6. Photograph gel under UV

Agarose gel

Photograph of an agarose gel including two PCR ladders (wells 1 and 16), 13 PCR samples (wells 2 to 14) and one negative control (well 15).

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16



Appendix 5

Problem Set 1

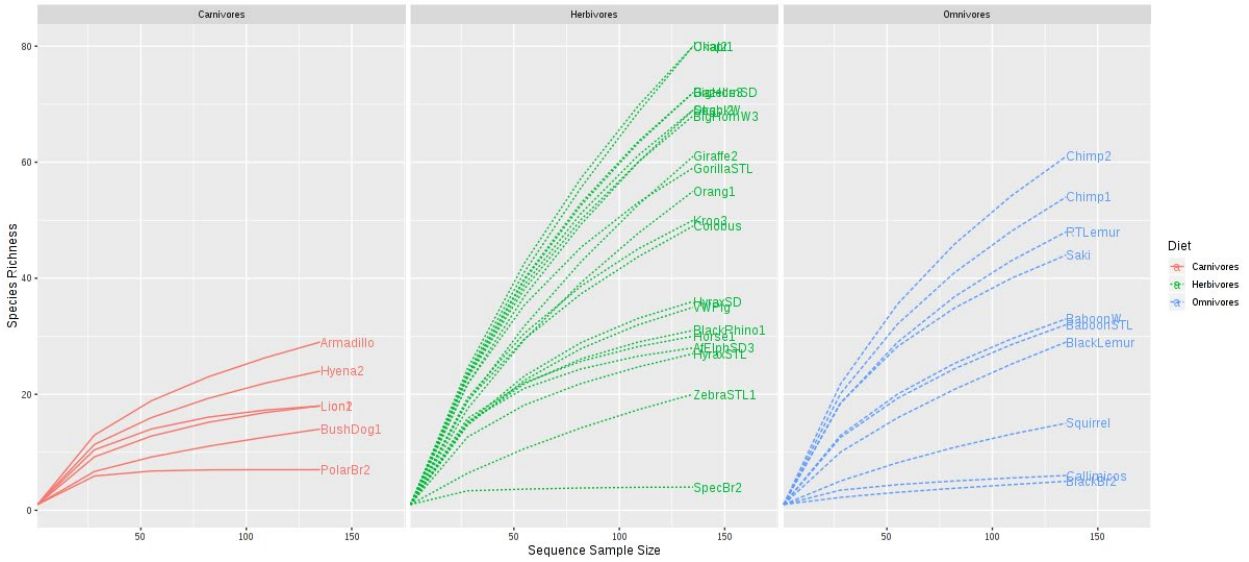
Time to completion: 60 minutes

This exercise is designed to explore differences in microbiome composition and diversity among 38 samples from different mammalian groups (herbivores, carnivores and omnivores) using 16S rRNA sequence data. The “mammalian dataset” is available at the Microbiome Analyst website. Data was processed using the Microbiome Analyst MDP tool and the SILVA database.

1. Data filtering: remove samples with <1000 reads first; then remove singletons; set `prevalence in samples` (%) and `percentage to remove` (%) with `inter-quantile range` to 10% each. How many features do remain in the data set after data filtering?

