PUBH 7445 Final Project: Investigating associations in microbiome composition and Antimicrobial Resistance (AMR) genes of cow metagenomic data

Jesse Elder and Minu Bhunia

12/18/2020

Abstract:

Antimicrobial resistance is a complex phenomenon that can lead to dire consequences in human and animal health. In an effort to elucidate relationships between the microbiome and antimicrobial resistance this project compared four metagenomic studies of cow rumens. These relationships were studied using Bayesian networks. Additionally, relationships between bacterial abundance and study design were investigated. It was found that Prevotella was the dominant bacterial genus in all studies other than Hess. In contrast, Tetracyclines were the dominant class of AMR genes in all studies other than Shi. In both analyses, these trends are recapitulated with both PCA and k-means clustering. Regarding the relationships between the microbiome and AMR, it was found that the presence of one type of AMR often coincided with other types. Moreover, the presence of Metronidazole resistance as well as the genus Ruminococcus correlated strongly with increased Tetracycline resistance. From these analyses, it was concluded that bacterial abundance and antimicrobial resistance often cluster together based on study design.

Introduction:

Antimicrobial resistance is a complex phenomenon that can lead to dire consequences in human and animal health [1]. One way to study this is to investigate the rumen of cattle and other livestock. Rumen comes from the first stomach of these types of animals and can be used to investigate the microbiome in depth. Luckily, with the use of Next Generation sequencing, analysis of rumen samples has become increasingly affordable and fast [2].

There are many publicly available metagenomic datasets from rumen-associated samples. This project compared the four studies Hess et. al 2011; Wallace et. al 2015; Shi et. al 2014; and Stewart et. al 2018 in order to elucidate relationships between the microbiome and antimicrobial resistance [3,4,5,2]. The Stewart study was based in Scotland using Scottish beef cattle. Hess used switchgrass as feed which is unusual. Wallace used Aberdeen-Angus or Limousin cross who had a concentrate-based or forage-concentrate based diet. Lastly, Shi used Sheep instead of cows. It is interesting to note that since the studies conducted by Hess and Shi are somewhat different from the others it makes sense for these studies to stand out from the rest in our investigation.

Using these datasets, the laboratory of Dr. Noelle Noyes has compiled bacteria and gene count data for common ruminants. We used these compiled files (microbiome_Genus_Normalized and AMR_Gene_Normalized) for our investigations.

Data and Methods:

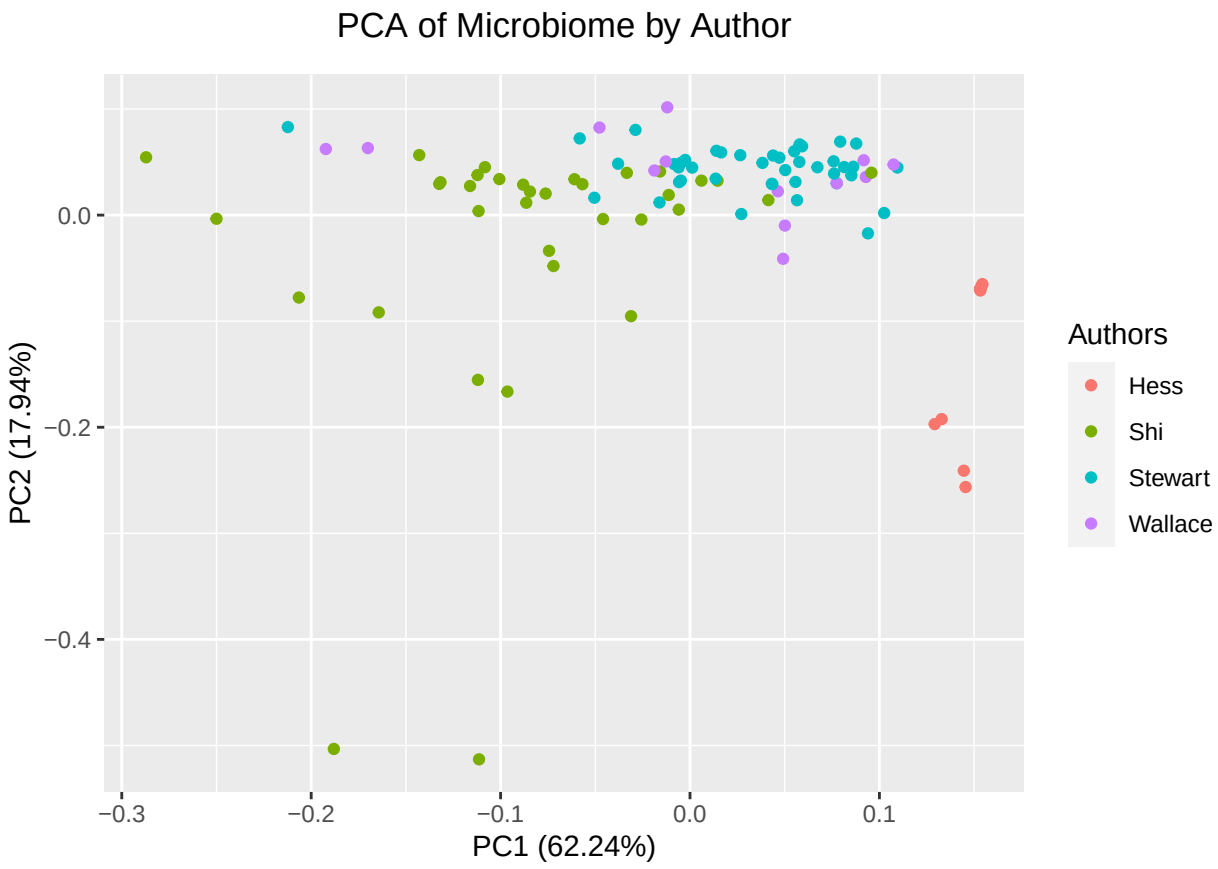
Data courtesy of Dr. Noelle Noyes. The investigated datasets included rumen microbiome/bacteria (microbiome_Genus_Normalized) and antimicrobial resistance gene composition (AMR_Gene_Normalized) data. Each of these files are formatted with all sample IDs from each study (Hess et. al 2011; Wallace et. al 2015; Shi et. al 2014; and Stewart et. al 2018) and the counts of each aligned read corresponding to either a bacterial genome or an antimicrobial resistance gene. The AMR_Gene_Normalized file had a few more complexities with the inclusion of Group, Mechanism, and Class information of each identified antimicrobial gene.

The analysis of these data was performed as follows. We planned a broad analysis to find any associations between bacteria and AMR in the microbiome, clustering between study samples, and constructing some type of classification with that information. We performed principal component analysis (PCA), multidimensional scaling (MDS), and t-distributed stochastic neighbor embedding (tSNE) for dimensionality reduction; K-means clustering for classification, and Bayesian network analysis to determine AMR-microbiome relationships.

Results:

Microbiome PCA

```
MicPCA<-prcomp(genus_t,scale=FALSE)
autoplot(MicPCA,data=Micauthors,colour="Authors") + ggtitle("PCA of Microbiome by Author") +
  theme(plot.title = element_text(hjust = 0.5,vjust=3))
```



AMR data loading and cleaning

PCA on AMR

```
PCAclass=prcomp(class_t)
autoplot(PCAclass, data=AMRauthors,colour="Authors")+ ggtitle("PCA of AMR Class by Author")+
  theme(plot.title = element_text(hjust = 0.5,vjust=3))
```


```
PCAgroup=prcomp(group_t)
autoplot(PCAgroup, data=AMRauthors, colour="Authors")+ ggtitle("PCA of AMR Group by Author")+
  theme(plot.title = element_text(hjust = 0.5,vjust=3))
```


```
PCAmech=prcomp(mech_t)
autoplot(PCAmech, data=AMRauthors, colour="Authors")+ ggtitle("PCA of AMR Mechanism by Author")
  theme(plot.title = element_text(hjust = 0.5,vjust=3))
```

