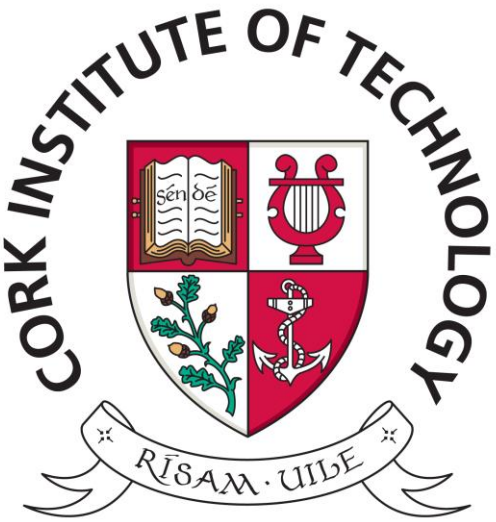
Cork Institute of Technology
Department of Mathematics



An Investigation into the Impact of a Stress Treatment on the Abundance of the Euryarchaeota Microbe in Cattle

---

Student: Shane O'Sullivan - R00181553

Supervisors: Professor Paul Walsh; Aengus Daly

Higher Diploma - Data Science & Analytics 2018/2019
23/05/2019

Declaration

This report was written entirely by the author, except where stated otherwise, and has not been submitted for another degree, at Cork Institute of Technology or elsewhere. The source of any material not created by the author has been clearly referenced. The work described in this report was conducted by the author, except where stated otherwise.

Signed:_____ Date:_____

Abstract

With the recent large-scale environmental protests and growing public awareness of methane's destructive contribution to global warming, agriculture and in particular, cattle farming has been highlighted as a vital area through which emissions reductions must be achieved.

The purpose of this project is to investigate whether the introduction of a stress treatment affected the abundance of the Euryarchaeota microbe within the digestive systems of Bos Taurus (cattle). The project sought to investigate this microbe's abundance due to its ability produce methane. To examine the effect of this treatment one control group and one testing group were established. The effect of varying diet on the abundance of the Euryarchaeota microbe was examined, given the success of numerous studies on this topic. Samples were taken from the cattle at three separate points: prior to the initial injection, after the third of three injections and a final sampling point. To determine the effect of the treatment, T-tests and ANCOVA modelling techniques were utilised to determine if a change in this microbe's abundance took place. The implications of using relative abundancies in analysis was discussed. The analysis did not find any statistically significant impact on the relative abundance of the Euryarchaeota microbe due to the treatment. However, the diet of the cattle was shown to have an impact, with a primarily concentrated feed diet resulting in a reduced Euryarchaeota microbe relative abundance compared to the primarily grazing diet. The residual feed intake variable was also recognised as being significant.

Acknowledgements

I would like to thank my supervisor Professor Paul Walsh for recommending such a challenging and interesting project, Dr. Catherine Palmer and Aengus Daly for their advice and feedback and finally, I would also like to thank Joana Lima and Miguel Somarriba for providing the data analysed in this report.

Contents

Introduction

The following report details the preparation and analysis of the data, presentation of results and conclusions drawn. The data used in this report came from Professor Paul Walsh, Joana Lima and Miguel Somarriba of the Department of Computer Science at Cork Institute of Technology.

The data was compiled using observations obtained from 37 cattle of species Limousin from the SRUC Easter Howgate farm located outside Edinburgh, Scotland. The aim of the study was to determine whether the volume of methane gas produced by cattle could be reduced by injecting cattle with a stress treatment. Similar to humans, dairy cattle release hormones such as cortisol and adrenaline when they are stressed. This in turn has been shown to reduce animal health in the form of a weaker immune system, an increased incidence of mastitis, higher Somatic Cell Counts, a reduced appetite. As a result, milk yields from the affected cattle will decline in both quantity and quality. (Agriculture & Horticulture Development Board, 2019) The objective of the stress treatment is to alter hormone production due to stress, thereby influencing the abundance of the microbes present.

Methane is produced by a process called enteric fermentation in the digestive system of domestic ruminant animals such as cattle, goats and sheep. According to Professor Peter Raynor of the Australian Academy of Science, Methane's Global Warming Potential (compared to CO2) is approximately 30. (Raynor, n.d.) This means that the impact on global warming of one tonne of Methane gas is equivalent to the impact of 30 tonnes of C02 being emitted.

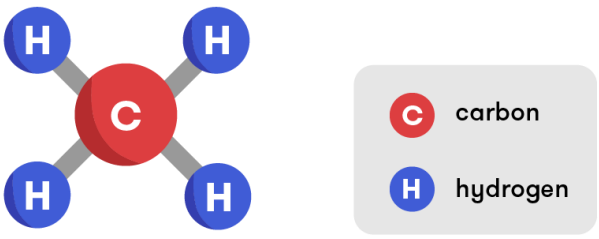


Figure 1: Methane Molecule (Source: Raynor, Peter)

Of the microbes present within the digestive system of cattle it is the Euryarchaeota microbe that is responsible for the production of methane in cattle. Therefore, any changes in the Euryarchaeota microbial abundance will be of crucial benefit in determining the effect, if any, the treatments have on methane production.

To examine whether there is a change in the relative abundance of the Euryarchaeota microbe, the sample size of 38 cattle was split into 2 groups for comparison. Members of Group1 (Treatment/DEX) were injected with the stress treatment 3 times over the course of 3 days. Members of the control group did not receive the stress treatment. Samples of the microbial abundance present within each bovine were obtained and the relative abundancies were calculated. The first sampling round took place prior to the first injection of the treatment. The second sampling round occurred immediately after the final injection of the treatment. The third sampling round took place a period after the final injection of the treatment. The sampling stages took place at the same time for all cattle, regardless of group. A description of each of the microbes recorded within the samples collected is featured further on in this report.

In the following sections of this report, the processes of preparing the data and conducting exploratory analysis of the data will be explained. Amongst the statistical analysis methods used were t-tests, tests for normality and variances and ANCOVA modelling techniques.

Within this report the implications of using relative abundancies as opposed to nominal values will be discussed. This includes the issues of analysing the relationships, or lack of, between variables and the impact on assessing changes in the abundance of microbes over the three sampling rounds. From the results of the analysis, a set of conclusions were reached which are detailed towards the end of this report.

Climate Change

Agriculture is seen as a key source of CO2 equivalent emissions as part of the broader human environmental impact. According to the Intergovernmental Panel on Climate Change," ***Human activities are estimated to have caused approximately 1.0°C of global warming above pre-industrial levels, with a likely range of 0.8°C to 1.2°C."*** (Masson-Delmotte, 2018)

In terms of the Irish context, the Environmental Protection Agency has determined that one-third of all greenhouse gas emissions in 2017 came from agriculture, almost equal to the sum of the next two largest sectors (Energy Industries (19.3%), Transport(19.8%)) combined. This can be seen in the pie chart below:

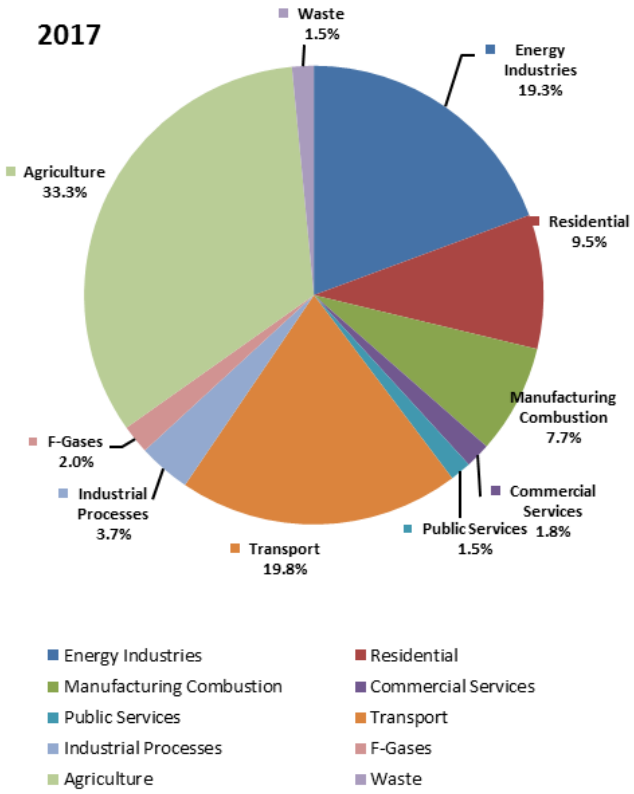


Figure 2:Ireland's Greenhouse gas emissions by sector for 2017 (Source: EPA 2019)

In relation to the impact of dairy and beef farming, Teagasc's Annual Sustainability Report 2019 through its survey of 90,000 farms across Ireland, has calculated that the average dairy farm produced .733Kg of CO2 equivalent for every 1 litre of milk produced. (Buckley, et al., 2017)

By 2020, Ireland is expected, under European Union targets, to decrease its carbon emissions by 20% from their 2005 levels or face potentially enormous fines. It is clear that insufficient progress has been made as in its 2018 annual review, the Climate Change Advisory Council, noted that Ireland's carbon emissions will have been reduced by less than 1% from 2005 levels. (Hilliard, 2018) According to the Minister for Communications, Climate Action & Environment, Ireland will be forced to spend €150million next year to purchase 16million climate credits. As the price of carbon credits and fines charged by the European Union on delinquent members will increase over time, it is in Ireland's best interest financially to curb carbon emissions.