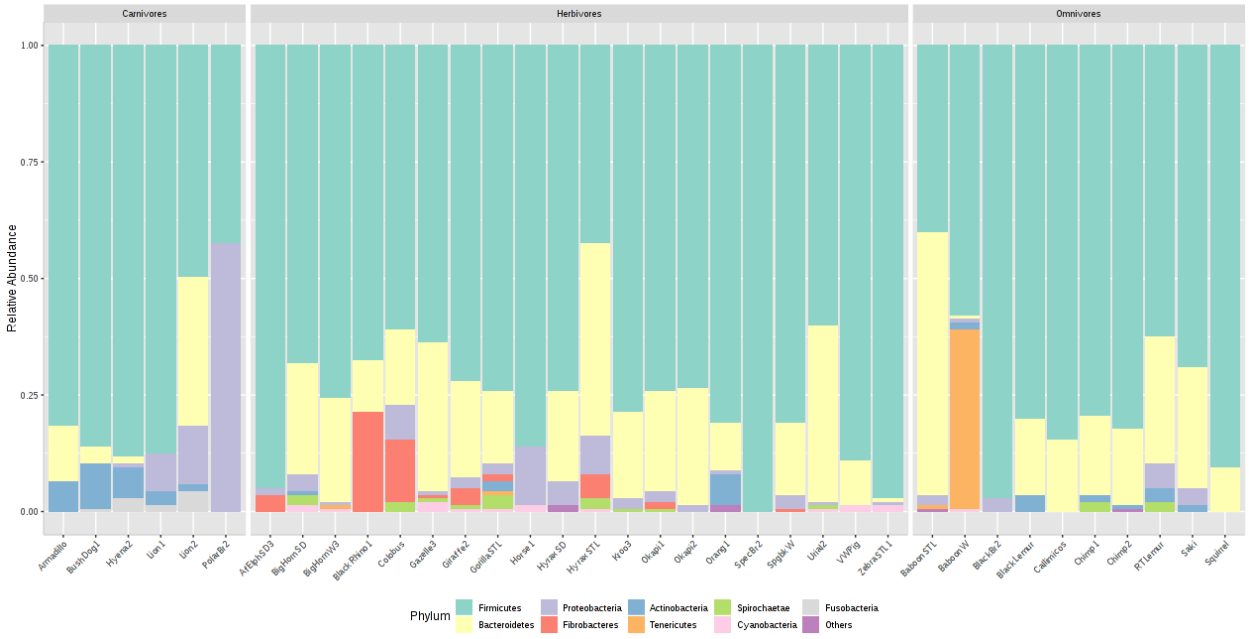
473 features remain in the data set after data filtering

2. Normalization: rarefy to the minimum library size; do not scale or transform the data. Make sure that samples with <1000 reads have been removed. Plot rarefaction curves for the three mammalian groups. Interpret plots.



Rarefaction plots show that carnivores curves have converged. Hence, the most common species were found, only the rarest species remain to be sampled. Herbivores and Omnivores curves, however, have not converged, so not all the common species have been found yet. Those latter samples should have to be further sequenced

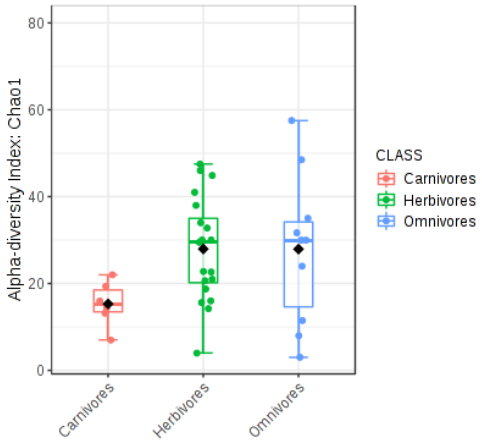
3. Create bar plots depicting Phyla relative abundances per sample as % for the three mammalian groups and paste below. What Phyla show a relative abundance >0.50 in any of the studied samples?



Proteobacteria and Bacteriodetes

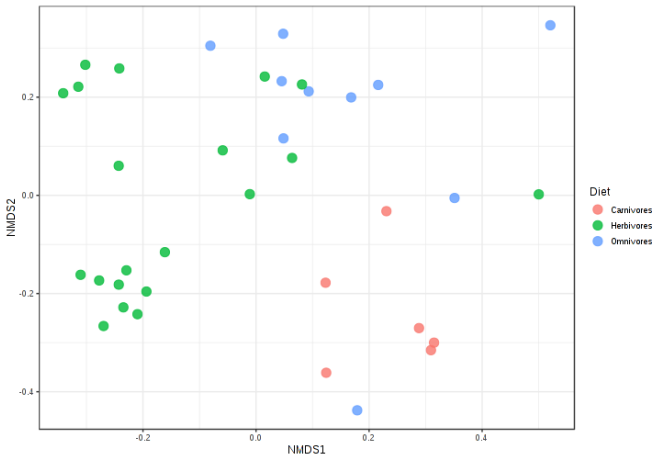
4. Estimate Chao1 alpha-diversity across mammalian groups at the species level. Which microbiome has the lowest diversity? Paste box plots below.

Carnivores



5. Estimate a NMDS plot at the OTU-level using the Jaccard Index Distance. Paste 2D plot and P value below. Do not show ellipses. Which microbiomes are more dissimilar?