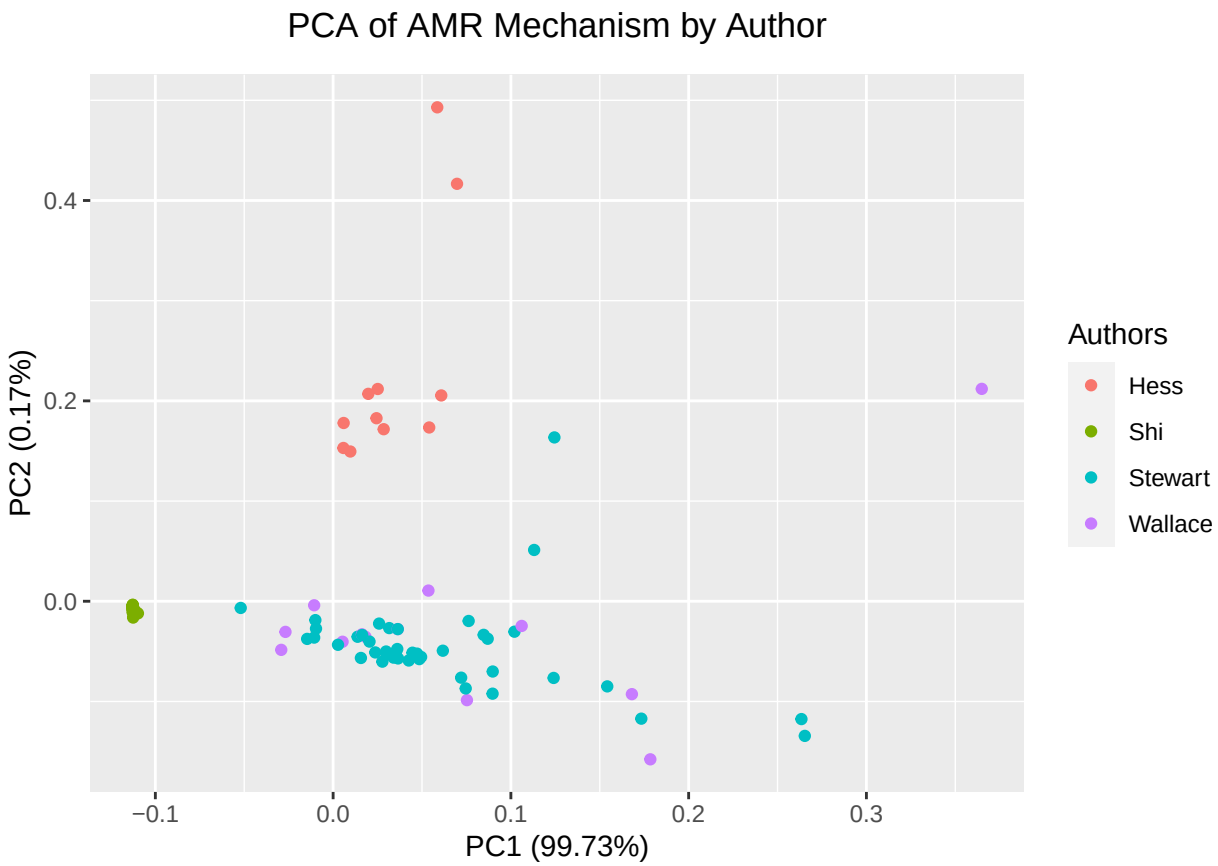


Figure 1: The PCA plots converts the correlations or lack of among the samples into a 2D graph. The samples that are highly correlated will cluster together. From our PCA plots without K means, there does not seem to be any clear clustering. Although Hess and Shi seem to be away from the other points.

Kmeans clustering on AMR Class data only

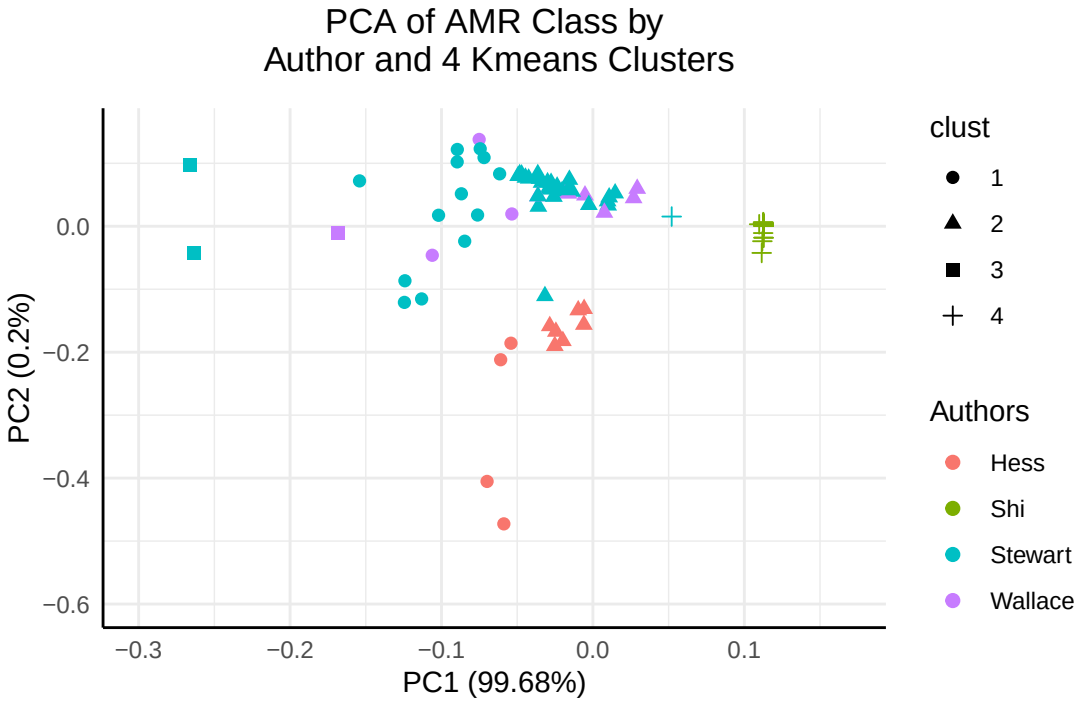
```
##########Kmeans

kenny<-kmeans(class_t,4)     #Kmeans cluster into 4 clusters
#aggregate(class_t,by=list(kenny$cluster),FUN=mean)     #Check out the means across the cluste

#fviz_cluster(kenny,data=class_t,geom="point",xlim=c(-2,2),ylim=c(-2,2),scale=FALSE) + theme(p

AMRauthorClust<-data.frame(class_t,Authors=AMRauthors,clust=as.factor(kenny$cluster))     #Com

#Final Figure
#Shape is the cluster it belongs to
#Colour is the Study it's from
AMRClassplot<-autoplot(PCAclass,data=AMRauthorClust,colour="Authors",shape="clust",size=2) +
  ggtitle("PCA of AMR Class by\n Author and 4 Kmeans Clusters") +
  theme_minimal() +
```

```
  theme(plot.title = element_text(hjust = 0.5,vjust=3)) +
  scale_y_continuous(limits=c(-0.6,0.15)) + scale_x_continuous(limits=c(-0.3,0.17)) +
  theme(plot.margin=unit(c(1,1,1.5,1.2),"cm"), axis.line = element_line(color="black"))
```

```
## Warning: `select_()` is deprecated as of dplyr 0.7.0.
## Please use `select()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

```
AMRClassplot
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```




```
kenny1<-kmeans(class_t,3)                          #Now try it with 3 clusters

AMRauthorClust<-cbind(AMRauthorClust,clust3=as.factor(kenny1$cluster))      #Merge old datafra

autoplot(PCAclass,data=AMRauthorClust,colour="Authors",shape="clust3")+ ggtitle("PCA of AMR Cla
```

Figure 2: A plot of the first two principal components of AMR class data. K-means clusters (4 clusters) are also labeled along with coloring the "correct" grouping by study. Principal components showed some notable clustering – Hess and Shi were distinct – but k-means clustering had limited predictive power – only Shi was correctly labeled.

