The Impact of Added Non-Protein Nitrogen on the Rumen Microbiome and Fermentation of Substrates

Mia Carmichael

2019-11-13

## Summary/Abstract

*Write a summary of your project.*

## Introduction (required for part 1)

# General Background Information

The symbiotic relationship between ruminants and their ruminal microbial population allows cattle to occupy an ecological niche in which they have the ability to convert fibrous feedstuffs into higher quality end products via fermentation. Ruminants themselves do not produce fiber degrading enzymes, but they harbor bacteria, fungi, and protozoa that can [1]. According to [2], feed inputs are the largest variable costs associated with producing beef, as they are directly related to 75% of the cost of producing finished cattle [3]. Continuing to improve animal efficiency can result in a decrease in feeding costs for producers and decrease environmental impacts. Residual Average Daily Gain (RADG) is one of the latest feed efficiency selection tools adopted by the American Angus Association in their selection programs. RADG is heritable, and it can be a predictor of a sire’s genetic ability for postweaning gain in future progeny compared to that of other sires, given a constant amount of feed consumed (American Angus Association, 2015). This allows producers to select for genetics that increase the efficiency of finishing cattle in a feedlot environment. Previous studies have shown de association between the ruminal microbiota and cattle feed efficiency (Hernandez-Sanabria et al., 2012; Shabat et al., 2016), so the current study was performed to further understand this relationship. More specifically, the objective of this study was to assess the relationship between cattle feed efficiency during the feedlot-finishing phase and their ruminal microbiotas at two distinct time points: at the beginning and end of the feedlot phase.

# Description of data and data source

This study was conducted as a smaller portion of a larger study [4]. Briefly, for 5 offspring generations, commercial Angus cows were bred to determine the effect of selection based on RADG and marbling expected progeny differences (EPDs) on animal productivity, performance, and carcass quality. The current study utilized the fifth generation of steers born in that selection program and analyzed their performance in a feedlot-finishing system. The study was conducted on a commercial feedlot (Ridgefield Farm L.L.C.; Brasstown, North Carolina; 35.0391° N, 83.9576°W) from March 8, 2018 to May 29, 2018. A total of 65 steers were used, and of those 65 steers, 12 were selected based on their predicted EPDs. Of those 12 steers, the 5 most efficient and 5 least efficient according to residual feed intake (RFI) were selected for analysis of their ruminal microbiomes. Data was further processed g # Questions/Hypotheses to be addressed In my project I am trying to answer what changes occur in the rumen microbiome of a feedlot cow on a high starch diet. It will be interesting to see the trends and how the microbial profile and end products change as the nutrient availability changes. In the rumen there are multiple classes of microbes, including fiber degrading, starch degrading, and methane producing bacteria. As the feedstuff changes, the abundance of specific microorganisms will as well, which dictate other outcomes as well, pH, VFA profile, ammonia production, and methane production. The implications of this study can aid in further manipulation of the diet and cow to improve feed efficiency, production costs, and envirornmental outputs.

## Methods and Results

Animals, experimental design, and management

All procedures involving animals were verified and approved by the University of Georgia’s Office of Animal Care and Use (AUP #A2012 11-006-R1).

The experimental period had a duration of 82 days. Prior to that, steers were adapted to the finishing diet for 3 weeks. The diets were formulated to meet all nutrient requirements for growing cattle according to the NRC [5]. The composition of those diets is shown in Table 1. During the entire study, steers were fed using a GrowSafe Systems® (GrowSafe Systems Ltd., Calgary, Canada) for individual collection of feed intake. Steers were weighted at the beginning, mid-point, and end of the experimental period. Because the individual dry matter intake (DMI) and body weight gains were known for each steer, computation of RFI was possible. RFI was calculated as the difference between actual and expected DMI using linear regression of midpoint metabolic body weight and average daily gain (ADG).

Sample collection, DNA extraction and sequencing

Ruminal samples were collected 2 times – upon entering and exiting the feedlot – using the methodology described by [6]. All samples were immediately refrigerated and transported to the laboratory, where they were frozen and stored at -20°C until further processing. DNA was extracted from the rumen fluid samples according to the semi-automated extraction protocol described by [7]. Briefly, the procedure started with 0.5 mL of homogenized rumen fluid and achieved extraction of DNA through a combination of mechanic and enzymatic methods. The mechanic portion of the protocol used a FastPrep 24 Instrument (MP Biomedicals LLC, Irvine, CA, USA) to disrupt the cells. The enzymatic portion used InhibitEX® Tablets (QIAGEN, Venlo, Netherlands). Elution and purification of DNA were performed using an automated robotic workstation (QIAcube; QIAGEN, Venlo, Netherlands). Determination of DNA concentration and purity were performed spectrophotometrically using the Synergy™ H4 Hybrid Multi-Mode Microplate Reader (BioTek Instruments Inc; Winooski, VT, USA). After DNA extraction, samples were taken to the Georgia Genomics and Bioinformatics Core facility (<https://dna.uga.edu>) for library preparation and 16S rRNA gene sequencing. Library preparation included PCR replications using the forward: S-D-Bact-0341-b-S-17 (5′-CCTACGGGNGGCWGCAG-3′); and reverse: S-D-Bact-0785-a-A-21 (5′-GACTACHVGGGTATCTAATCC-3′) primer pair [8]. Samples were sequenced using the Illumina MiSeq system (Illumina Inc., San Diego, CA, USA).