Due to the volume of C02 equivalent emissions by agriculture and in particular cattle farming, even a small change in the CO2 equivalent emissions per cow can significantly reduce Ireland's impact on climate change. This will also reduce and prevent further, larger government expenditure on either carbon alleviation measures and carbon credits. (Lee, 2019)

Methodology

To conduct an examination and analysis of the data provided, several statistical software packages were used. In terms of data cleaning and preparation, Microsoft Excel and Jupyter Notebook were utilised. For the exploratory data analysis section of this project, both Jupyter Notebook and RStudio were used. With regards to further analysis and visualizations, RStudio was the primary software package utilised.

Literature Review

With regards to the observing of methane production by cattle, there have been numerous studies testing hypothetical methods to reduce the volume of methane produced.

Residual Feed Intake

A 2007 study sought to quantify the relationship between residual feed intake and the daily methane emissions of Angus steers. The research concluded that animals who were recorded having lower residual feed intakes had lower daily rates of production of methane. However, it was acknowledged that the residual feed intake only explained a small proportion of the observed variation in the daily production of methane. (Hegarty, 2007)

Diet

In relation to the diet sources for cattle, a paper was published in the Journal of Dairy Science examining the effect of adding various seeds to the diets of dairy cattle. The research concluded that Crushed sunflower, flax, and canola seeds all had an effect in reducing the volume of methane produced by cattle. (Beauchemin, 2009)

McCaughey et al. examined the effect of pasture type on the volume of methane produced by lactating beef cows. (W. P. McCaughey, 1999) It was observed that the methane production by lactating cows was greater for those grazing on alfalfa-grass pastures than for those grazing only on grass pastures. Additionally, it was noted that the dry matter intake was greater for cows grazing on alfalfa-grass pastures.

Model Building

In terms of model building to predict methane output, Ellis et al. utilised 172 datasets to develop models to estimate methane production for beef and dairy cows. The response variable used for each model was volume of methane produced per day (expressed as Mejajoules/day). Numerous variables were considered in the building of these models including: Dry Matter Intake, Metabolizable Energy Intake and forage proportion of diet amongst others. The research resulted in the creation of 2 out of the 9 models tested adequately predicting the amount of Methane produced. (J.L.Ellis, et al., 2007) In terms of the RMSPE (Root Mean Squared Percentage Error) calculated, the best performing model

for the beef datasets contained the variables: metabolizable energy intake, acid detergent fibre and lignin. With regards to dairy cattle, forage as a percentage of diet was significant.

Compositional Data

The dataset provided for this project was in relative abundancies with the sum of the relative abundancies for each observation equalling one. This presents a number of implications regarding analysing the data. Compositional Data that has been transformed into a proportions basis with the aim of having a constant sum (e.g. 1) risks elements appearing more/ less abundant than they should be. This can lead to incorrect inferences. Gloor & Reid noted in a 2016 paper that much of the distortion can be removed using a ratio transformation involving dividing proportional/ nominal values by the geometric means of the sample. (Gloor & Reid, 2016) The three graphs below are from Gloor & Reid's 2016 paper. The proportions graph would enable false inferences of the data as the filled in circle appears to be more abundant than it should be. The reverse is true for the empty circles.

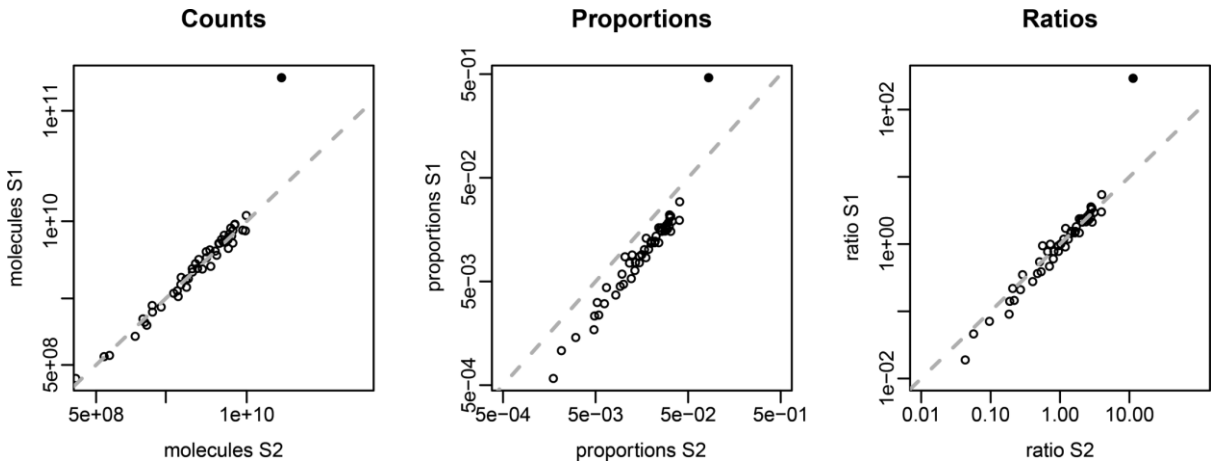


Figure 3: Compositional Data Visualisation (Gloor & Reid, 2016)

Lord's Paradox

This arises due to differing conclusions drawn from the use of ANCOVA and t-tests when comparing two groups in a study involving data obtained prior to and post an event occurring. Kim (2018) noted that when the initial selection process is randomised and pre and post values are positively correlated, the statistical power to conclude that the alternative hypothesis is significant is greater in an ANCOVA model than it would be in a t-test. (Kim, 2018) In Van Breukelen (2006) it was concluded that the ANCOVA model as

opposed to the ANOVA model should be used for randomised studies that are based on relative values. (Van Breukelen, 2006)

Table 1: Description of Substances Recorded During Sampling

Name	**Description**
Firmicutes	Firmicutes is the main bacterial phylum, comprising over 250 genera, including Lactobacillus, Streptococcus, Mycoplasma, and Clostridium which are able to produce several short chain fatty acids (SCFAs) like butyrate. (Tissue, 2017)
Bacteroides	A genus of gram-negative anaerobic bacteria that belong to the family Bacteroidaceae, that have rounded ends, produce no endospores and no pigment, and that occur usually in the normal intestinal flora. (Merriam-Webster, n.d.)
Euryarchaeota	Archaebacteria; Ability to perform cellular respiration using carbon as their electron acceptor and capable of producing methane. Often found in the stomachs of ruminants including cows. (Biology Dictionary, n.d.)
Proteobacteria	The Proteobacteria are a major group (phylum) of bacteria. They include a wide variety of pathogens, such as Escherichia, Salmonella, Vibrio, Helicobacter, and many other notable genera. Others are free-living and include many of the bacteria responsible for nitrogen fixation. All proteobacteria are Gram-negative, with an outer membrane mainly composed of lipopolysaccharides. (Lumen Learning, n.d.)
Actinobacteria	Actinobacteria constitute one of the largest phyla among Bacteria and represent gram-positive bacteria with a high G+C content in their DNA. This bacterial group includes microorganisms exhibiting a wide spectrum of morphologies, from coccoid to fragmenting hyphal forms, as well as possessing highly variable physiological and metabolic properties. Furthermore, Actinobacteria members have adopted different lifestyles, and can be pathogens (e.g., Corynebacterium, Mycobacterium, Nocardia, Tropheryma, and Propionibacterium), soil inhabitants (Streptomyces), plant