Kmeans clustering on Microbiome

```
kennyMic<-kmeans(genus_t,4)
Micauthorframe<-cbind(genus_t,Authors=Micauthors,clust=as.factor(kennyMic$cluster))
Micplot<-autoplot(MicPCA,data=Micauthorframe,colour="Authors",shape="clust",size=2) +
  ggtitle("PCA of Microbiome by\n Author and Kmeans Cluster") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5,vjust=3)) +
  scale_y_continuous(limits=c(-0.6,0.15)) + scale_x_continuous(limits=c(-0.3,0.17)) +
  theme(plot.margin=unit(c(1,1,1.5,1.2),"cm"), axis.line = element_line(color="black"))
Micplot
```

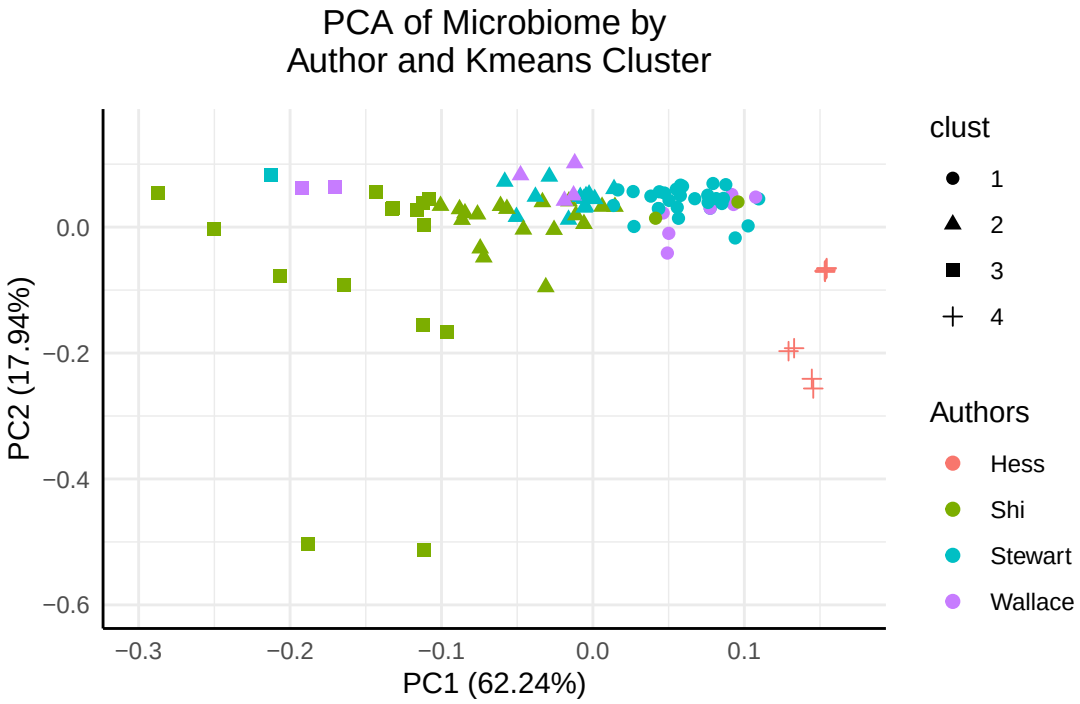
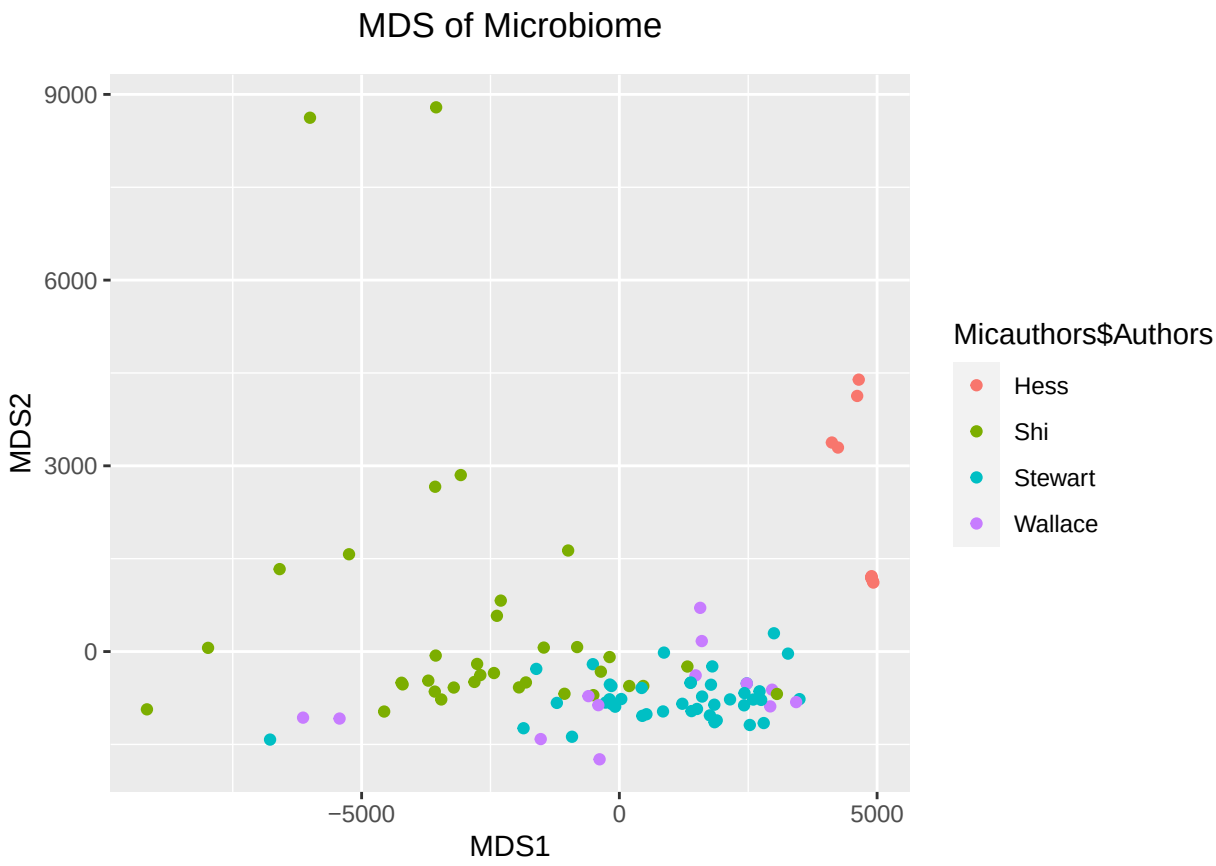


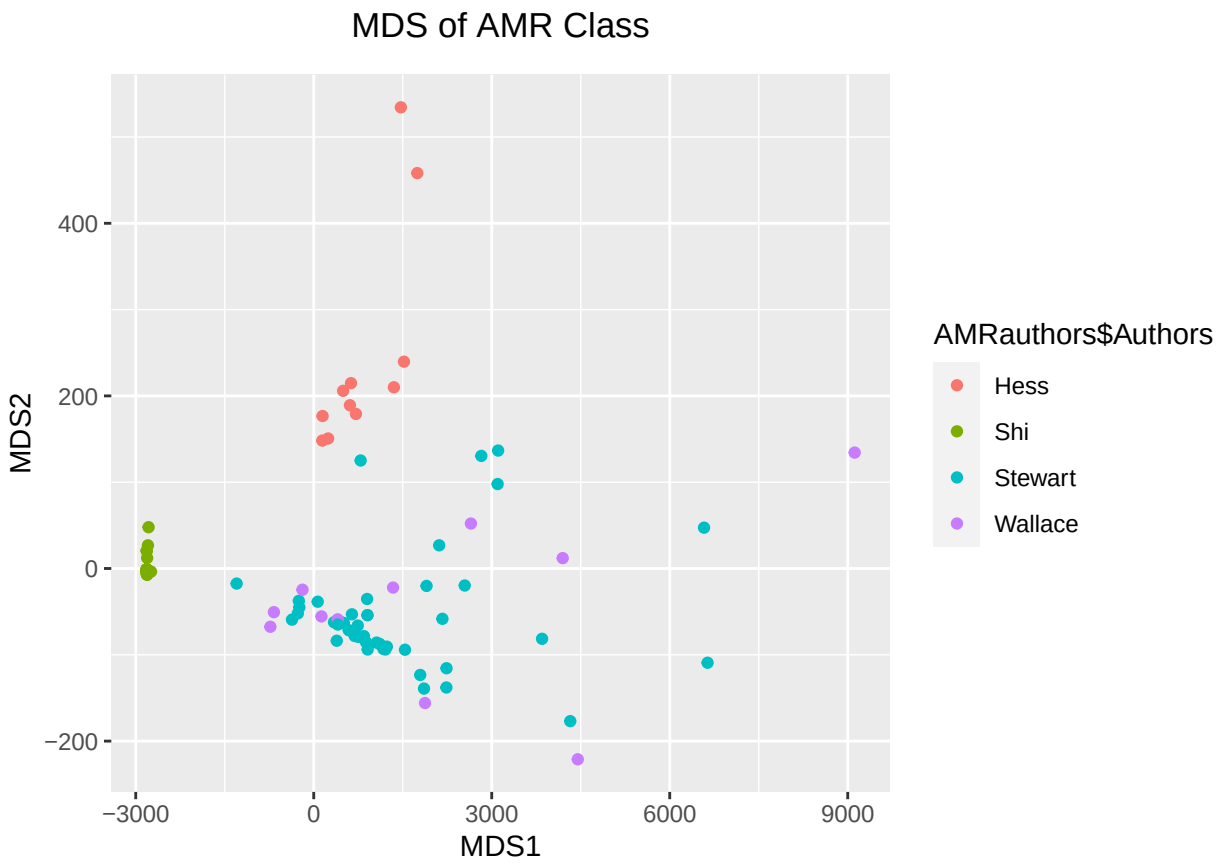
Figure 3: A plot of the first two principal components of bacterial genera count data. K-means clusters (4 clusters) are also labeled along with coloring the "correct" grouping by author. Again, principal components showed some notable clustering – in this case only Shi was distinct – but k-means clustering had limited predictive power – only Hess was correctly labeled.