Sequencing data

Sequencing data were demultiplexed and downloaded as FASTQ files. Pair-end reads were merged using BBMerge Paired Read Merger v37.64, and files were further analyzed using the QIIME pipeline v1.9.1 [9]. Files were converted into FASTA files and quality-filtered. The sequences were clustered into operational taxonomic units (OTUs) at 97% similarity according to the Greengenes database. Singleton OTUs were excluded from the analysis. A sampling depth of 28,662 sequences per sample was used for the diversity analyses. The computed alpha-diversity indexes were: number of observed OTUs, Chao1, species evenness, Shannon index, and Faith’s phylogenetic diversity.

Statistical Analysis

## Data aquisition

The data described was collected from a total of 65 steers, and of those 65 steers, 12 were selected based on their predicted EPDs. Of those 12 steers, the 5 most efficient and 5 least efficient according to residual feed intake (RFI) were selected for analysis of their ruminal microbiomes.

## Data import and cleaning

Here I have imported and cleaned data from the microbiome (phylum, family, and genus) of High and Low RFI steers, as well as their alpha diversity and p-values.

Processing Script: This script loads the raw data, processes and cleans it and saves it as Rds file in the processed\_data folder ` Load needed packages. make sure they are installed.

library(readxl)  
library(readr)  
library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(here)

## here() starts at /Users/miacarmichael/Desktop/epid8060fall2019/MiaCarmichael-Project

library(visdat)  
library(dplyr)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ tibble 2.1.3 ✔ purrr 0.3.2  
## ✔ tidyr 1.0.0 ✔ stringr 1.4.0  
## ✔ tibble 2.1.3 ✔ forcats 0.4.0

## ── Conflicts ────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

#Phylum:High RFI

Load data, path is relative to project directory. Select a file and read the data into a data frame.

phyladata <- read\_csv(here("data", "raw\_data", "Abundant\_Phyla\_MADA.csv"))

## Parsed with column specification:  
## cols(  
## `Bacterial Phyla` = col\_character(),  
## Beginning = col\_double(),  
## End = col\_double(),  
## `P-value` = col\_double()  
## )

Take a look at the data

head(phyladata)

## # A tibble: 6 x 4  
## `Bacterial Phyla` Beginning End `P-value`  
## <chr> <dbl> <dbl> <dbl>  
## 1 Firmicutes 35.0 55.1 0.02  
## 2 Bacteroidetes 48.7 33.0 0.12  
## 3 Actinobacteria 7.21 1.29 0.34  
## 4 TM7 1.81 2.06 0.85  
## 5 Proteobacteria 1.48 2.31 0.43  
## 6 Euryarchaeota 1.28 1.39 0.83

dplyr::glimpse(phyladata)

## Observations: 13  
## Variables: 4  
## $ `Bacterial Phyla` <chr> "Firmicutes", "Bacteroidetes", "Actinobacteria…  
## $ Beginning <dbl> 34.95, 48.73, 7.21, 1.81, 1.48, 1.28, 1.31, 0.…  
## $ End <dbl> 55.14, 33.04, 1.29, 2.06, 2.31, 1.39, 0.81, 0.…  
## $ `P-value` <dbl> 0.02, 0.12, 0.34, 0.85, 0.43, 0.83, 0.42, 0.74…

There are only 11 identified Phylum in this data, but there are currently 13 observations. Remove any observations with NA

phyladata <- drop\_na(phyladata)  
dim(phyladata)

## [1] 11 4

Rename the columes, so when creating a graph things can be plotted easier

names(phyladata)[1:4]<-c("Phyla", "Beginning", "End", "P")

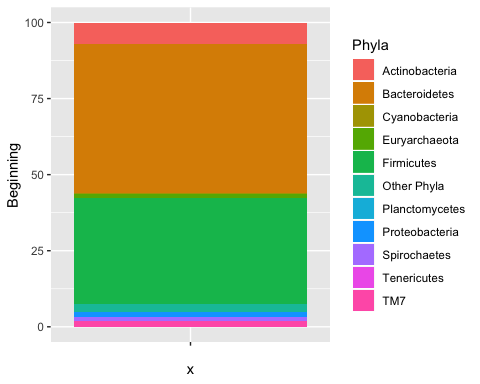
Check name of columns to see what data looks like.

names(phyladata)

## [1] "Phyla" "Beginning" "End" "P"

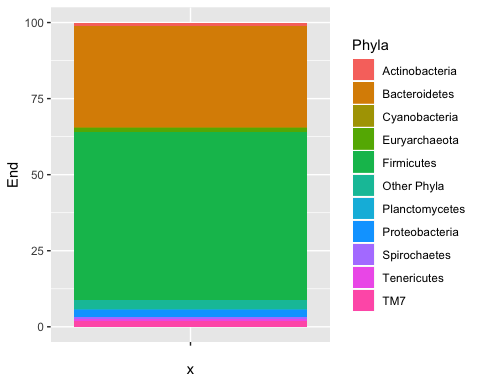
An initial plot is created showing the composition of phyla at the beginning of the feedlot period in High RFI/low efficiency steers. Save plot to a file.