	commensals (Leifsonia), or gastrointestinal commensals (Bifidobacterium). (Anon., 2007)
Verrucomicrobia	The phylum Verrucomicrobia is a divergent phylum within domain Bacteria including members of the microbial communities of soil and fresh and marine waters; recently extremely acidophilic members from hot springs have been found to oxidize methane. (Lee, et al., 2009)
TM7	TM7 is a recently described subgroup of Gram-positive uncultivable bacteria originally found in natural environmental habitats. An association of the TM7 bacterial division with the inflammatory pathogenesis of periodontitis has been previously shown. (Kuehbacher, et al., n.d.)
Spirochaetales	Spirochete, (order Spirochaetales), also spelled spirochaete, any of a group of spiral-shaped bacteria, some of which are serious pathogens for humans, causing diseases such as syphilis, yaws, Lyme disease, and relapsing fever. Spirochetes are gram-negative, motile, spiral bacteria, from 3 to 500 m (1 m = 0.001 mm) long. (Encyclopædia Britannica, Inc, 2019)
Cyanobacteria	Cyanobacteria are aquatic and photosynthetic, that is, they live in the water, and can manufacture their own food. Because they are bacteria, they are quite small and usually unicellular, though they often grow in colonies large enough to see. They have the distinction of being the oldest known fossils, more than 3.5 billion years old. (Berkeley, n.d.)
Fibrobacteres	The phylum Fibrobacteres currently comprises one formal genus, Fibrobacter, and two cultured species, Fibrobacter succinogenes and Fibrobacter intestinalis, that are recognised as major bacterial degraders of lignocellulosic material in the herbivore gut. (Ransom-Jones, et al., 2012)

SR1	SR1 includes cosmopolitan bacteria that are found in marine and terrestrial high-temperature environments, fresh-water lakes, and subsurface aquifers. SR1 bacteria also associate with animals and exist in termite and mammalian digestive tracts as well as in the human oral cavity. SR1 is in low abundance in healthy oral microbiota (∼0.1% on average) (Campbell, et al., 2013)
Lentisphaerae	the phylum Lentisphaerae comprised the orders Lentisphaerales (Cho et al., 2004) and Victivallales (Cho et al., 2004) and five subphyla that contain no cultured representatives (Hedlund et al., 2011). Lentisphaera accommodates Gram-negative, nonmotile, non-pigmented cocci that produce extracellular polymeric substances in oligotrophic seawater medium. (Choi, et al., 2013)
Elusimicrobia	Organisms of the candidate phylum termite group 1 (Elusimicrobia) are regularly encountered in termite hindguts but are present also in many other habitats. (DP, et al., 2009)
Planctomycetes	Planctomycetes are a unique divergent phylum of the domain Bacteria. Members display a number of unusual properties, such as cell compartmentalization among many species examined electron microscopically, the presence of unusual or unique lipids, such as sterols and ladderane lipids in some species, and unique physiology in some species, such as the anammox planctomycetes performing ammonium oxidation anaerobically. (Kurtböke, 2017)
Tenericutes	A phylum of gram-negative bacteria consisting of cells bounded by a plasma membrane. Its organisms differ from other bacteria in that they are devoid of cell walls. This phylum was formerly the class Mollicutes. Mollicutes is now the sole class in the phylum Tenericutes. (U.S. National Library of Medicine, 2011)
Synergistetes	Members of the phylum Synergistetes have been demonstrated in several environmental ecosystems and mammalian microflorae by culture-independent methods. In the past few years, the clinical relevance of some uncultivated phylotypes has been demonstrated in

	endodontic infections, and uncultured Synergistetes have been demonstrated in human mouth, gut and skin microbiota. (H1, et al., 2010)
Chloroflexi	The *Chloroflexi* is one of the most common and diverse bacterial phyla in sponges and contains many sponge-specific lineages. (Schmitt, et al., 2011)
Fusobacteria	Fusobacteria are non–spore-forming, nonmotile, pleomorphic, gram-negative, obligate anaerobic bacilli that can cause a wide spectrum of human disease ranging from mild pharyngitis to sepsis, and these organisms are most notorious for causing septic thrombophlebitis of the internal jugular vein, commonly referred to as Lemierre syndrome.1 (Rellosa & Vodzak, 2018)
Acidobacteria	Acidobacteria is a very abundant and ubiquitous bacterial phylum in natural ecosystems. The dominance of Acidobacteria in acidic environments and chemically polluted sites (e.g. where heavy metal5,6,7, petroleum compounds8, linear alkylbenzene sulfonate9 and p-nitrophenol10 are major contaminants) is related to the ability of these bacteria to produce large amounts of EPS. (Kielak, et al., 2017)
Chlamydiae	Chlamydiae are obligate intracellular bacteria. They lack several metabolic and biosynthetic pathways and depend on the host cell for intermediates, including ATP. Chlamydiae exist as two stages: (1) infectious particles called elementary bodies and (2) intracytoplasmic, reproductive forms called reticulate bodies. The chlamydiae consist of three species, C trachomatis, C psittaci, and C pneumoniae. (Becker, 1996)

Exploratory Data Analysis

Data Cleaning

After an initial examination of the data, it was clear that some observations would need to be removed. In one case this was due to a cow falling ill and therefore taken out of the study and another due to an incorrect sampling procedure taking place on one occasion. All observations pertaining to the former were removed and the observation obtained from the incorrect sampling procedure was also removed. Given the large number of microbes observed with relatively small abundancies, the analysis will primarily focus on the following microbes:

- Firmicutes
- Bacteroides
- Euryarchaeota
- Proteobacteria
- Actinobacteria
- Verrucomicrobia

Summary Statistics & Visualizations

To examine the distributions in recorded microbial abundance a series of boxplots and other visualisations have been created to highlight aspects of the data. Summary tables containing the minimum, maximum, median ,mean and standard deviation for the Control and Treatment groups were created:

Table 2:Summary Statistics for Treatment Group

	Firmicutes	Bacteroidetes	Euryarchaeota	Proteobacteria	Actinobacteria	Verrucomicrobia
count	69.000000	69.000000	69.000000	69.000000	69.000000	69.000000
mean	0.312679	0.277785	0.202550	0.163429	0.022472	0.004693
std	0.073513	0.048274	0.059769	0.090142	0.016773	0.007248
min	0.210687	0.162351	0.095437	0.008091	0.003418	0.000000
25%	0.257988	0.248819	0.157606	0.104278	0.010326	0.000000
50%	0.293375	0.278678	0.200746	0.144562	0.016313	0.000910
75%	0.366618	0.307430	0.240530	0.201768	0.033766	0.006533
max	0.568774	0.406940	0.349385	0.417493	0.089589	0.030233

Table 3: Summary Statistics for Control Group

	Firmicutes	Bacteroidetes	Euryarchaeota	Proteobacteria	Actinobacteria	Verrucomicrobia
count	46.000000	46.000000	46.000000	46.000000	46.000000	46.000000
mean	0.308098	0.272150	0.211127	0.159287	0.025944	0.005148
std	0.046423	0.060482	0.064685	0.072819	0.018425	0.005793
min	0.218879	0.163368	0.028028	0.016745	0.001723	0.000000
25%	0.273576	0.238753	0.179947	0.115736	0.014725	0.000007
50%	0.309301	0.267030	0.208951	0.154618	0.023417	0.003805
75%	0.334356	0.291760	0.261442	0.188589	0.032388	0.008513
max	0.427148	0.492790	0.327739	0.357084	0.094520	0.023871

Comparing the two tables above, several interesting insights can be obtained including:

- For the Euryarchaeota microbe, the mean for the Treatment Group (0.2026) is less than the mean calculated for the Control Group (0.2111).
- The standard deviation of the Firmicutes relative abundance varies significantly between the two groups. (Treatment: 0.0735 V. 0.0464: Control) This may be due to the larger sample size for the treatment group of 23 cattle.