MDS

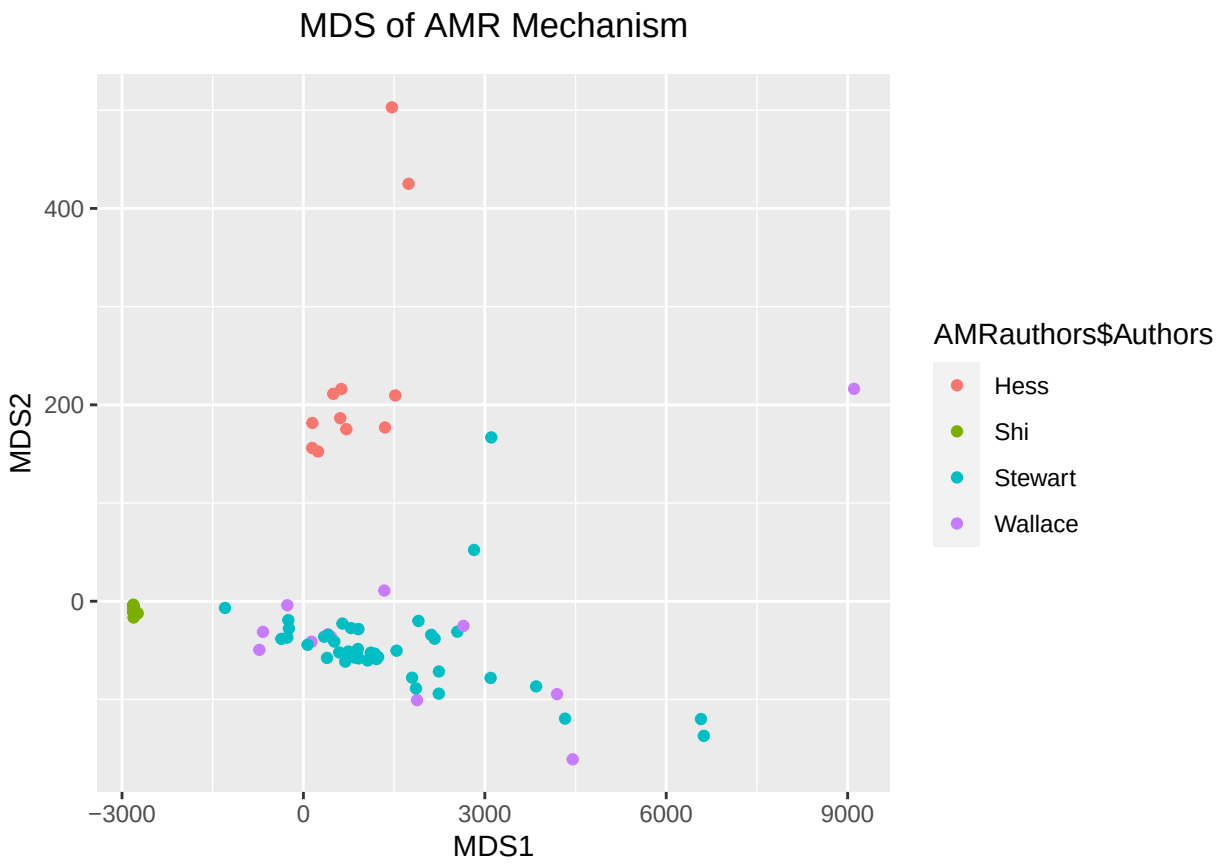
```
#microbiome
micro_dist<-dist(genus_t)
micro<-cmdscale(micro_dist)
x<-micro[,1]
y<-micro[,2]
p<-data.frame(x,y,Micauthors)
ggplot(p, aes(x=x, y=y, color=Micauthors$Authors)) + geom_point()+ggtitle("MDS of Microbiome")+
  theme(plot.title = element_text(hjust = 0.5,vjust=3)) +xlab("MDS1") +ylab("MDS2")
```




```
#AMR
#distance matrix, MDS, plot
#class
c_dist<-dist(class_t)
c<-cmdscale(c_dist)
x<-c[,1]
y<-c[,2]
p<-data.frame(x,y,AMRauthors)
ggplot(p, aes(x=x, y=y, color=AMRauthors$Authors)) + geom_point()+ggtitle("MDS of AMR Class")+
  theme(plot.title = element_text(hjust = 0.5,vjust=3))+xlab("MDS1") +ylab("MDS2")
```




```
#mechanism
m_dist<-dist(mech_t)
m<-cmdscale(m_dist)
x<-m[,1]
y<-m[,2]
p<-data.frame(x,y,AMRauthors)
ggplot(p, aes(x=x, y=y, color=AMRauthors$Authors)) + geom_point()+ggtitle("MDS of AMR Mechanism
  theme(plot.title = element_text(hjust = 0.5,vjust=3))+xlab("MDS1") +ylab("MDS2")
```




```
#group
g_dist<-dist(group_t)
g<-cmdscale(g_dist)
x<-g[,1]
y<-g[,2]
p<-data.frame(x,y,AMRauthors)
ggplot(p, aes(x=x, y=y, color=AMRauthors$Authors)) + geom_point()+ggtitle("MDS of AMR Group")+
  theme(plot.title = element_text(hjust = 0.5,vjust=3))+xlab("MDS1") +ylab("MDS2")
```

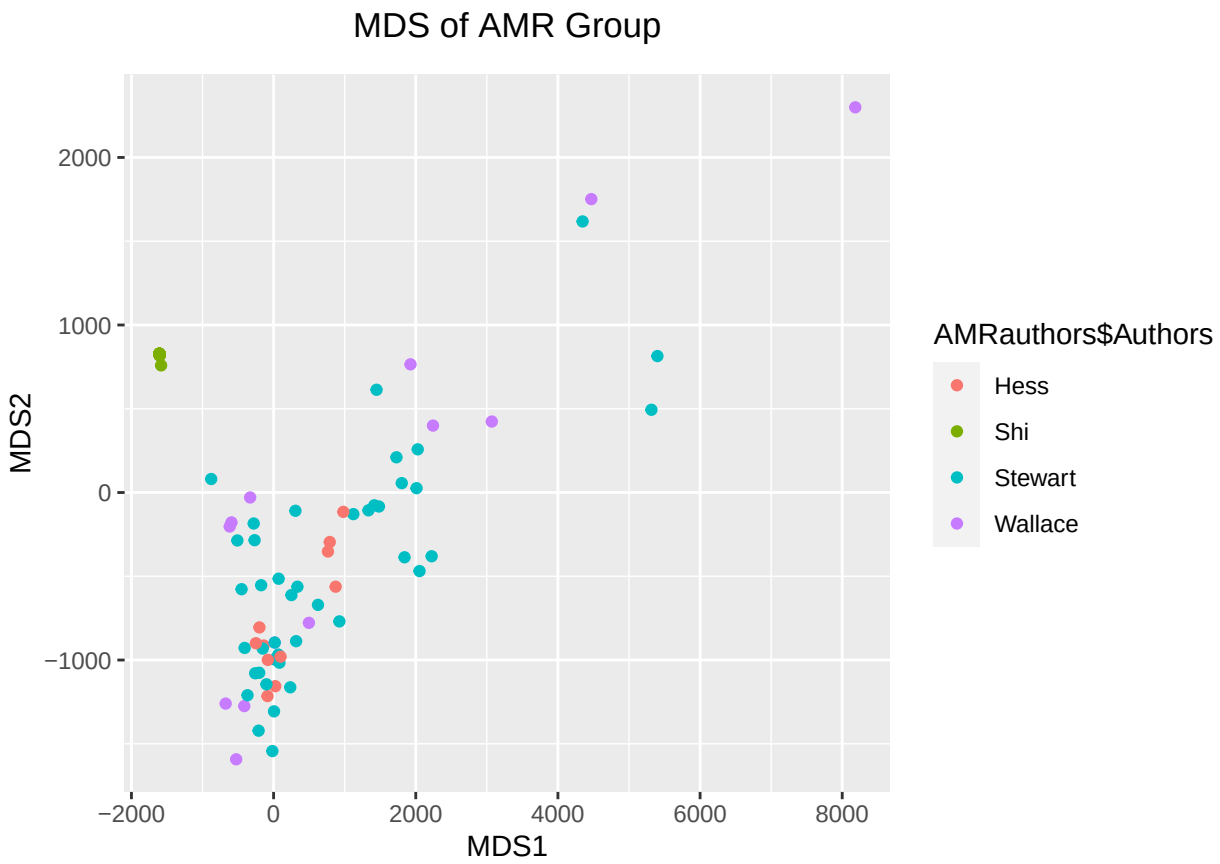


Figure 4: Plot of multidimensional scaling (MDS) using the microbiome data shows Hess clusters alone, but still is spread out.