ggplot(phyladata,aes(x="", y=Beginning, fill=Phyla))+ geom\_bar(width = 1, stat = "identity")



A second plot is created showing the composition of families at the end of the feedlot period in High RFI/low efficiency steers.Save the plot to a file.

ggplot(phyladata,aes(x="", y=End, fill=Phyla))+ geom\_bar(width = 1, stat = "identity")



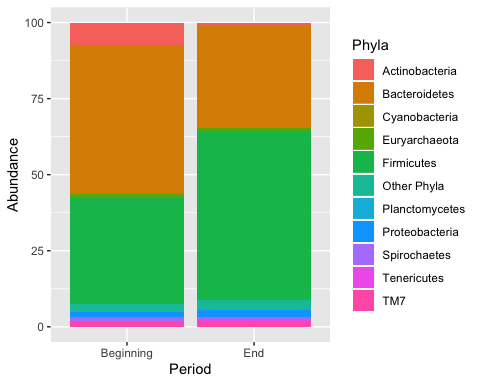
The axis labels aren’t quite right for both graphs. That should be considered when creating them.

It is easier to see these plots side by side. To do this, create subsets to divide by period, “Beginning” and “End”.

#Create datasets   
  
Period <- c(rep("Beginning",11), rep("End",11))  
  
Phyla <- rep(c("Firmicutes", "Bacteroidetes",   
"Actinobacteria", "TM7", "Proteobacteria", "Euryarchaeota", "Spirochaetes", "Cyanobacteria",   
"Tenericutes", "Planctomycetes", "Other Phyla"),2)   
  
Abundance <- c(34.95, 48.73, 7.21, 1.81, 1.48, 1.28, 1.31, 0.39, 0.16, 0.21, 2.46, 55.14, 33.04, 1.29, 2.06, 2.31, 1.39, 0.81, 0.33, 0.46, 0.27, 2.91)  
  
  
   
#Combine datasets   
phylacomb <- data.frame(Period, Phyla, Abundance)  
  
#Remove missing data  
phylacomb <- drop\_na(phylacomb)  
dim(phylacomb)

## [1] 22 3

#Graph combined dataset  
ggplot(phylacomb, aes(fill=Phyla, y=Abundance, x=Period)) +  
 geom\_bar(stat = "identity")

 The graph above displays the composition of phyla over the course of the feedlot period in High RFI/low efficiency steers Repeat for Family and Genus

#Phylum: Low RFI

Load data, path is relative to project directory. Select a file and read the data into a data frame.

phyladata2 <- read\_csv(here("data", "raw\_data", "Phyla\_Low\_MADA.csv"))

## Parsed with column specification:  
## cols(  
## `Bacterial Phyla` = col\_character(),  
## Beginning = col\_double(),  
## End = col\_double(),  
## `P-value` = col\_double()  
## )

Take a look at the data

head(phyladata2)

## # A tibble: 6 x 4  
## `Bacterial Phyla` Beginning End `P-value`  
## <chr> <dbl> <dbl> <dbl>  
## 1 Firmicutes 35.0 55.1 0.02  
## 2 Bacteroidetes 48.7 33.0 0.12  
## 3 Actinobacteria 7.21 1.29 0.34  
## 4 TM7 1.81 2.06 0.85  
## 5 Proteobacteria 1.48 2.31 0.43  
## 6 Euryarchaeota 1.28 1.39 0.83

dplyr::glimpse(phyladata2)

## Observations: 11  
## Variables: 4  
## $ `Bacterial Phyla` <chr> "Firmicutes", "Bacteroidetes", "Actinobacteria…  
## $ Beginning <dbl> 34.95, 48.73, 7.21, 1.81, 1.48, 1.28, 1.31, 0.…  
## $ End <dbl> 55.14, 33.04, 1.29, 2.06, 2.31, 1.39, 0.81, 0.…  
## $ `P-value` <dbl> 0.02, 0.12, 0.34, 0.85, 0.43, 0.83, 0.42, 0.74…

There are only 11 identified Phylum in this data, but there are currently 13 observations. Remove any observations with NA

phyladata2 <- drop\_na(phyladata2)  
dim(phyladata2)

## [1] 11 4

Rename the columes, so when creating a graph things can be plotted easier

names(phyladata2)[1:4]<-c("Phyla", "Beginning", "End", "P")

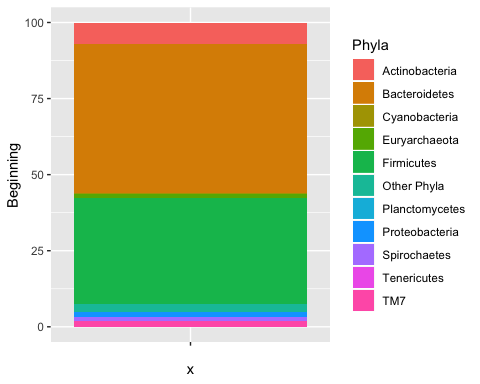
Check name of columns to see what data looks like.