To examine the distributions of the variables across different sampling rounds the following boxplots were created. The Seaborn module in the Jupyter Notebook software program was utilised to create the following:

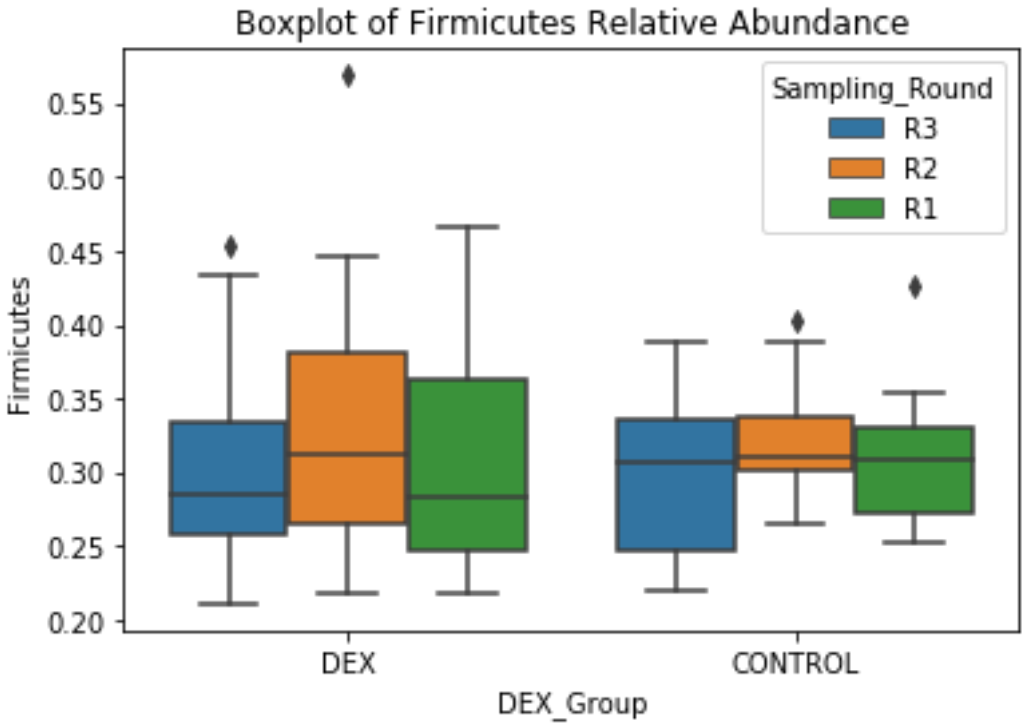


Figure 4: Boxplot of Firmucutes Microbe Relative Abundance

In terms of the Firmicutes microbe it appears that there was a slight difference in the median values between the Control and Treatment groups at the time of the first sampling round. However, the difference appears to disappear for the second sampling round. For the third sampling round, visually, it appears broadly the same for both the Control and Treatment groups.

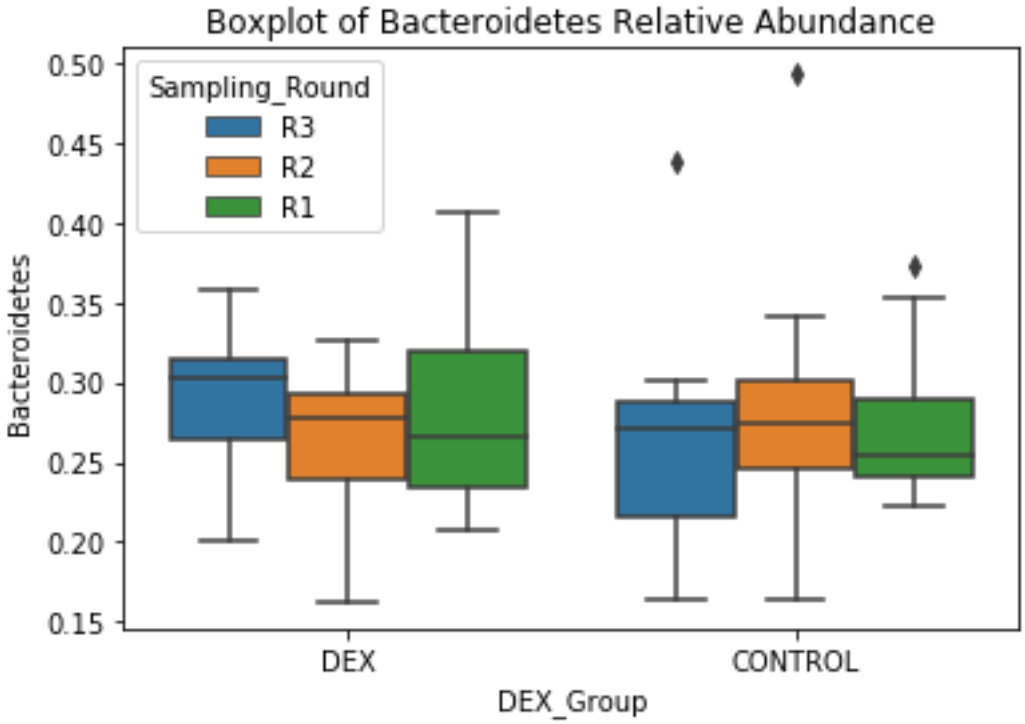


Figure 5: Boxplot of Bacteriodetes Microbe Relative Abundance

For the Bacteroides microbe it appears that their relative abundance increases due to the introduction of the treatment with a slight lead in the 25th percentile, median and 75th percentile values between the Treatment and Control groups recordings for sampling round 3.

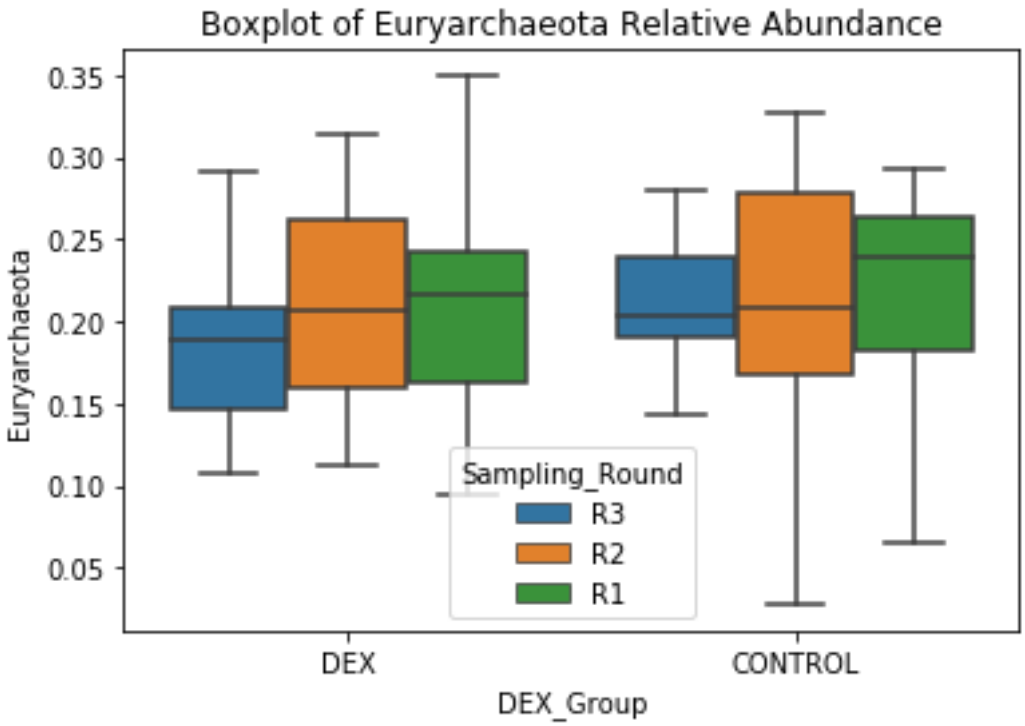


Figure 6: Boxplot of Euryarchaeota Microbe Relative Abundance

In the case of the Euryarchaeota microbe which is responsible for the production of methane, the boxplot highlights a small difference between the Treatment and Control groups for the R3 sampling round. It appears from the boxplot that the Treatment group has a lower 25th percentile, median and 75th percentile than its Control counterpart.