Figure 5: Plot of multidimensional scaling (MDS) using the AMR class data shows Hess clusters alone, but still is spread out. Shi also clusters alone.

tSne on Microbiome Data

```
#tsne w/distance matrix (Also acceptable)
plot(SneezeDist,col=cool[AMRauthors$Authors],pch=11,main="tSNE of AMR Class Sample Distance",x
legend("topright",legend=c("hess","shi","stewart","wallace"),col = cool,pch=1)
```

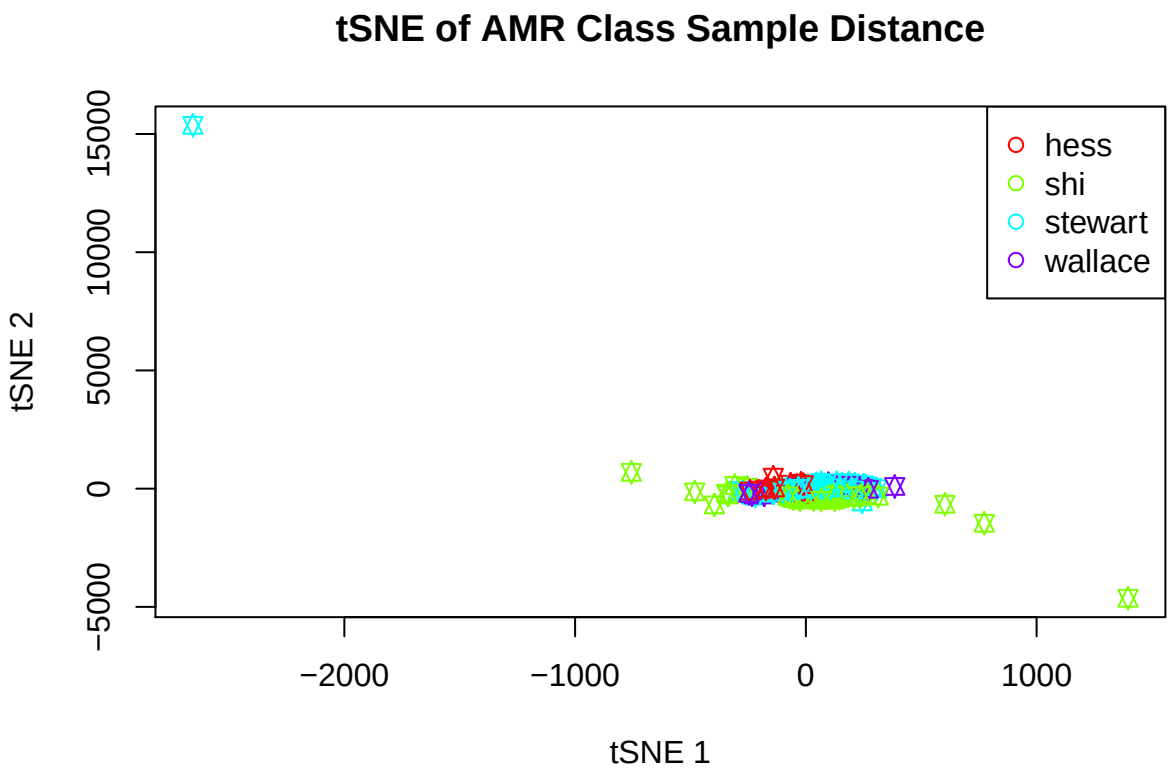
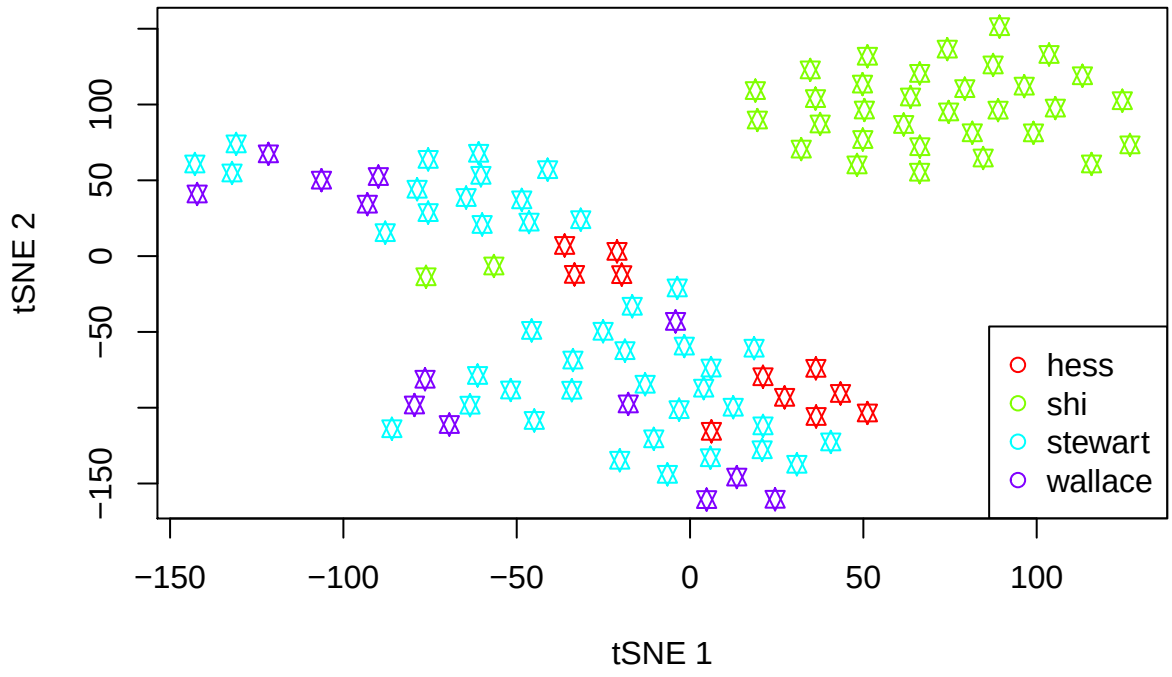


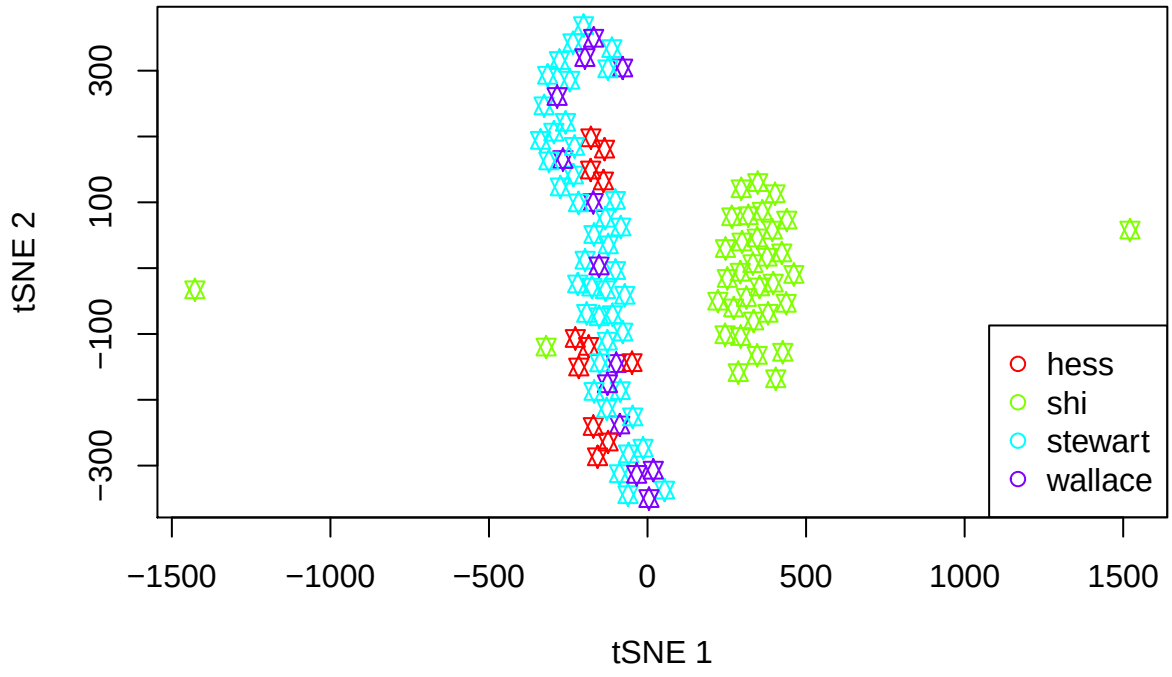
Figure 6: Plot of t-distributed stochastic neighbor embedding (tSNE) using microbiome data shows that Shi clusters alone.