names(phyladata2)

## [1] "Phyla" "Beginning" "End" "P"

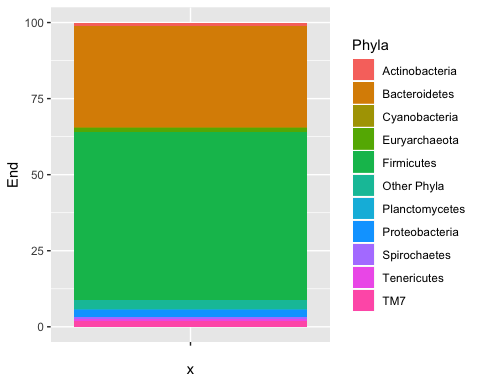
An initial plot is created showing the composition of phyla at the beginning of the feedlot period in Low RFI/High efficiency steers. Save plot to a file.

ggplot(phyladata2,aes(x="", y=Beginning, fill=Phyla))+ geom\_bar(width = 1, stat = "identity")



A second plot is created showing the composition of families at the end of the feedlot period in Low RFI/High efficiency steers.Save the plot to a file.

ggplot(phyladata2,aes(x="", y=End, fill=Phyla))+ geom\_bar(width = 1, stat = "identity")



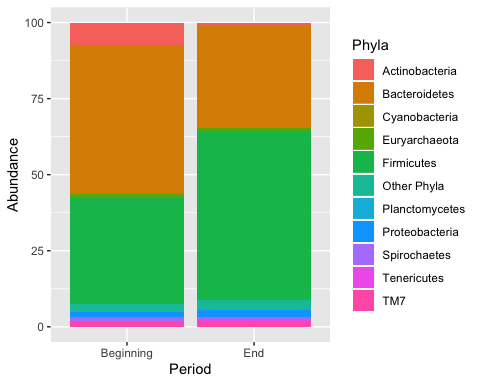
The axis labels aren’t quite right for both graphs. That should be considered when creating them.

It is easier to see these plots side by side. To do this, create subsets to divide by period, “Beginning” and “End”.

#Create datasets   
  
Period <- c(rep("Beginning",11), rep("End",11))  
  
Phyla <- rep(c("Firmicutes", "Bacteroidetes",   
"Actinobacteria", "TM7", "Proteobacteria", "Euryarchaeota", "Spirochaetes", "Cyanobacteria",   
"Tenericutes", "Planctomycetes", "Other Phyla"),2)   
  
Abundance <- c(34.95, 48.73, 7.21, 1.81, 1.48, 1.28, 1.31, 0.39, 0.16, 0.21, 2.46, 55.14, 33.04, 1.29, 2.06, 2.31, 1.39, 0.81, 0.33, 0.46, 0.27, 2.91)  
  
  
   
#Combine datasets   
phylacomb2 <- data.frame(Period, Phyla, Abundance)  
  
#Remove missing data  
phylacomb2 <- drop\_na(phylacomb)  
dim(phylacomb2)

## [1] 22 3

#Graph combined dataset  
ggplot(phylacomb2, aes(fill=Phyla, y=Abundance, x=Period)) +  
 geom\_bar(stat = "identity")



The graph above displays the composition of phyla over the course of the feedlot period in Low-RFI/High efficiency steers Repeat for Family and Genus

#Family: High RFI Load data, path is relative to project directory. Select a file and read the data into a data frame.

familydata <- read\_csv(here("data", "raw\_data", "Abundant\_FamiliesHighRFI\_MADA.csv"))

## Parsed with column specification:  
## cols(  
## Family = col\_character(),  
## Beginning = col\_double(),  
## End = col\_double(),  
## `P-VALUE` = col\_double()  
## )

Take a look at the data

head(familydata)

## # A tibble: 6 x 4  
## Family Beginning End `P-VALUE`  
## <chr> <dbl> <dbl> <dbl>  
## 1 Ruminococcaceae 26.0 40.3 0.17   
## 2 Prevotellaceae\* 36.9 14.3 0.05   
## 3 Order Clostridiales 5.93 8.40 0.08   
## 4 Order Bacteroidales 5.94 7.49 0.15   
## 5 Lachnospiraceae 5.19 7.70 0.290  
## 6 RF16 1.94 1.36 0.14

dplyr::glimpse(familydata)

## Observations: 18  
## Variables: 4  
## $ Family <chr> "Ruminococcaceae", "Prevotellaceae\*", "Order Clostridi…  
## $ Beginning <dbl> 25.9771285, 36.9109508, 5.9340251, 5.9409425, 5.189480…  
## $ End <dbl> 40.3397858, 14.3071369, 8.4018357, 7.4860880, 7.701011…  
## $ `P-VALUE` <dbl> 0.17, 0.05, 0.08, 0.15, 0.29, 0.14, 0.01, 0.09, 0.49, …

There are only 18 identified Phylum in this data, but there are currently observations. Remove any observations with NA

familydata <- drop\_na(familydata)  
dim(familydata)