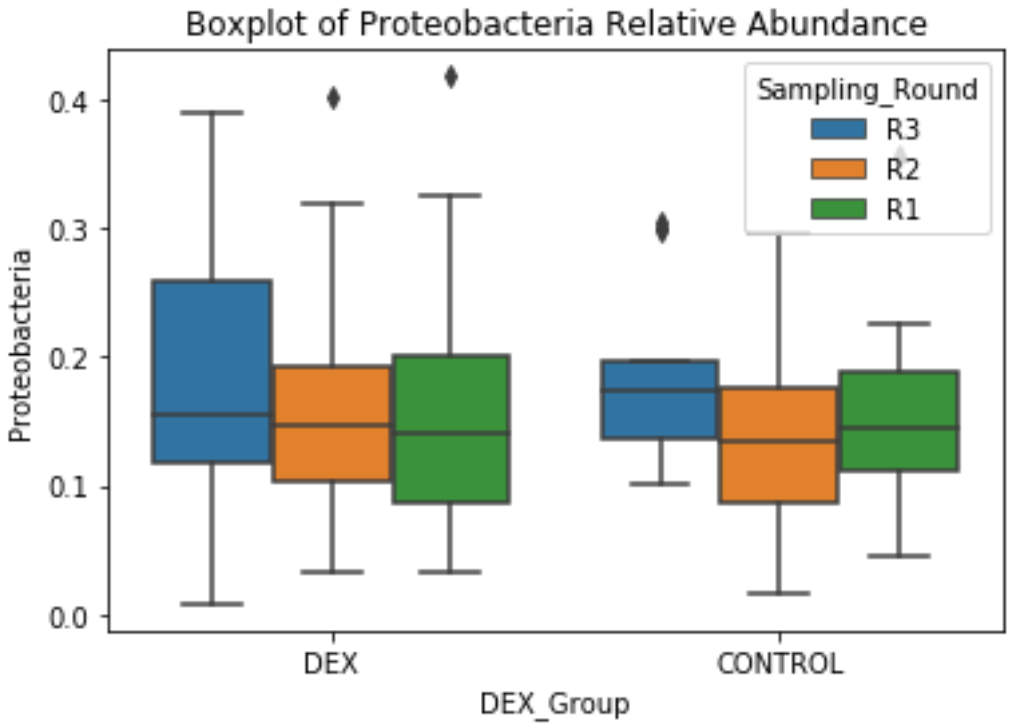


Figure 7: Boxplot of Proteobacteria Microbe Relative Abundance

Although the Treatment/ DEX median value (R3) for the Proteobacteria Microbe is lower than its Control counterpart, the interquartile range has doubled. This would indicate that although the values for many observations are lower than their counterparts, a small proportion of values appear to have risen significantly causing the interquartile range to grow.

Effect of Diet Variable

Each bovine either had a diet of primarily forage based (e.g. grazing based) or concentrated feed based. Given the research conducted by Beauchemin et al. examining the effect of feed-based diets versus grazing based diets, the relationship between the two diet options was examined.

Table 4: Cattle Frequency Per Treatment and Diet

	Control	**Treatment**	**Total**
PBconc	7	12	19
PBforg	8	11	19
Total	15	23	38

To examine any potential differences in the Euryarchaeota microbe relative abundance, the following boxplots were created showing the mean abundance at each sampling round.

T-Tests

To compare the changes in abundance of the various microbes over each sampling round, the data was subdivided into the Control group and the Treatment group. The Control group and Treatment group were sampled at the same time. To examine whether there has been a change in the relative abundances of the microbes due to the treatment, an independent samples t-test is conducted. For this test to be conducted, the following two assumptions must hold:

Normality

Both samples being examined must be normally distributed. This is tested using the Shapiro-Wilkes test. For each round of sampling for both the control and Treatment groups, the Shapiro Wilkes test was conducted to examine normality.

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Equation 1: Shapiro Wilk Test Formula (Statistics How To)

The following is an example of the output produced in R studio:

```
        Shapiro-Wilk normality test

data:  control$Firmicutes[control$Sampling_Round == "R1"]
W = 0.90894, p-value = 0.1119
```

Figure 8: Shapiro Wilk Normality Test - RStudio Output (Firmicutes)

In the above example, the normality of the relative abundance recorded for the first sampling round of the Firmicutes microbe was examined. The null hypothesis is that the data is not normally distributed. The alternative hypothesis is that the data is normally distributed. As the p value is greater than 0.05 (5% level of significance), we fail to reject the null hypothesis. The data is normally distributed.

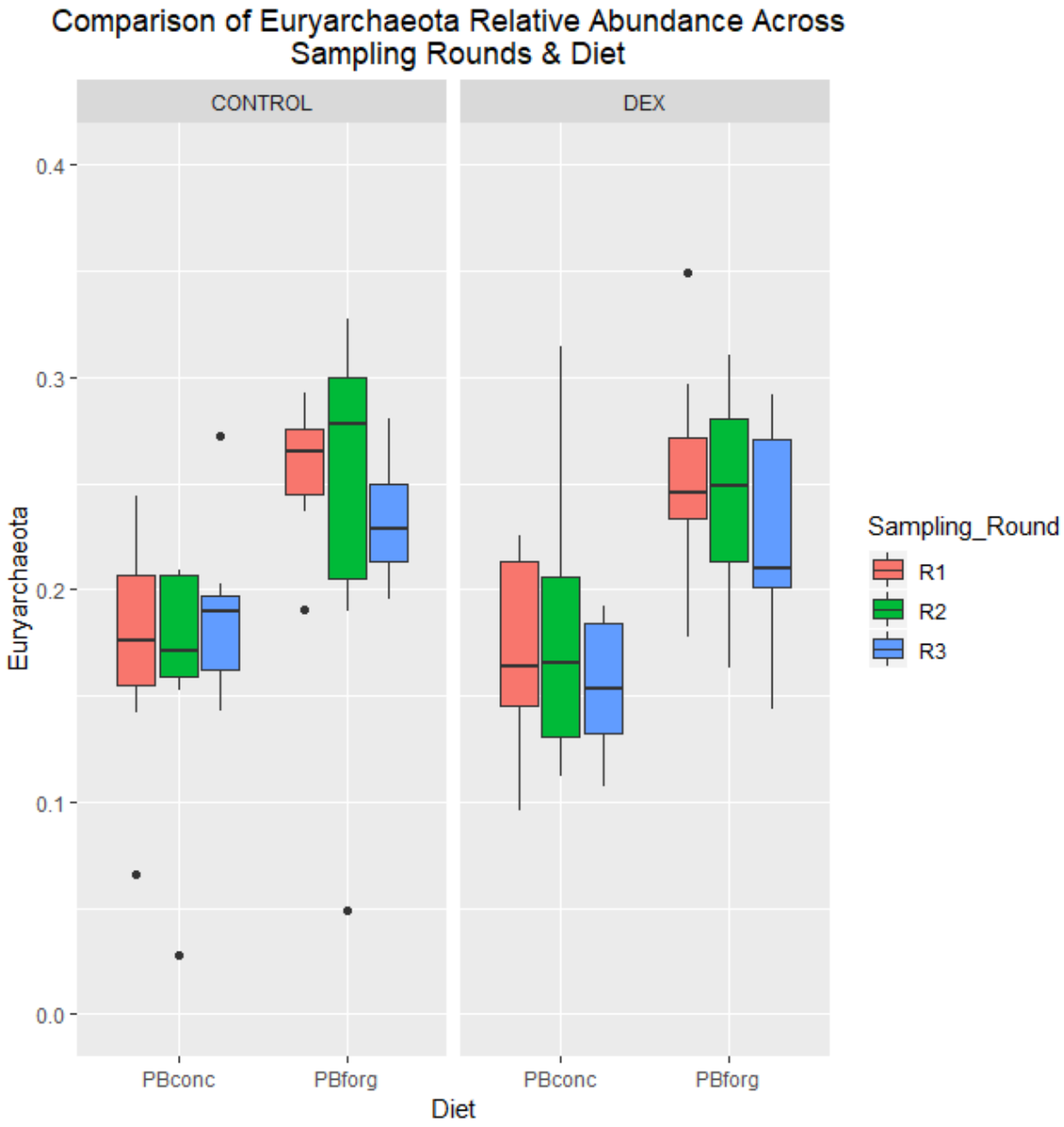


Figure 9:Boxplot Comparison of Euryarchaeota Relative Abundance per Diet per Sampling Round

From examining the boxplot representation of the data, it appears likely that the Euryarchaeota relative abundance varies with the bovine's diet. The median values for the foraging diet for each sampling round are higher than any median value for the concentrated feed group regardless of which Treatment Group it belonged to. This would seem to align with the conclusions reached by Beauchemin et al. that a concentrated feed diet has a significant impact on the lowering the relative abundance of the Euryarchaeota microbe, potentially leading to a reduction in methane emissions.

Variances

To examine whether the variances between the microbe populations are equal, f-tests were conducted on the data for each sampling round, comparing the Control group with the Treatment group.