tSne on AMR Data

```
plot(Sneeze,col=cool[AMRauthors$Authors],pch=11,xlab="tSNE 1",ylab="tSNE 2")
legend("bottomright",legend=c("hess","shi","stewart","wallace"),col = cool,pch=1)
```




```
plot(Sneeze,col=cool[AMRauthors$Authors],pch=11,xlab="tSNE 1",ylab="tSNE 2")
legend("bottomright",legend=c("hess","shi","stewart","wallace"),col = cool,pch=1)
```




```
plot(Sneeze,col=cool[AMRauthors$Authors],pch=11,xlab="tSNE 1",ylab="tSNE 2")
legend("bottomright",legend=c("hess","shi","stewart","wallace"),col = cool,pch=1)
```

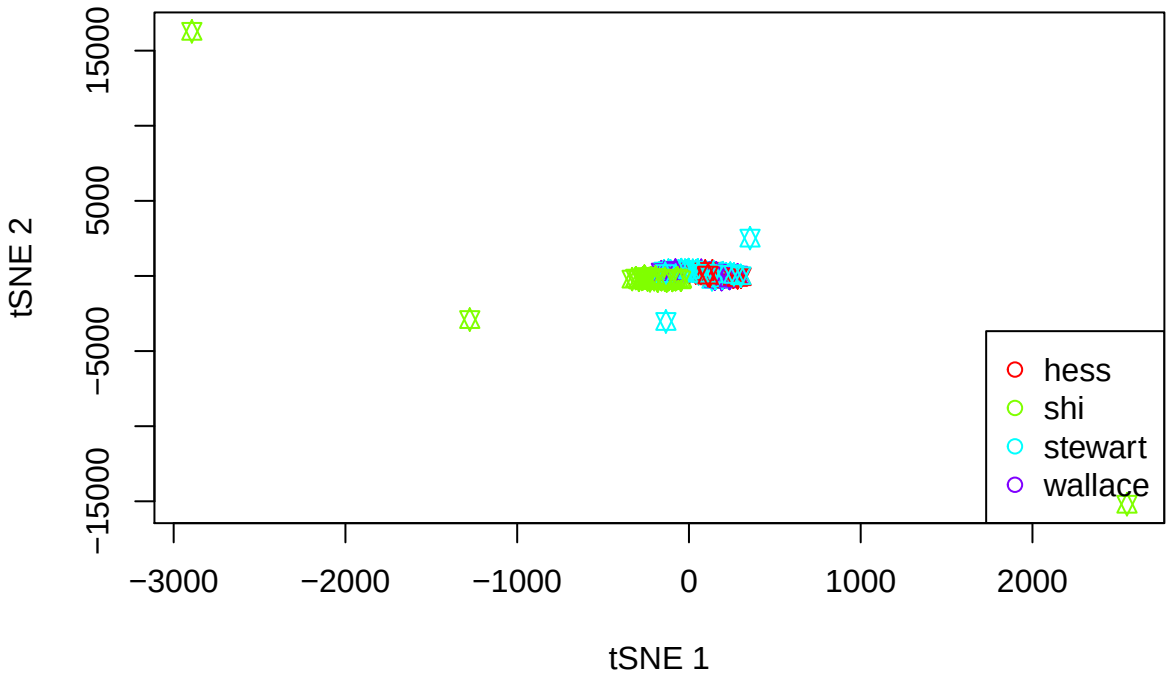


Figure 7: Plots of t-distributed stochastic neighbor embedding (tSNE) using AMR class, mechanism, and group data shows that Shi always clusters alone.

Relative Abundance

```
coul <- brewer.pal(10, "Spectral")
par(oma=c(1,1,1,10))
barplot(as.matrix(MicRelAbsClean), col=coul,xlab="Rumen Studies",xaxt="n",cex.lab=1,ylab="Rela
legend(123,0.9,legend=rownames(as.matrix(MicRelAbsClean)),col = coul,box.lty=0,lty=1,lwd=3,xpd=
```

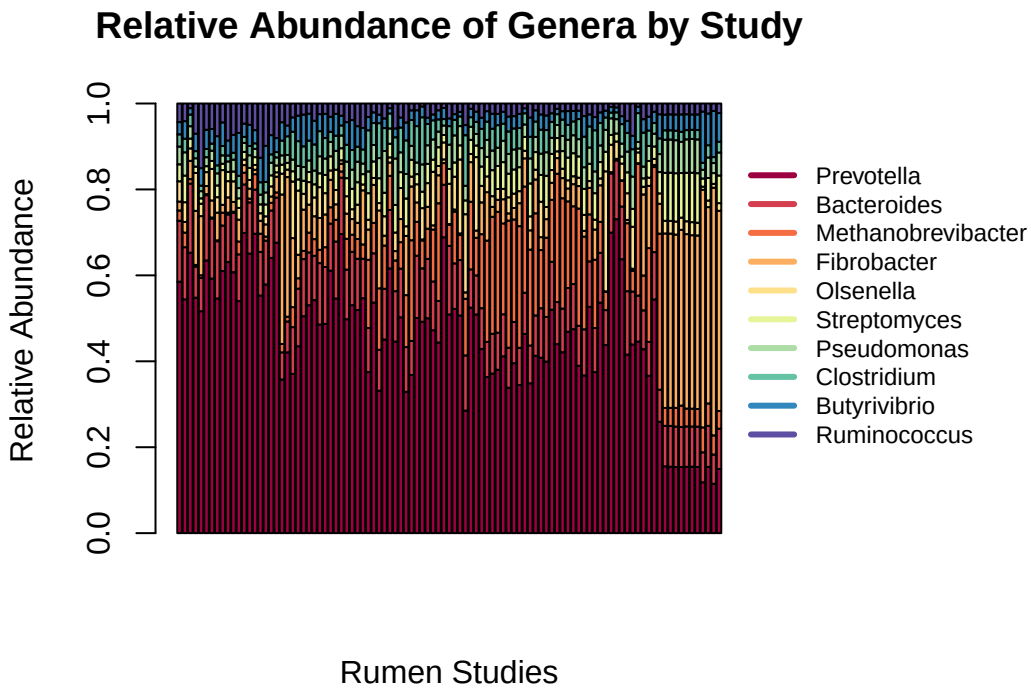


Figure 8: Relative abundance was calculated using the microbiome count data. From this, the top 10 most abundant bacterial genera were extracted and plotted over all the samples. Provetella was the most abundant in most of the samples, but Hess seemed to have Fibrobacter as the most abundant genera. Therefore this sample group was appended to the right end of the figure for clarity.

```
#class
coul <- brewer.pal(10, "Spectral") #change the color order to make it easier to see
coul<-c(coul[7], coul[5], coul[4], coul[1], coul[2], coul[3], coul[6], coul[8:10])
par(oma=c(1,1,1,10))
barplot(as.matrix(ClassRelAbsClean), col=coul,xlab="Rumen Studies",xaxt="n",cex.lab=1,ylab="Rel
legend(123,0.9,legend=rownames(as.matrix(ClassRelAbsClean)),col = coul, box.lty=0, lty=1, lwd=3
```