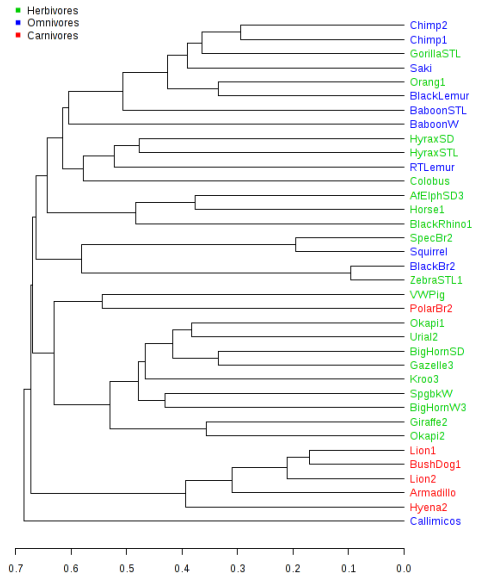
P<0.001



Herbivores and omnivores

6. Estimate an OTU dendrogram using the Jensen-Shannon Divergence Distance and the Average clustering algorithm. Paste below. Are the microbiomes fully segregated by mammalian group?

No



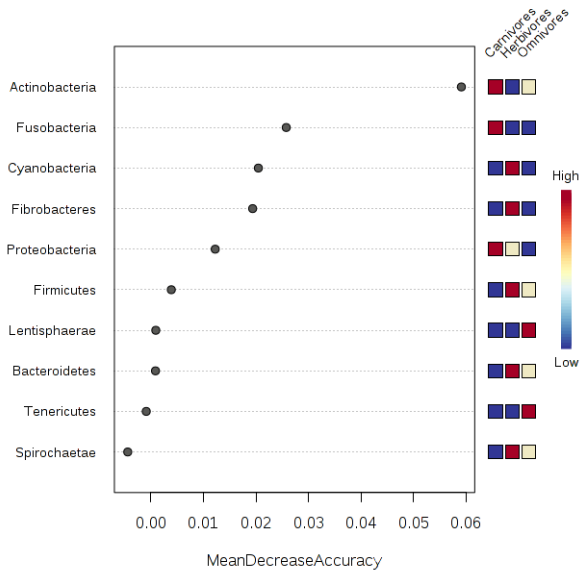
7. How many microbial Phyla show significant differences in abundance between mammalian groups after FDR correction for the Mann-Whitney/Kruskal-Wallis test at the 0.05 adjusted p-value cutoff?

Four

9. How many microbial Phyla show significant differences in abundance between mammalian group after FDR correction as indicated by the Linear Discriminant Analysis test at Log LDA score of 0.5 – use default settings for all other parameters?

One

10. Perform a random forest analysis at the Phylum level using 1000 trees. Paste plot below. Which Phylum contributes the most to the classification of the compared mammalian microbiomes.



Actinobacteria

Grading key: A fully correct answer is worth one point, hence the maximum score for the problem set is 10 points.

Problem Set 2

Time to completion: 60 minutes

This exercise is designed to explore differences in microbiome composition and diversity among 38 samples from different mammalian groups (herbivores, carnivores and omnivores) using 16S rRNA sequence data. The “mammalian dataset” is available at the Microbiome Analyst website. Data was processed using the Microbiome Analyst MDP tool and the SILVA database.

1. Data filtering: first remove samples with <500 reads; then remove singletons; set `prevalence in samples` (%) and `percentage to remove` (%) with `inter-quantile range` to 10% each. How many features remain in the data set after data filtering?

481 features remain in the data set after data filtering

2. Normalization: rarefy to the minimum library size; do not scale or transform the data. Make sure that samples with <500 reads have been already removed. What is the most abundant Phylum and Order?

Firmicutes and Clostridiales

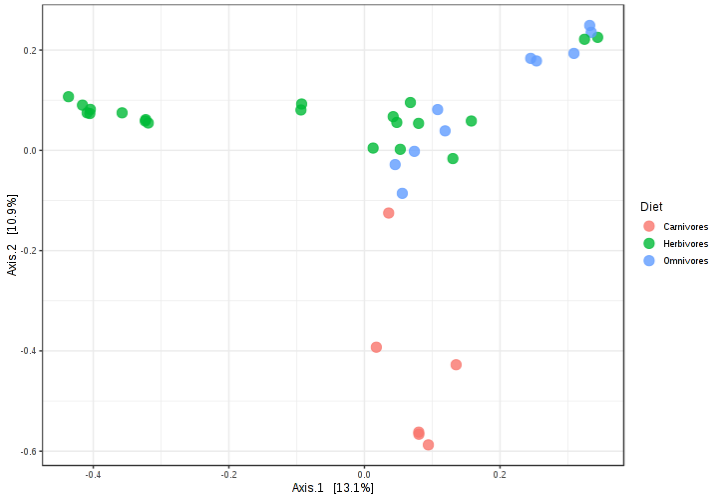
3. What mammalian species shows the highest abundance (percentage) of Proteobacteria?