To streamline this process, the original dataset was subsetted into three smaller groups based on the sampling round. The var.test() function was then used to compare the variances for each microbe at each round. An example of the output is as follows:

```
        F test to compare two variances

data:  firstround$Firmicutes by firstround$DEX_Group
F = 0.44091, num df = 15, denom df = 22, p-value = 0.1067
alternative hypothesis: true ratio of variances is not equal to
 1
95 percent confidence interval:
 0.1764754 1.2018915
sample estimates:
ratio of variances
          0.440907
```

Figure 10: F Test - RStudio Output (Firmicutes)

The null hypothesis for the f test is that the true ratio for the Firmicutes relative abundance recorded at the time of the first sampling round is equal to 1 for the Control and Treatment groups. The alternative hypothesis is that the true ratio of the variances is not equal to 1. The function returned a value of 0.4409 with a p value of .1067. At the 5% level of significance we fail to reject the null hypothesis that the true ratio of the variances is equal to 1.

This process is repeated for each of the main variables we are interested in each of the sampling rounds.

As the variances for some of the Control and Treatment groups are not equal at the 5% level of significance, the t-test and welch tests were used.

The formula for the unpaired two-samples t-test is as follows:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$s^2 = \frac{\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2 + \sum_{j=1}^{n_2}(x_j - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

Equation 2: Two Samples T-Test - StatsDirect Limited, 2016

The null hypothesis for the unpaired t-test is that the true difference in means between the Control and Treatment groups for the third round of testing, is equal to 0.

The alternative hypothesis for the unpaired t-test is that the true difference in means between the Control and Treatment groups for the third round of testing is not equal to zero. The level of significance for the t-test is 5%.

For the Firmicutes microbe, as the p-value of 0.6019 is greater than 0.05 we fail to reject the null hypothesis. We cannot conclude that the mean relative abundance of the firmicutes microbe is different between the Control and Treatment groups.

For the Bacteroides microbe, the variance test's p-values for the third sampling round comparing the variances of the Control and Treatment group was significant. Therefore, the Welch Test must be used instead. In R this is completed by setting the var.equal parameter to FALSE. As the p-value is .1859 and therefore greater than 0.05, the null hypothesis of the true difference in means is equal to zero, is not rejected. For the remainder of the microbes tested, the variances were equal, and the data follows a normal distribution.

For the Proteobacteria, Verrucomicrobia and Actinobacteria microbes analysed using the Two-Sample T-Test above, the p-values were not deemed to be significant at the 5% level. Therefore, the null hypothesis of the true difference in means being equal is not rejected.

Based on the data obtained from the t-test outputs collected in RStudio, we can conclude that the difference in means between the Control and Treatment groups is not equal to zero.

Comparison of Round One and Round Three Sampling Results (Treatment Group):

As the Firmicutes, Actinobacteria and Verrucomicrobia microbes are not normally distributed for one or more of their sampling rounds, they cannot be compared using the standard t-test. Instead the Wilcox Signed Rank Test is used to examine whether there has been a statistically significant change in the means of the microbes recorded between the sampling rounds.

The hypotheses tested for the Wilcoxon Rank Sum tests for each of the microbes are as follows:

Null Hypothesis: The true location shift is equal to zero.

Alternative Hypothesis: The true location is not equal to zero.

$$z = \frac{W - 0.5}{\sqrt{\frac{n(n+1)(2n+1)}{6}}}$$

$$W = |\sum[\mathrm{sgn}(x_2 - x_1) \cdot R]|$$

Equation 3:Wilcoxon Rank Sum (Educational Research Techniques)

The output for each of the three microbes was as follows:

Firmicutes

```
        Wilcoxon rank sum test

data:  firstround$Firmicutes[firstround$DEX_Group == "DEX"] and third
round$Firmicutes[thirdround$DEX_Group == "DEX"]
W = 260, p-value = 0.9307
alternative hypothesis: true location shift is not equal to 0
```

Figure 11: Wilcoxon Rank Sum Test - Firmicutes

As the p value is greater than 0.05, the null hypothesis is not rejected: The true location shift between sampling rounds one and three is equal to zero at the 5% level of significance.

Actinobacteria

```
        Wilcoxon rank sum test

data:  firstround$Actinobacteria[firstround$DEX_Group == "DEX"] and
thirdround$Actinobacteria[thirdround$DEX_Group == "DEX"]
W = 381, p-value = 0.00987
alternative hypothesis: true location shift is not equal to 0
```

Figure 12: Wilcoxon Rank Sum Test - Actinobacteria

As the p-value is significant at the 5% level the null hypothesis is rejected, the true location shift is not equal to zero. This means that there is a statistically significant difference between the mean recorded for the first sampling round and the third sampling round.

Verrucomicrobia

```
        Wilcoxon rank sum test with continuity correction

data:  firstround$Verrucomicrobia[firstround$DEX_Group == "DEX"] and
 thirdround$Verrucomicrobia[thirdround$DEX_Group == "DEX"]
W = 256, p-value = 0.8562
alternative hypothesis: true location shift is not equal to 0
```

Figure 13: Wilcoxon Rank Sum Test - Verrucomicrobia

As the p-value calculated is greater than 0.05, the null hypothesis is not rejected, the true location shift is equal to zero.

For the other three microbes, the observations are approximately normally distributed and so the Paired Samples T-Test will be used to analyse the microbes. To use the paired samples t-test formula, the assumption that the differences between the microbes are normally distributed must be examined. To do so, the Shapiro Wilks test is performed on the differences between the relative abundancies recorded for the first and third sampling rounds of the microbes. The null and alternative hypotheses for the tests are as follows:

Null Hypothesis: The data is not normally distributed

Alternative Hypothesis: The data is normally distributed.

Euryarchaeota

```
        Shapiro-Wilk normality test

data:  dexthird$Euryarchaeotadifference
W = 0.97594, p-value = 0.8268
```

Figure 14: Shapiro Wilk Test - Euryarchaeota

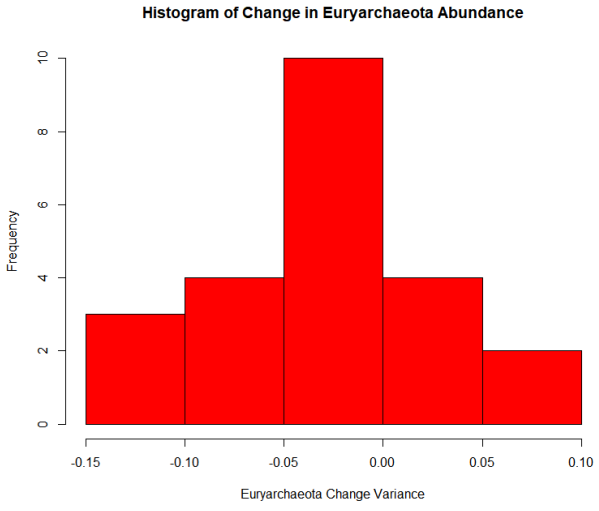


Figure 15: Histogram of Change in Euryarchaeota Relative Abundance

In the case of the Euryarchaeota microbe the differences in recordings for the first and third sampling rounds are normally distributed. We reject the null hypothesis given that the p value calculated of 0.8268 is greater than 0.05 (5% level of significance).

Proteobacteria

```
        Shapiro-Wilk normality test

data:  dexthird$Proteobacteriadifference
W = 0.96654, p-value = 0.6069
```

Figure 16: Shapiro Wilk Test - Proteobacteria

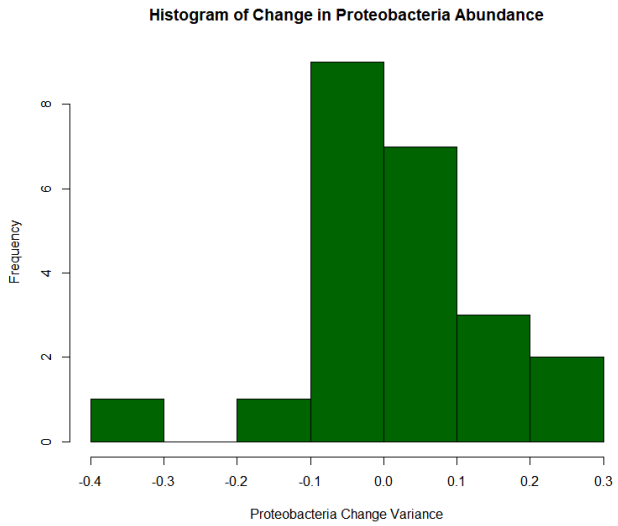


Figure 17: Histogram of Change in Proteobacteria Microbe Relative Abundance

As the p-value recorded is greater than 0.05, we can reject the null hypothesis at the 5% level of significance and can conclude that the values have a normal distribution.