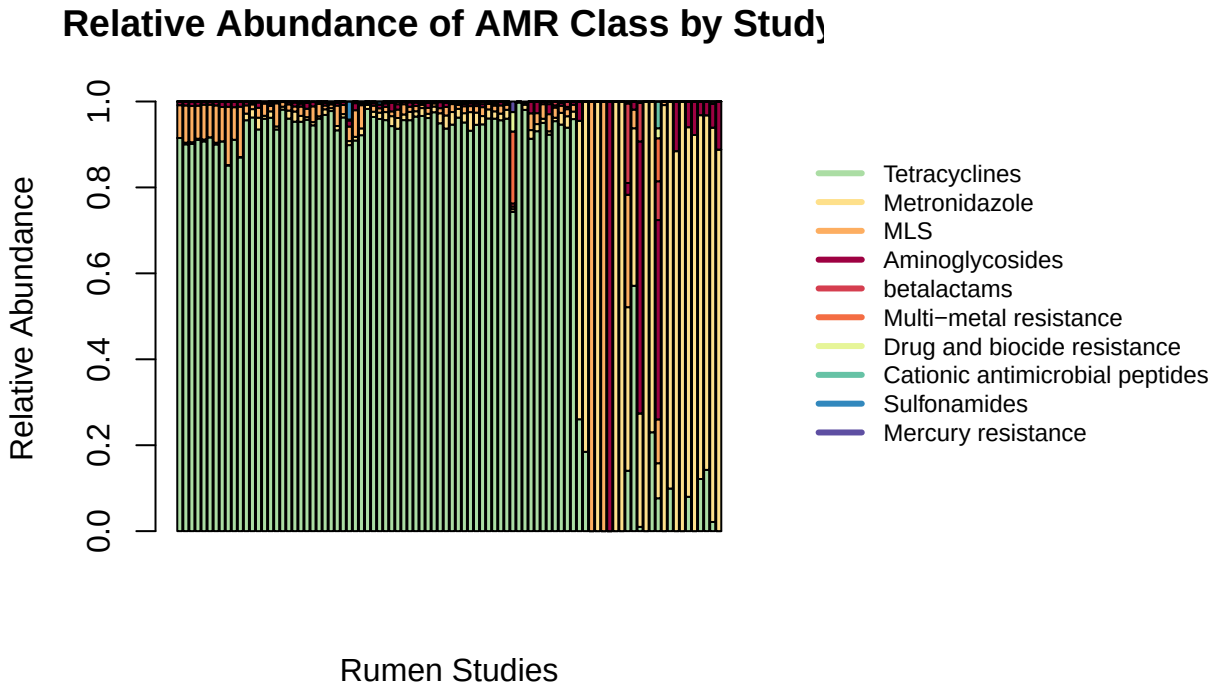


Figure 9a: Relative abundance was calculated using the AMR Class count data. From this, the top 10 most abundant AMR classes were extracted and plotted over all the samples. Tetracyclines were the msot abundant for most studies, but not in the Shi studies. Therefore this sample group was appended to the right end of the figure for clarity.

```
#mech
coul <- brewer.pal(10, "Spectral") #change the color order to make it easier to see
coul<-c(coul[7], coul[5], coul[4], coul[1], coul[2], coul[3], coul[6], coul[8:10])
par(oma=c(1,1,1,10))
barplot(as.matrix(MechRelAbsClean), col=coul,xlab="Rumen Studies",xaxt="n",cex.lab=1,ylab="Rela
legend(123,0.9,legend=rownames(as.matrix(MechRelAbsClean)),col = coul, box.lty=0, lty=1, lwd=3
```

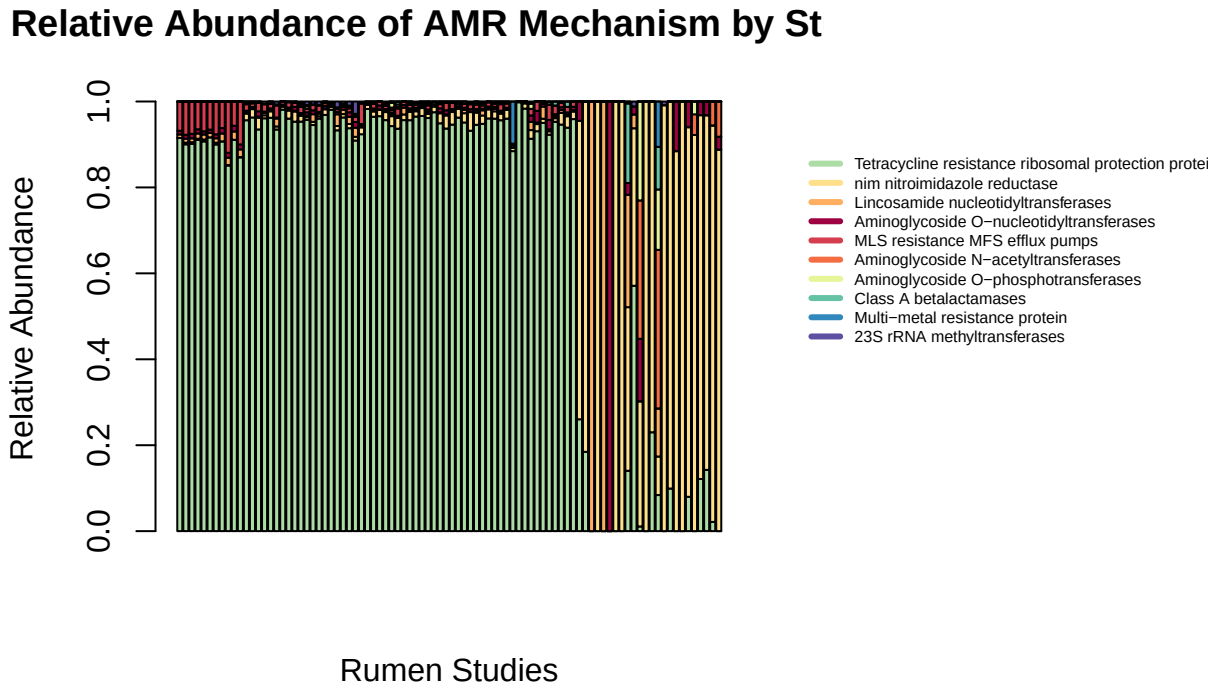


Figure 9b: Relative abundance was calculated using the AMR Mechanism count data. From this, the top 10 most abundant AMR mechanisms were extracted and plotted over all the samples. Tetracycline resistance ribosomal protection protein was the msot abundant for most studies, but not in the Shi studies. Therefore this sample group was appended to the right end of the figure for clarity.

```
#group
coul <- brewer.pal(10, "Spectral") #change the color order to make it easier to see
coul<-c(coul[7], coul[5], coul[4], coul[1], coul[2], coul[3], coul[6], coul[8:10])
par(oma=c(1,1,1,10))
barplot(as.matrix(GroupRelAbsClean), col=coul, xlab="Rumen Studies",xaxt="n",cex.lab=1,ylab="R
legend(123,0.9,legend=rownames(as.matrix(GroupRelAbsClean)),col = coul, box.lty=0, lty=1, lwd=
```

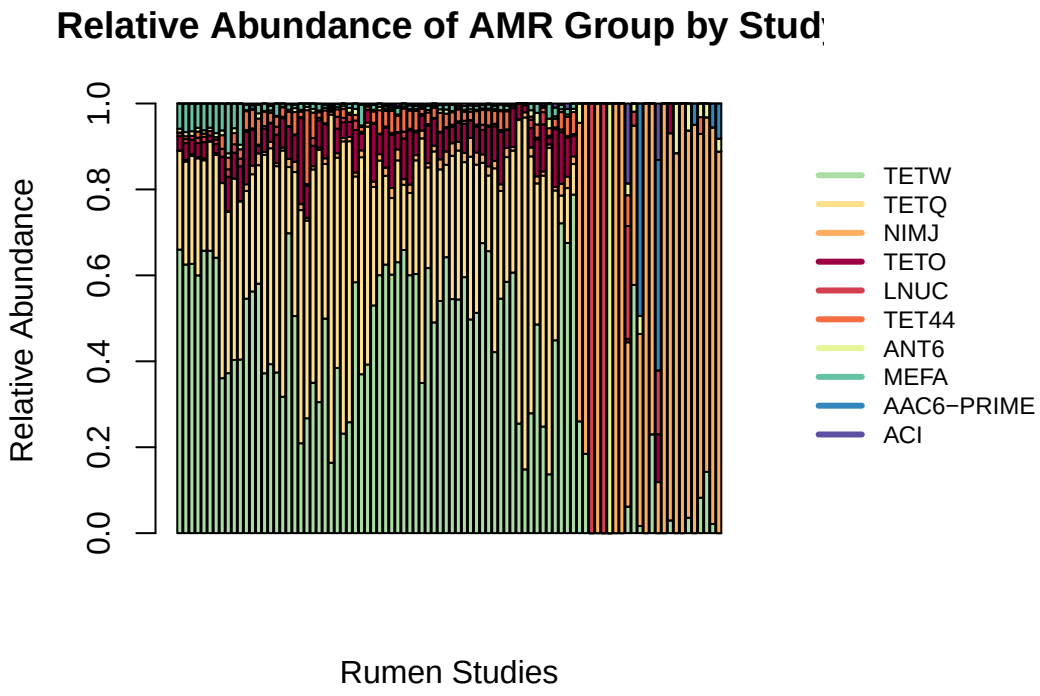


Figure 9c: Relative abundance was calculated using the AMR Group count data. From this, the top 10 most abundant AMR groups were extracted and plotted over all the samples. TETW and TETQ were among the msot abundant groups for most studies, but not in the Shi studies. Therefore this sample group was appended to the right end of the figure for clarity.