Polar Br2

4. Estimate Shannon alpha-diversity index across mammalian groups at the OTU-level. Interpret P value.

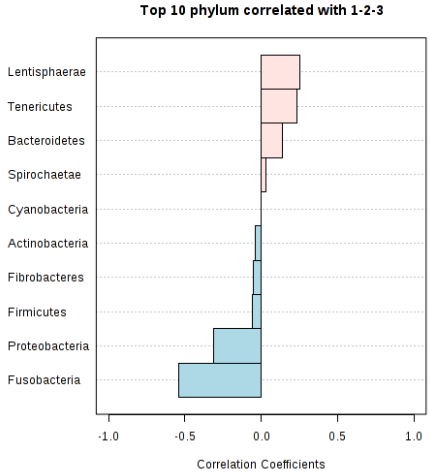
P values are < 0.05. Hence groups are significantly different in richness (OTU counts) and evenness (OUT distribution)

5. Using the PCoA ordination method, the Bray-Curtis Index Distance and the ANOSIM test determine if the beta-diversity varies significantly across mammalian groups at the OTU-level. Paste 2D PCoA plot and P value for the mammalian groups and interpret it.



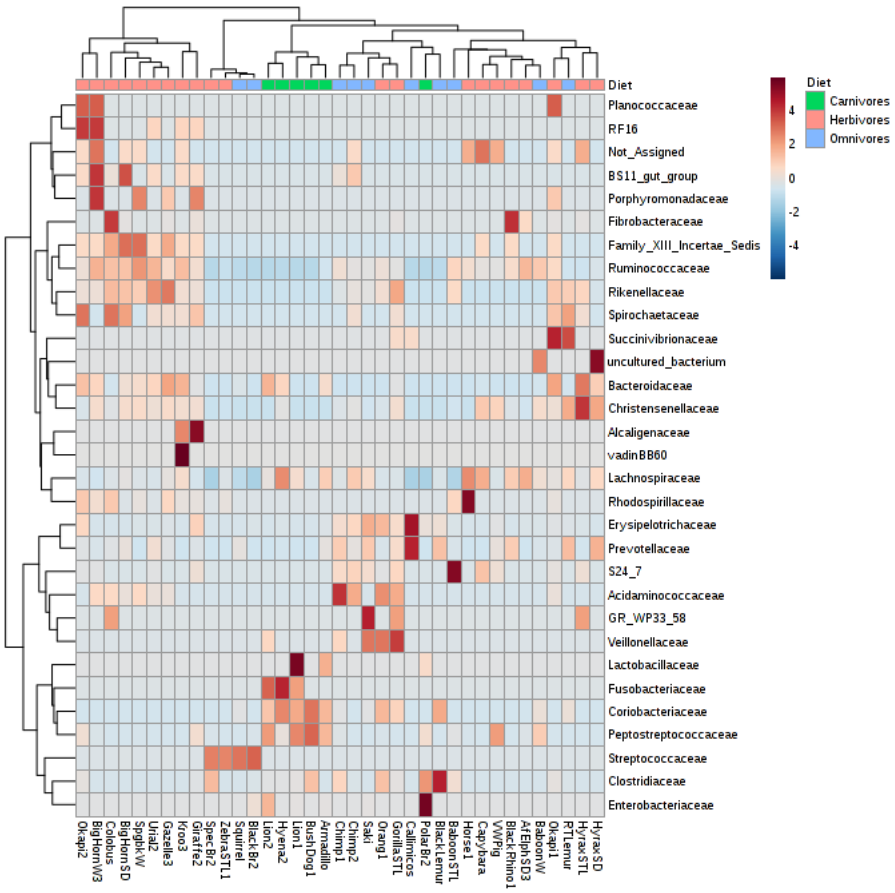
P<0.001. The plot shows significant differences in dissimilarity between groups. Herbivores and omnivores are less dissimilar to one another. Carnivores are the most dissimilar group of the three.

6. Perform a correlation analysis of the Phyla using the Pearson r correlation distance. Paste plot below. What microbial Phyla show the highest negative correlation?