## [1] 18 4

Rename the columes, so when creating a graph things can be plotted easier

names(familydata)[1:4]<-c("Family", "Beginning", "End", "P")

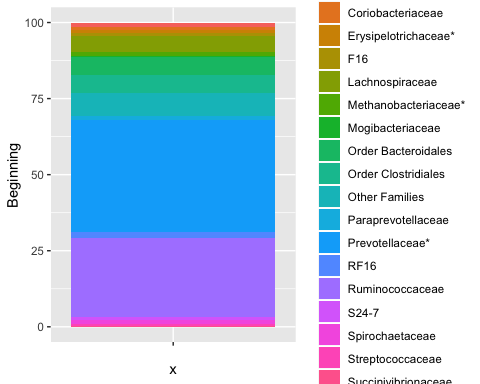
Check name of columns to see what data looks like.

names(familydata)

## [1] "Family" "Beginning" "End" "P"

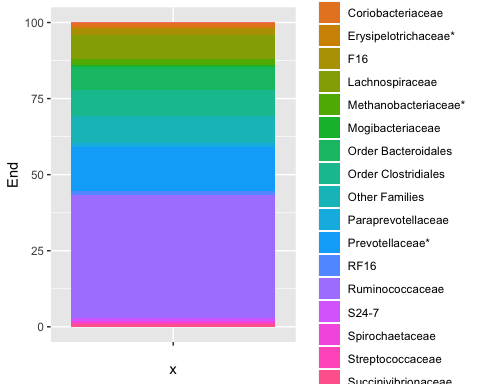
An initial plot is created showing the composition of families at the beginning of the feedlot period in High RFI/low efficiency steers. Save plot to a file.

ggplot(familydata,aes(x="", y=Beginning, fill=Family))+ geom\_bar(width = 1, stat = "identity")



A second plot is created showing the composition of families at the end of the feedlot period in High RFI/low efficiency steers.Save the plot to a file.

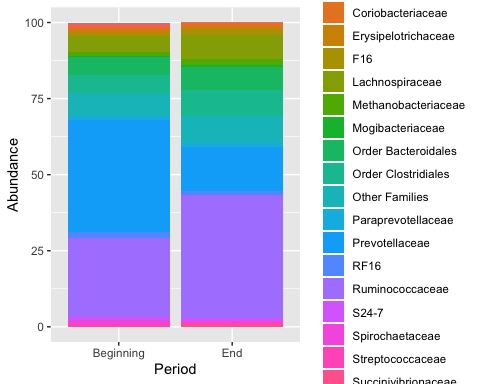
ggplot(familydata,aes(x="", y=End, fill=Family))+ geom\_bar(width = 1, stat = "identity")



The axis labels aren’t quite right for both graphs. That should be considered when creating them.

It is easier to see these plots side by side. To do this, create subsets to divide by period, “Beginning” and “End”.

#Create datasets   
  
Period <- c(rep("Beginning",18), rep("End",18))  
  
Family <- rep(c("Ruminococcaceae", "Prevotellaceae", "Order Clostridiales", "Order Bacteroidales", "Lachnospiraceae", "RF16", "Methanobacteriaceae", "F16", "Paraprevotellaceae", "S24-7", "Coriobacteriaceae", "Erysipelotrichaceae", "Bifidobacteriaceae", "Spirochaetaceae", "Streptococcaceae", "Mogibacteriaceae", "Succinivibrionaceae", "Other Families"),2)   
  
Abundance <- c(25.97712848, 36.91095082, 5.934025064, 5.940942506, 5.189480948, 1.937621255, 1.235079234, 1.049483052, 1.214503215, 0.956874771, 0.860283125, 1.158815891, 1.451193422, 1.135264464, 0.802975589, 0.476774167, 0.255410691, 7.513193298, 40.33978584, 14.30713689, 8.401835665, 7.486088, 7.701011161, 1.357351894, 2.004700597, 2.111417947, 1.348409159, 1.177308762, 1.058027748, 0.733590889, 0.311866263, 0.47494151, 0.461737838, 0.773816017, 0.916878216, 9.034095609)  
   
#Combine datasets   
familycomb <- data.frame(Period, Family, Abundance)  
  
#Graph combined dataset  
ggplot(familycomb, aes(fill=Family, y=Abundance, x=Period)) +  
 geom\_bar(stat = "identity")

 The graph above displays the abundance of families over the course of the feedlot period in High RFI/low efficiency steers

#Family: Low RFI

Load data, path is relative to project directory. Select a file and read the data into a data frame.

familydata2 <- read\_csv(here("data", "raw\_data", "Family\_Low\_MADA.csv"))

## Parsed with column specification:  
## cols(  
## `Bacterial Family` = col\_character(),  
## Beginning = col\_double(),  
## End = col\_double(),  
## `P-VALUE` = col\_double()  
## )

Take a look at the data

head(familydata2)