Bacteriodetes

```
        Shapiro-Wilk normality test

data:  dexthird$Bacteriodetesdifference
W = 0.96894, p-value = 0.6638
```

Figure 18: Shapiro Wilk Test - Bacteriodetes

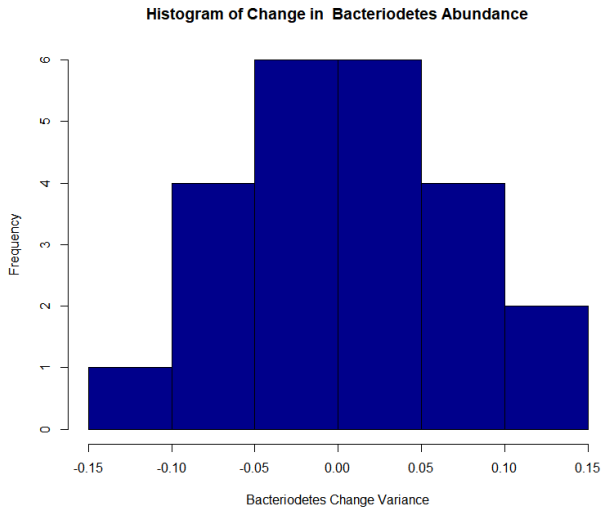


Figure 19: Histogram of Change in Bacteriodes Microbe Relative Abundance

As the p value is significant for the Bacteriodetes sampling rounds value differences, we can reject the null hypothesis.

As all three of the microbe relative abundance differences have been shown to be normally distributed, the paired samples t-test can now be conducted. The level of significance is set at 5% for the paired samples t-test. The null and alternative hypotheses for the three tests are as follows:

Null Hypothesis: The true difference in means between the two samples is equal to zero.

Alternative Hypothesis: The true difference in means between the two means is not equal to zero.

Bacteroidetes

As the p-value of 0.6009 is greater than 0.05 the null hypothesis is not rejected. The true difference in means is equal to zero.

Proteobacteria

As the p-value calculated of 0.384 > 0.05 the null hypothesis is not rejected. The true difference in means between the two samples is not statistically different from zero.

Euryarchaeota

At the 95% confidence level the null hypothesis cannot be rejected as the p-value calculated (0.0622) is greater than 0.05. However, it is statistically significant at the 90% confidence level. The following is the output when the t-test function's confidence level is set to 90%:

```
        Paired t-test

data:  dexfirst$Euryarchaeota and dexthird$Euryarchaeota
t = 1.9646, df = 22, p-value = 0.06222
alternative hypothesis: true difference in means is not equal to
0
90 percent confidence interval:
 0.002703564 0.040225895
sample estimates:
mean of the differences
             0.02146473
```

Figure 20: Paired T-Test - Euryarchaeota

Modelling Techniques

To examine the effect of the other microbes, stress treatment and dietary factors on the Euryarchaeota microbe, which is responsible for methane production, the ANCOVA (Analysis of Covariance) model comparison technique will be used. This is due to the use of continuous and factor variables in the model.

Due to the conclusions of the research carried out by McCaughey et al, the effect of diet on the relative abundance of the Euryarchaeota microbe was examined. The diet of an individual cattle remained constant throughout the study.

This section will focus on using the Euryarchaeota microbe as the dependent variable. Using the RStudio output from earlier in this report and some exploratory data analysis, several models were created and tested:

Model 1: Euryarchaeota ~ Sampling_Round + DEX_Group

The purpose of this model is incorporate both the time aspect of the sampling rounds and the binary variable,"DEX_Group" which indicates the group the cow belonged to. Dummy variables were created to consider the 3 sampling rounds and the 2 treatment groups. I expected both variables to be significant at the 5% level of significance.

However, none of the dummy variables created by the model turned out to be significant. Also, the Coefficient of Determination (R-squared) is quite low at 0.0146. In addition, the model's F statistic is not as useful as its calculated p value is not significant.

Model 2: Euryarchaeota ~ Sampling_Round + DEX_Group + Diet

The Diet variable was included in this model given both prior research which examined the effect of alternative diets on cattle and the exploratory data analysis which appeared to show large differences in the Euryarchaeota microbe's relative abundance. The RStudio output indicates that the DietPBforg dummy variable is significant in building a suitable model as the p-value calculated was 5.49e^-12. This indicates that a cow with a foraging based diet will be predicted to have an increased relative abundance of the Euryarchaeota microbe by 7.24%. When compared to model 1, model 2 is significant with an F statistic value of 15.56 and a p-value (4.048e-10) significant at the 5% level of significance.

Model 3: Euryarchaeota ~ Sampling_Round + DEX_Group + RFI_category

The RFI (Residual Feed Intake) variable was introduced to the model to examine whether a change in the feed intake of a cow would result in a change in its Euryarchaeota microbe relative abundance. The RFI variable is used as a dummy variable in the regression model as it is binary (High or Low). The low residual feed intake dummy variable is significant at the 5% level of significance. Its coefficient of -0.023819 means that if a subject is observed as having a low residual feed intake then it is predicted to have a reduction of 2.38% in its Euryarchaeota microbe relative abundance.

Model 4: Euryarchaeota ~ Sampling_Round + DEX_Group + Diet + RFI_category

Before testing the above model, I would expect some correlation to exist between the residual feed intake status and the dietary options for a cow.

Testing the model in RStudio, returns the result that both Diet and RFI Category have an impact on the predicted Euryarchaeota relative abundance. This is at the 5% level of significance. To compare the explanatory capabilities of the models, the Adjusted R Squared is analysed. This calculation is used as it considers that by using an additional variable our model's R Squared will at worst stay the same and is almost certain to increase. The Adjusted R Squared aids in enforcing Occam's principle that a simpler model (i.e. less parameters) is preferred over a more complex model (i.e. more parameters) that has a similar explanatory capability. Between models 2 & 3 there is an increase in the Adjusted R Squared of 0.0309 which indicates that having the RFI_Category variable included in the model will improve on its predictive performance.

To compare the 4 models built the ANCOVA test is utilised. Out of the 4 models, only two were calculated to be significant at the 5% level of significance. The baseline model is model1. The two models which are significant are Models 2 & 4. The only difference between these two models is the addition of the RFI_Category parameter. The Residual Sum of Squares (The statistic showing the amount of variation not explained by the model) for model 2 was 0.27666 and for model 4 it was 0.26126.

To compare these models more definitively, the ANOVA test was conducted on just those two models. From the output it can be concluded that at the 5% level of significance, Model

4 which includes the RFI_Category is significant. With a lower Residual Sum of Squares value, it would be of better use in predicting a change in the baseline Euryarchoeta rate.

Areas for Future Research

With regards to potential research opportunities, the influence of the weight of a cow versus the methane level produced could be examined. As the weight of a cow is carefully measured for beef production, it could be worthwhile to examine microbe abundance on a per kg basis.

From the data analysed in this report it is highly probable that the diet of a cow has an enormous impact on the methane produced. In this study, the two categories for feeding were a primarily grazing based diet versus a primarily concentrated feed diet. Further studies could examine alternative diets such as the introduction of seaweed on this breed of cattle.

Additionally, with regards to replicating this study, it may be useful to increase the sample size to examine the results on a larger hard whilst collecting more data with regards to height and age.

Conclusions

Based on the analysis undertaken above, at the 5% level of significance, there is not a statistically significant difference in the values for the relative abundance of most microbes. The Euryarchaeota microbe which is responsible for the production of methane, has not been shown to be significantly altered by the introduction of the treatment at the 5% level of significance. However, it has been shown that there is a statistically significant difference at the 10% level of significance. With a larger sample size and more frequent sampling it may be possible to definitively conclude whether the treatment can lead to a statistically significant decrease in the amount of methane produced by cattle.

Bibliography

Agriculture & Horticulture Development Board, 2019. *Stress - Health & Productivity.* [Online]
Available at: https://dairy.ahdb.org.uk/technical-information/animal-health-welfare/mastitis/symptoms-of-mastitis/stress-health-and-productivity/#.XOHo78hKi1s

Anon., 2007. Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum. *Microbiol & Molecular Biology Reviews,* 71(3), pp. 495-548.

Beauchemin, K. e. a., 2009. Crushed sunflower, flax, or canola seeds in lactating dairy cow diets: Effects on methane production, rumen fermentation, and milk production. *Journal of Dairy Science,* 92(5), pp. 2118-2127.