Baysian Network

```
renderGraph(sviz)
```

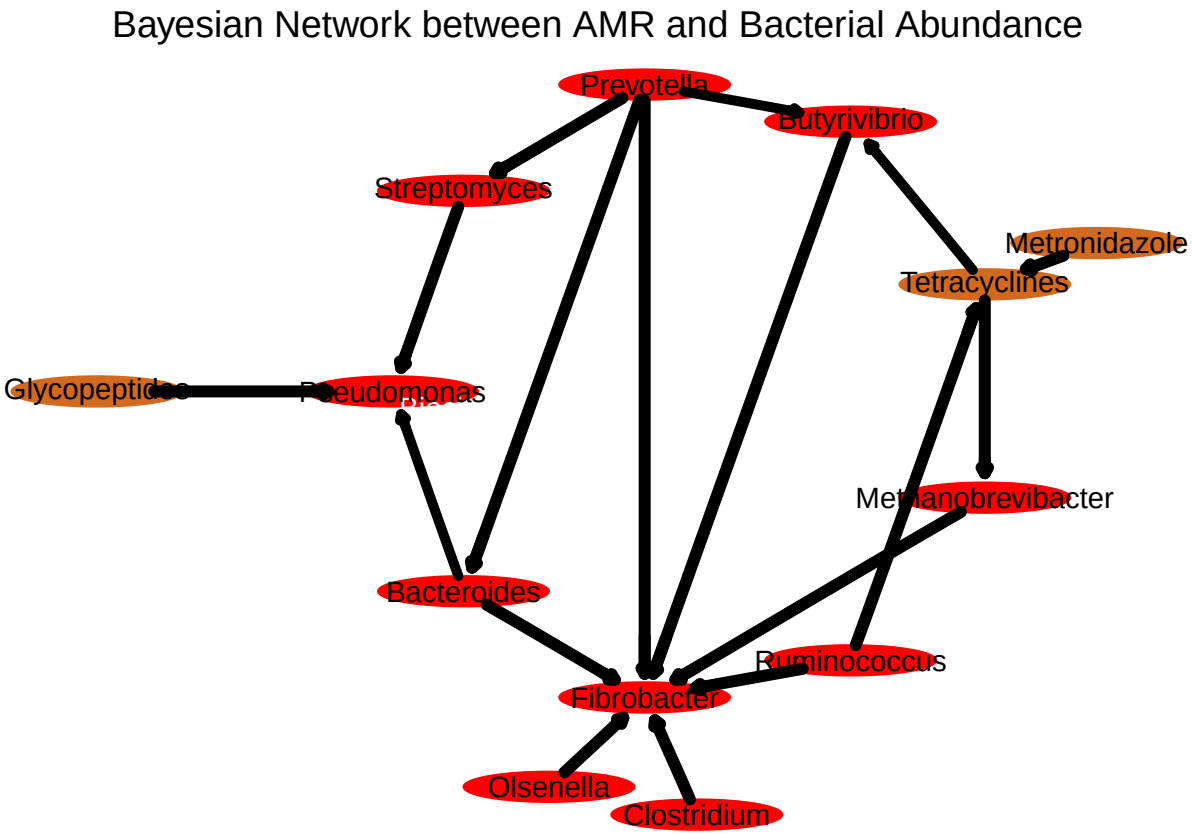


Figure 10: Bayesian Network between relative abundances of the AMR classes and the 10 most abundant genera of bacteria. AMR and bacterial genera data were derived from the same sequence data. In red are bacteria and in brown are the classes of AMR.

One notable trend in this graph is that the abundance of many of the bacterial genera were correlated with Fibrobacter. Additionally, Prevotella had a direct influence on Streptomyces, Butyrivibrio, Fibrobacter, and Bacteroides. These relationships were interesting; however, the key points to note here were the edges between genera and AMR. Notably, Glycopeptide-class AMR had a direct influence on the relative abundance of Pseudomonas. Additionally, Ruminococcus has a direct influence on Tetracycline-class AMR which in turn has a direct influence on Methanobrevibacter.

Conclusions:

In investigating the relationships between bacterial abundance and antimicrobial resistance as well as the design of the studies from which data was collected, it becomes clear that each of these factors influenced the others. More specifically, AMR and bacterial abundance data are notably similar across studies with the exception of Hess in which Fibrobacter was the dominant genus and Shi in which Metronidazole was the dominant class of AMR.

It was found that principal components analysis, multidimensional scaling, and tSNE all revealed similar results. The bacterial abundance data did not separate into distinct clusters except for the Hess study data. Conversely, the AMR data did cluster somewhat better with the Shi and Hess study data being distinct. This was an expected outcome due to the fact that Hess used an unusual feed for cattle [3], and Shi collected rumen samples from sheep [5].

It was also found that tetracycline resistance was the most abundant AMR mechanism in almost all samples other than Shi study. There are two types of tetracycline resistance: tetracycline efflux and ribosomal protection [6]. Our investigation found ribosomal protection as the most commonly occurring mechanism. Interestingly, it has been found that the Bacteroides group often uses this type of resistance [6]. However, this was not reflected in our Bayesian network analysis.

It's important to note that AMR and bacterial abundance influence one another as evidenced by Bayesian network analysis. In particular, Glycopeptide resistance has a strong influence on the relative abundance of Pseudomonas. This is an interesting yet not unexpected result – glycopeptide antibiotics are known to be effective against multi-drug resistant Gram-negative bacteria such as Pseudomonas aeruginosa [7]. Another important relationship to consider is that between Metronidazole and Tetracycline resistance. It is possible that the influence the former has on the latter is due to the fact that metronidazole-resistant bacteria are common, and, consequently, tetracyclines are used as a first-line defense [8].

We used rumen data to investigate relationships between rumen microbiome and antimicrobial resistance genes. We found that it is possible to conclude such relationships with the use of clustering techniques and Bayesian network analysis. This type of modeling could be useful in tentative predictions of types of antibiotic treatments to avoid livestock that may be resistant. Another application is guiding the understanding of bacterial resistance to antimicrobials. More knowledge on antimicrobial resistance will allow researchers to improve human and animal health.

References:

1. Pal, C., Bengtsson-Palme, J., Kristiansson, E. & Larsson, D. G. J. The structure and diversity of human, animal and environmental resistomes. Microbiome 4, 54 (2016).
2. Stewart, R. D. et al. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. Nat. Commun. 9, 870 (2018).
3. Hess, M. et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331, 463–467 (2011).
4. Wallace, R. J. et al. The rumen microbial metagenome associated with high methane production in cattle. BMC Genomics 16, 839 (2015).
5. Shi, W. et al. Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. Genome Res. 24, 1517–1525 (2014).
6. Speer, B., Shoemaker, N., & Salyers, A. Bacterial resistance to tetracycline: mechanisms, transfer, and clinical significance. Clin Microbiol Rev 5, 4 (1992).
7. Yarlagadda, V. et al. Glycopeptide Antibiotic To Overcome the Intrinsic Resistance of Gram-Negative Bacteria. ACS Infect Dis 2, 2 (2016).
8. Dailidiene, D. et al. Emergence of Tetracycline Resistance in Helicobacter pylori: Multiple Mutational Changes in 16S Ribosomal DNA and Other Genetic Loci. Mechanisms of Resistance 46, 12 (2002).