## # A tibble: 6 x 4  
## `Bacterial Family` Beginning End `P-VALUE`  
## <chr> <dbl> <dbl> <dbl>  
## 1 Ruminococcaceae\* 20.1 39.3 0.04  
## 2 Prevotellaceae\* 38.0 20.1 0.05  
## 3 Order Bacteroidales 6.84 8.47 0.53  
## 4 Order Clostridiales 5.51 7.12 0.37  
## 5 Lachnospiraceae 5.90 5.86 0.98  
## 6 Bifidobacteriaceae 6.23 0.317 0.32

dplyr::glimpse(familydata2)

## Observations: 17  
## Variables: 4  
## $ `Bacterial Family` <chr> "Ruminococcaceae\*", "Prevotellaceae\*", "Order…  
## $ Beginning <dbl> 20.1494728, 38.0286712, 6.8358601, 5.5082703,…  
## $ End <dbl> 39.2707821, 20.0930152, 8.4717829, 7.1198175,…  
## $ `P-VALUE` <dbl> 0.0400000, 0.0500000, 0.5300000, 0.3700000, 0…

There are only 17 identified families in this data, but there are currently observations. Remove any observations with NA

familydata2 <- drop\_na(familydata2)  
dim(familydata2)

## [1] 17 4

Rename the columes, so when creating a graph things can be plotted easier

names(familydata2)[1:4]<-c("Family", "Beginning", "End", "P")

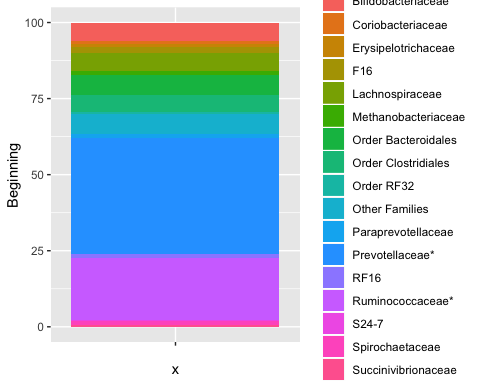
Check name of columns to see what data looks like.

names(familydata2)

## [1] "Family" "Beginning" "End" "P"

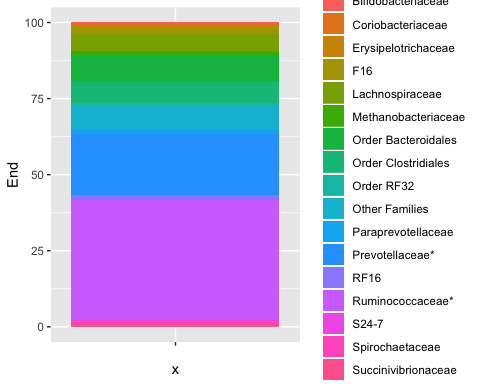
An initial plot is created showing the composition of families at the beginning of the feedlot period in Low RFI/High efficiency steers. Save plot to a file.

ggplot(familydata2,aes(x="", y=Beginning, fill=Family))+ geom\_bar(width = 1, stat = "identity")



A second plot is created showing the composition of families at the end of the feedlot period in Low RFI/High efficiency steers.Save the plot to a file.

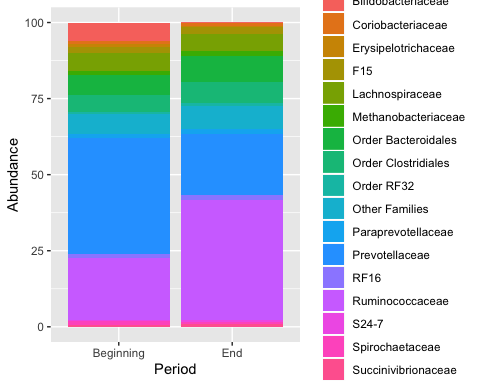
ggplot(familydata2,aes(x="", y=End, fill=Family))+ geom\_bar(width = 1, stat = "identity")



The axis labels aren’t quite right for both graphs. That should be considered when creating them.

It is easier to see these plots side by side. To do this, create subsets to divide by period, “Beginning” and “End”.

#Create datasets   
  
Period <- c(rep("Beginning",17), rep("End",17))  
  
Family <- rep(c("Ruminococcaceae", "Prevotellaceae", "Order Bacteroidales", "Order Clostridiales", "Lachnospiraceae", "Bifidobacteriaceae", "F15", "RF16", "Paraprevotellaceae", "Methanobacteriaceae", "Spirochaetaceae", "Erysipelotrichaceae", "Coriobacteriaceae", "Succinivibrionaceae", "Order RF32", "S24-7", "Other Families"),2)   
  
Abundance <- c(20.1494728, 38.02867118, 6.835860087, 5.508270337, 5.904255838, 6.231525756, 1.812183786, 1.50086736, 1.331658729, 1.280539877, 1.254985451, 1.199951243, 0.707032815, 0.587611249, 0.446050088, 0.49438202, 6.726681384, 39.27078212, 20.09301518, 8.471782946, 7.119817496, 5.863717733, 0.317316787, 2.053210621, 1.752576479, 1.496544772, 1.392298259, 0.649711389, 0.529057072, 0.757975929, 0.820005408, 0.932485624, 0.862008182, 7.617694)  
   
#Combine datasets   
familycomb2 <- data.frame(Period, Family, Abundance)  
  
#Graph combined dataset  
ggplot(familycomb2, aes(fill=Family, y=Abundance, x=Period)) +  
 geom\_bar(stat = "identity")

 The graph above displays the abundance of families over the course of the feedlot period in Low RFI/High efficiency steers

Notice that the abundance and composition of families differs between Low-RFI and High-RFI steers

Repeat for Genus

#Genus: High RFI For my dataset, there was litte variation over the course of the feedlot amongst genra. Howver there were noticible changes amongst methanogen populations.