Fusobacteria

7. Perform a heatmap clustering analysis of the microbial family using the Euclidean distance and the Ward clustering algorithm. Are the microbiomes fully separated (clustered) by mammalian group? Paste the heatmap below.



No

8. How many microbial Phyla show significant differences in abundance among mammalian groups after FDR correction for the ANOVA test at the 0.05 adjusted p-value cutoff?

Two

9. How many microbial Phyla show significant differences in abundance among mammalian groups after FDR correction as indicated by the Linear Discriminant Analysis test at Log LDA score of 0.5 – use default settings for all the other parameters?

Two

10. What microbial Phylum constitutes the core microbiome across all mammalian species for a sample prevalence of 90% and a relative abundance of 0.2%?

Firmicutes

Grading key: A fully correct answer is worth one point, hence the maximum score for the problem set is 10 points.

Appendix 6

Scientific Report

Time to completion: 2 weeks

GENERAL FORMAT

Five pages long

Pages 1 and 2 will include the main text. Pages 3 and 4 will include tables and figures. Page 5 will be for references only. Make sure tables and figures are legible.

Use single line spacing, Arial 11pt font and 0.5-inch margins.

Report should adhere to the format below. The following sections of the scientific report describe what you need to include and address in each of them. You will be evaluated and graded according to your ability to complete and address all of them. Instructors will review the reports according to the instructions below and email comments to students.

Pages 1 and 2

TITLE (2 points)

Include a short and catchy (if possible) title. The title is expected to summarize the aim of the study **(1 point)** and the main outcome **(1 point)** to score 2 points.

Name: your name

Submission date:

INTRODUCTION (4 points)

In a few sentences state the question you want to investigate and its importance **(1 point)**. Provide a conceptual framework for your study **(1 point)**. Then describe what knowledge or research is missing and needed in relation to that question (i.e., knowledge gap) and explain why addressing that knowledge gap

is important **(1 point)**. Then state the aim(s) of your study **(1 point)**. You can afterwards (optional) propose hypotheses - do not copy literally from the proposal. Include references.

MATERIALS AND METHODS (4 points)

Describe briefly how you are going to address the aim(s) above including the methods used to collect the samples, generate the sequences and analyze the data **(2 points)**. Include info about sampling and samples, molecular preparation, sequencing protocols, pipelines and statistical tools, taxonomic database, statistical tests (alpha-, beta-div and taxon comparisons), QC, normalization **(2 points)**. Make sure your methods are detailed enough to make the study reproducible and justify your methodological choices if needed.

RESULTS (5 points)

Describe the results of the data analysis. The results should address microbial composition **(1 point)**, alpha-diversity **(1 point),** beta-diversity **(1 point),** taxonomic differences **(1 point)** and core microbiome **(1 point)**. Do not repeat the information already displayed on figures or tables, instead summarize what that table or figure represents or means.

DISCUSSION (7 points)

Draw conclusions from each of the results above following the same order: microbial composition **(1 point)**, alpha-diversity **(1 point),** beta-diversity **(1 point),** taxonomic differences **(1 point)** and core microbiome **(1 point)**. Explain how your results relate to previous knowledge (i.e., peer-review publications) on the matter **(1 point)**. Do your results agree or disagree with previous information and how so? **(1 point)**.

CONCLUSIONS (2 points)

Summarize in two sentences the main conclusions about the microbiome questions you have studied and the results above (**1 point per sentence**). This is not a repetition of the results or discussion.

Pages 3 and 4

TABLES AND FIGURES (4 points)

Include 4 tables or figures to support your results and discussion. Cite tables and figures in the main text. Provide a short explanatory head (top) for each of them. **Each table and figure is worth 1 point**.

Pages 5

REFERENCES (2 points)

List all the references cited in the main text at the end. Use a minimum of 15 references **(1 point)** including at least 10 studies related to skin microbiome diversity **(1 point)**. See PubMed https://www.ncbi.nlm.nih.gov/pubmed/ for biomedical literature. Follow the PLoS One guidelines at https://journals.plos.org/plosone/s/submission-guidelines#loc-references for reference format and citation.

Rubric: A maximum score on the report is 40 points. Ten points are given for just presenting a full draft by session 14 (Appendix 1). Once that first draft is returned with comments from the instructors, students have two weeks to submit the final version. There is a maximum of 30 points if the final draft is submitted on time. Those 30 points are distributed as indicated above.