Becker, Y., 1996. Chapter 39 Chlamydia. In: S. Baron, ed. *Medical Microbiology.* Galveston: University of Texas Medical Branch at Galveston; .

Berkeley, n.d. *Introduction to the Cyanobacteria.* [Online]
Available at: https://ucmp.berkeley.edu/bacteria/cyanointro.html

Biology Dictionary, n.d. *Archaebacteria.* [Online]
Available at: https://biologydictionary.net/archaebacteria/

Buckley, C. et al., 2017. *National Farm Survey 2017 Sustainability Report,* Athenry, Co. Galway: Teagasc.

Campbell, J. et al., 2013. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Procededings of the National Academy of Sciences,* 110(14), pp. 5540-5545.

Choi, A., Yang, S.-J., Rhee, K.-H. & Cho, J.-C., 2013. Lentisphaera marina sp. nov., and emended description of the genus Lentisphaera. *International Journal of Systematic and Evolutionary Microbiology,* Issue 63, p. 1540–1544.

DP, H. et al., 2009. Genomic analysis of "Elusimicrobium minutum," the first cultivated representative of the phylum "Elusimicrobia" (formerly termite group 1).. *Applied and Envrionmental Microbiology,* 75(9), pp. 2841-2849.

educational research techniques, n.d. *Wilcoxon Signed Rank Test in R.* [Online]
Available at: https://educationalresearchtechniques.com/2016/03/30/wilcoxon-signed-rank-test-in-r/

Encyclopædia Britannica, Inc, 2019. *Spirochete.* [Online]
Available at: https://www.britannica.com/science/spirochete

Environmental Protection Agency, 2019. *What are Irelands greenhouse gas emissions ?.* [Online]
Available at: https://www.epa.ie/climate/communicatingclimatescience/whatisclimatechange/whatareirelandsgreenhousegasemissionslike/

GetClac.com, n.d. *Z, t, F & χ^2 Statistic & Formulas.* [Online]
Available at: https://getcalc.com/statistics-statistic-calculator.htm

GHG Online, 2006. *Methane Sources - Ruminants.* [Online]
Available at: http://www.ghgonline.org/methaneruminants.htm

Gloor, G. & Reid, G., 2016. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. *Canadian Journal of Microbiology,* 62(8), pp. 692-703.

H1, M. et al., 2010. Phylogeny, diversity and host specialization in the phylum Synergistetes with emphasis on strains and clones of human origin.. *Research in Microbiology,* 2(161), pp. 91-100.

Hegarty, R. &. G. J. &. M. H. R. &. M. B., 2007. Cattle selected for lower residual feed intake have reduced daily methane production. *Jurnal of Animal Science,* 85(6), pp. 1479-1486.

Hilliard, M., 2018. *Ireland 'can't reach' target to cut carbon emissions by 2020.* [Online]
Available at: https://www.irishtimes.com/news/environment/ireland-can-t-reach-target-to-cut-carbon-emissions-by-2020-1.3576192

J.L.Ellis, et al., 2007. Prediction of Methane Production from Dairy and Beef Cattle. *Journal of Dairy Science,* 90(7), pp. 3456-3466.

Kielak, A. et al., 2017. Characterization of novel Acidobacteria exopolysaccharides with potential industrial and ecological applications. *Scientific Reports,* Volume 7.

Kim, S., 2018. Explaining Lord's Paradox in Introductory Statistical Theory Courses. *International Journal of Statistics and Probability,* 7(4).

Kuehbacher, T., Rehman, A. & L, n.d.

Kurtböke, I., 2017. *Microbial Resources: From Functional Existence in Nature to Application.* 1 ed. s.l.:Academic Press.

Lee, G., 2019. *€150m carbon credit cost for State over failure to meet 2020 emissions target.* [Online]
Available at: https://www.rte.ie/news/environment/2019/0515/1049589-carbon-credits/

Lee, K.-C.et al., 2009. Phylum Verrucomicrobia representatives share a compartmentalized cell plan with members of bacterial phylum Planctomycetes. *BioMed Central,* 9(5).

Lumen Learning, n.d. *Proteobacteria.* [Online]
Available at: https://courses.lumenlearning.com/boundless-microbiology/chapter/proteobacteria/

Masson-Delmotte, V. P. Z. H.-O. P. D. R. J. S. P. S. A. P. W. M.-O. C. P. R. P. S. C. J. M. Y. C. X. Z. M. G. E. L. M. M. T. a. T. W. (., 2018. *IPCC, 2018: Summary for Policymakers. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global,* Geneva, Switzerland: Intergovernmental Panel on Climate Change.

Merriam-Webster, n.d. *bacteroides.* [Online]
Available at: https://www.merriam-webster.com/medical/bacteroides

Ransom-Jones, E., Jones, D., McCarthy, A. & McDonald, J., 2012. The Fibrobacteres: An Important Phylum of Cellulose-Degrading Bacteria. *Microbial Ecology,* 63(2), pp. 267-281.

Raynor, P. P., n.d. *Excuse me! The problem with methane.* [Online]
Available at: https://www.science.org.au/curious/earth-environment/methane

Rellosa, N. & Vodzak, J., 2018. *Principles and Practice of Pediatric Infectious Diseases (Fifth Edition).* 1 ed. s.l.:Elsevier.

Schmitt, S. et al., 2011. hloroflexi bacteria are more diverse, abundant, and similar in high than in low microbial abundance sponges. *Federation of European Microbiological Societies,* 78(3), pp. 497-510.

StatsDirect Limited, 2016. *Unpaired (Two Sample) t Test.* [Online]
Available at: https://www.statsdirect.co.uk/help/parametric_methods/utt.htm

Tissue, G., 2017. *Science Direct.* [Online]
Available at: https://www.sciencedirect.com/topics/medicine-and-dentistry/firmicutes

U.S. National Library of Medicine, 2011. *Tenericutes.* [Online]
Available at: https://www.ncbi.nlm.nih.gov/mesh?term=Tenericutes

Van Breukelen, G., 2006. ANCOVA versus change from baseline had more power in Randomized Studies, more bias in non-Randomized Studies. *Journal of Clinical Epidemiology,* 2(7).

W. P. McCaughey, K. W. D. C., 1999. Impact of pasture type on methane production by lactating beef cows. *Canadian Journal of Animal Science,* 79(1), pp. 221-226.

Appendices

- Table of p-values and f-values for the microbes (Sampling Round Comparison)
- Table of p-values and w-values for the microbes (Sampling Round and Diet Comparison)
- R Code

	Firmicutes		Bacteroidetes		Euryarchaeota		Proteobacteria		Actinobacteria		Verrucomicrobia	
	P-Value	F-Test	P-Value	F-Test	P-Value	F-Test	P-Value	F-Test	P-Value	F-Test	P-Value	F-Test
Sampling Round 1	1.0670	0.4409	0.2741	0.5758	0.9668	1.0060	0.3149	0.6021	0.7030	0.8196	0.8483	0.8988
Sampling Round 2	0.0060	0.2250	0.0329	2.7498	0.1785	1.8819	0.6311	0.7742	0.1013	2.1662	0.0405	0.3382
Sampling Round 3	0.3215	0.5969	0.0314	2.7753	0.3006	0.5836	0.1989	0.5122	0.3420	0.6098	0.8275	0.8823

Sampling Round		Firmicutes		Bacteroidetes		Euryarchaeota		Proteobacteria		Actinobacteria		Verrucomicrobia	
		P-Value	W	P-Value	W	P-Value	W	P-Value	W	P-Value	W	P-Value	W
Control	1	0.1119	0.9084	0.0225	0.8645	0.1493	0.9167	0.1323	0.9135	0.0666	0.8949	0.0007	0.7554
	2	0.1665	0.9159	0.0301	0.8666	0.2468	0.9271	0.8875	0.9721	0.0314	0.8679	0.0339	0.8701
	3	0.1653	0.9157	0.0642	0.8876	0.8476	0.9693	0.0272	0.8636	0.1246	0.9076	0.0024	0.7854
DEX	1	0.0963	0.9247	0.0592	0.9176	0.9705	0.9849	0.0614	0.9192	0.0037	0.8572	0.0000	0.7102
	2	0.0765	0.9228	0.1869	0.9408	0.2940	0.9501	0.0688	0.9206	0.0147	0.8883	0.0000	0.6746
	3	0.0125	0.8848	0.7324	0.9718	0.2630	0.9478	0.3710	0.9551	0.0048	0.8633	0.0000	0.7037