Load data, path is relative to project directory. Select a file and read the data into a data frame .

genusdata <- read\_csv(here("data", "raw\_data", "Genera\_MADA.csv"))

## Parsed with column specification:  
## cols(  
## `HIGH-RFI` = col\_character(),  
## Beginning = col\_double(),  
## End = col\_double(),  
## `P-VALUE` = col\_double()  
## )

Take a look at the data

head(genusdata)

## # A tibble: 2 x 4  
## `HIGH-RFI` Beginning End `P-VALUE`  
## <chr> <dbl> <dbl> <dbl>  
## 1 Methanobrevibacter 1.2 2 0.01  
## 2 Methanosphaera 0.01 0.03 0.09

dplyr::glimpse(genusdata)

## Observations: 2  
## Variables: 4  
## $ `HIGH-RFI` <chr> "Methanobrevibacter", "Methanosphaera"  
## $ Beginning <dbl> 1.20, 0.01  
## $ End <dbl> 2.00, 0.03  
## $ `P-VALUE` <dbl> 0.01, 0.09

Remove any observations with NA

genusdata <- drop\_na(genusdata)  
dim(genusdata)

## [1] 2 4

Rename the columes, so when creating a graph things can be plotted easier

names(genusdata)[1:4]<-c("Genus", "Beginning", "End", "P")

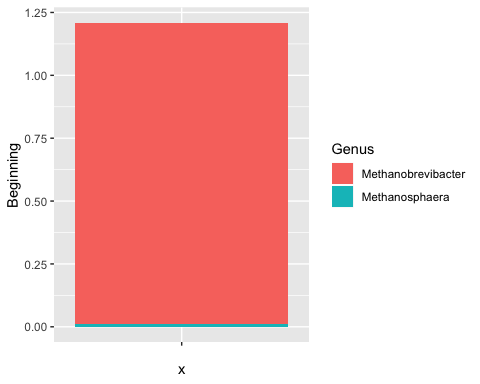
Check name of columns to see what data looks like.

names(genusdata)

## [1] "Genus" "Beginning" "End" "P"

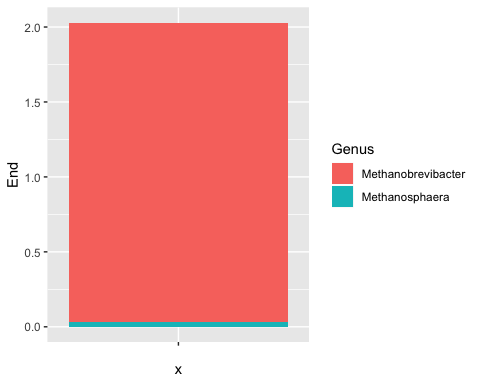
An initial plot is created showing the composition of two genera at the beginning of the feedlot period in High RFI/low efficiency steers. Save plot to a file.

ggplot(genusdata,aes(x="", y=Beginning, fill=Genus))+ geom\_bar(width = 1, stat = "identity")



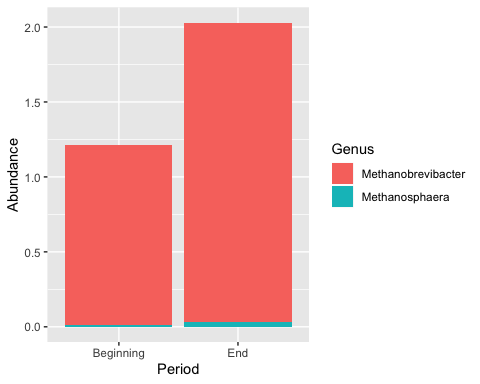
A second plot is created showing the composition of two genera at the end of the feedlot period in High RFI/low efficiency steers.Save the plot to a file.

ggplot(genusdata,aes(x="", y=End, fill=Genus))+ geom\_bar(width = 1, stat = "identity")



It is easier to see these plots side by side. To do this, create subsets to divide by period, “Beginning” and “End”.

Period <- c(rep("Beginning",2), rep("End",2))  
  
Genus <- rep(c("Methanobrevibacter",  
"Methanosphaera"),2)   
  
Abundance <- c(1.2, 0.01, 2.0, 0.03)  
   
#Combine datasets   
genuscomb <- data.frame(Period, Genus, Abundance)  
  
#Graph combined dataset  
ggplot(genuscomb, aes(fill=Genus, y=Abundance, x=Period)) +  
 geom\_bar(stat = "identity")



#Genus: Low RFI For my dataset, there was litte variation over the course of the feedlot amongst genra. Howver there were noticible changes amongst methanogen populations.

Load data, path is relative to project directory. Select a file and read the data into a data frame .

genusdata2 <- read\_csv(here("data", "raw\_data", "Genus\_Low\_MADA.csv"))

## Parsed with column specification:  
## cols(  
## `Bacterial Genus` = col\_character(),  
## Beginning = col\_double(),  
## End = col\_double(),  
## `P-VALUE` = col\_double()  
## )

Take a look at the data

head(genusdata2)

## # A tibble: 2 x 4  
## `Bacterial Genus` Beginning End `P-VALUE`  
## <chr> <dbl> <dbl> <dbl>  
## 1 Methanobrevibacter 1.3 1.4 0.83  
## 2 Methanosphaera 0.02 0.02 0.86

dplyr::glimpse(genusdata2)

## Observations: 2  
## Variables: 4  
## $ `Bacterial Genus` <chr> "Methanobrevibacter", "Methanosphaera"  
## $ Beginning <dbl> 1.30, 0.02  
## $ End <dbl> 1.40, 0.02  
## $ `P-VALUE` <dbl> 0.83, 0.86

Remove any observations with NA

genusdata2 <- drop\_na(genusdata2)  
dim(genusdata2)

## [1] 2 4

Rename the columes, so when creating a graph things can be plotted easier

names(genusdata2)[1:4]<-c("Genus", "Beginning", "End", "P")

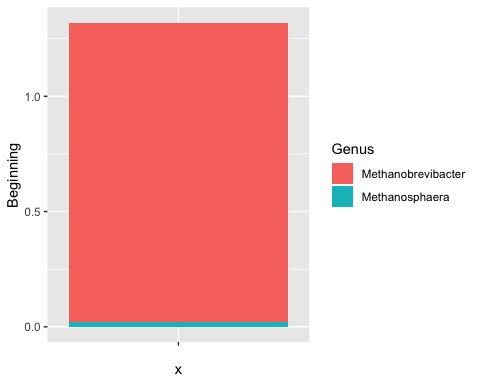
Check name of columns to see what data looks like.

names(genusdata2)

## [1] "Genus" "Beginning" "End" "P"

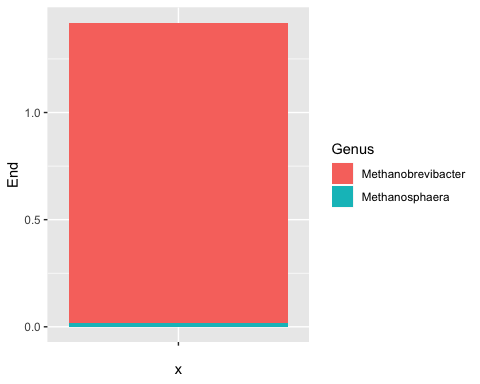
An initial plot is created showing the composition of two genera at the beginning of the feedlot period in Low RFI/High efficiency steers. Save plot to a file.

ggplot(genusdata2,aes(x="", y=Beginning, fill=Genus))+ geom\_bar(width = 1, stat = "identity")



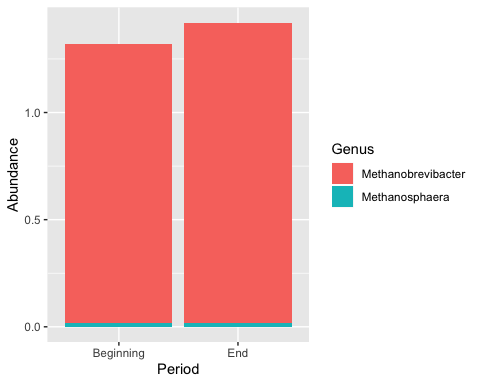
A second plot is created showing the composition of two genera at the end of the feedlot period in Low RFI/High efficiency steers.Save the plot to a file.

ggplot(genusdata2,aes(x="", y=End, fill=Genus))+ geom\_bar(width = 1, stat = "identity")



It is easier to see these plots side by side. To do this, create subsets to divide by period, “Beginning” and “End”.

Period <- c(rep("Beginning",2), rep("End",2))  
  
Genus <- rep(c("Methanobrevibacter",  
"Methanosphaera"),2)   
  
Abundance <- c(1.3, 0.02, 1.4, 0.02)  
   
#Combine datasets   
genuscomb2 <- data.frame(Period, Genus, Abundance)  
  
#Graph combined dataset  
ggplot(genuscomb2, aes(fill=Genus, y=Abundance, x=Period)) +  
 geom\_bar(stat = "identity")



## Univariate analysis

*Use a combination of text/tables/figures to explore and describe your data. You should produce plots or tables or other summary quantities for most of your variables. You definitely need to do it for the important variables, i.e. if you have main exposure or outcome variables, those need to be explored. Depending on the total number of variables in your dataset, explore all or some of the others.*

## Bivariate analysis

*Create plots or tables and compute simple statistics (e.g. t-tests, simple regression model with 1 predictor, etc.) to look for associations between your outcome(s) and each individual predictor variable*

## Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You can then load the results produced by this code*

# Discussion

## Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

# References