TAXONOMIC AND METAGENOMIC APPROACHES
TO DIET, HEALTH, AND BREEDS OF RUMINANTS

A Dissertation
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

By
ROCKY PATIL
Dr. Gavin Conant, Dissertation Supervisor
December 2020

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

TAXONOMIC AND METAGENOMIC APPROACHES
TO DIET, HEALTH, and BREEDS OF RUMINANTS

presented by Rocky Patil,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Gavin Conant

Professor William Lamberson

Professor Christine Elsik

Professor J. Chris Pires

ACKNOWLEDGEMENTS

I would like to sincerely thank my Ph.D. adviser, Dr. Gavin Conant for giving me the opportunity to join his evolutionary systems lab group at the University of Missouri – Columbia. Dr. Conant has been a supportive mentor by providing me with constructive criticism and guidance in my research projects. I would like to thank my former lab members: Dr. Yue Hao, Dr. Huan Truong, and Sara Wolff for being great lab mates by always being supportive in the lab and always willing to help me whenever I needed it.

I would like to thank all the members of my committee: Dr. William Lamberson, Dr. Christine Elsik, Dr. J. Chris Pires, and Dr. Tieming Ji for their support and guidance in my research projects. Furthermore, I would like to thank our research collaborators: Dr. Kristi Cammack and Dr. Melinda Ellison for providing us with all our rumen samples and their expertise in helping us in completing our projects.

Finally, I would like to thank my parents for all their support and unconditional love over the years as they were always there for me. I am forever grateful for my wife, Bebi, who has been patient and understanding as I finish my Ph.D. I am thankful for everything she had done for me.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	II
LIST OF TABLES.....	V
LIST OF FIGURES.....	VI
ABSTRACT.....	VII
Chapters	
1. INTRODUCTION.....	1
1.1 References.....	8
2. POOR FEED EFFICIENCY IN SHEEP IS ASSOCIATED WITH SEVERAL STRUCTURAL ABNORMALITIES IN THE COMMUNITY METABOLIC NETWORK OF THEIR RUMINAL MICROBES.....	14
2.1 Abstract.....	15
2.2 Introduction.....	16
2.3 Materials and Methods.....	17
2.4 Results.....	25
2.5 Discussion.....	28
2.6 References.....	41
3. A METAGENOMIC ANALYSIS OF THE EFFECT OF ANTIBIOTIC FEED ADDITIVES ON THE OVINE RUMEN METABOLISM.....	50
3.1 Abstract.....	51
3.2 Introduction.....	52
3.3 Materials and Methods.....	53
3.4 Results.....	60
3.5 Discussion.....	63
3.6 References.....	75

4. HOST GENETICS AND THE TAXONOMIC AND FUNCTIONAL STRUCTURE OF THE BOVINE RUMIAL MICROBIAL ECOSYSTEM.....	83
4.1 Abstract.....	83
4.2 Introduction.....	83
4.3 Materials and Methods.....	87
4.4 Results.....	94
4.5 Discussion.....	97
4.6 References.....	105
5. SUMMARY.....	114
VITA.....	115

LIST OF TABLES

2.1. RFI group animal read statistics.....	33
2.2. RFI group network structure.....	34
2.3. RFI group and network position.....	35
3.1. ANT/NON group animal read statistics.....	67
3.2. Differences in network layer and antibiotic treatment.....	69
3.3. ANT/NON network structure.....	70
3.4 Pearson's and Spearman's correlation test.....	71
4.1 Breed group animal read statistics.....	100
4.2 Breed group and network position.....	102
4.3 Breed group network structure.....	103

LIST OF FIGURES

2.1. The interaction between the host and microbial metabolic networks for efficient and inefficient sheep.....	36
2.2. First two principal components (x and y -axes, respectively) from PCAs of the OTUs.....	38
2.3. Shannon entropies of the distribution of microbial reads onto OTUs (x -axis) and onto metabolic network nodes (y -axis) for the low and high RFI animals.....	39
3.1. The distribution of Shannon entropies of the mapped OTUs (x -axis) and metabolic network nodes (y -axis) for ANT and NON-fed animals.....	72
3.2. Microbial diversity between ANT and NON-fed sheep.....	73
3.3. The interaction between the host and microbial metabolic network for ANT and NON-fed sheep.....	74
4.1. Mean layer of the individual animal samples for ANG and CHA for N_{50}	104

TAXONOMIC AND METAGENOMIC APPROACHES TO DIET, HEALTH, AND BREEDS OF RUMINANTS

Rocky Patil

Dr. Gavin Conant, Dissertation Supervisor

ABSTRACT

Ruminant mammals maintain a symbiotic relationship with their gut microbiome, the organisms of which break complex plant matter into metabolites that are absorbed by the host animal for its nutrition. Building on prior work describing the taxonomic composition of this ecosystem, we sought to explore its biochemical operation with a systems biology approach. We constructed a merged host/microbial metabolic network by using shotgun metagenomic data matched to a reference enzyme database. We used this approach to explore the complexities of the rumen ecosystem and to assess the effects of feed additives in diet, feed efficiency, and breed-differences on this ecosystem. We linked the metabolic network of the microbial community to that of the host animal by defining a set of interface metabolites shared between them. The value of our approach is illustrated with our comparison of the rumen microbiomes of animals of low and high feed efficiencies. Strikingly, there was no individual enzyme that was statistically different in its abundance between these two groups. However, by analyzing the network as a whole, we observed that the high efficiency animals had a microbial network that was closer in network space to the host's reactions. In our antibiotics study, we observed that there were changes in the taxonomic and functional structure when sheep were fed a diet supplemented with antibiotics. Sheep fed antibiotics had a set of microbial enzymes that were closer in in the network space to the host metabolism than

were the enzymes of the microbes of sheep not fed antibiotics. There were 546 individual microbial reactions that were significantly different between the two groups. Finally in the third study described, we found that breed differences in these community metabolic networks between Angus and Charolais cattle were only detected by analyzing the global network structure: there were no individual microbial taxa or enzymes that differed significantly between the breeds. These studies have shown the importance of using a metabolic network approach to better characterize the interactions between gut microbiota and host organisms as compared to solely studying the individual taxonomic and metabolic reactions.

INTRODUCTION

Ruminant animals depend on symbiotic microbes in their gut microbiome to break down cellulose and other complex plant material into simpler metabolites that can be used for nutrition by the host animal (Stevens & Hume, 1998). This symbiosis is a critical adaption for these mammals, giving them access to an ecological niche unavailable to other mammals because mammalian biochemistry on its own cannot access such carbon sources. Several studies have shown that changes to microbial taxa that make up the microbiota of the gastrointestinal tract are linked to an increased susceptibility to several diseases for the host such as IBD and colorectal cancer (Marchesi et al., 2016). In studies in agricultural animals, these taxonomic studies have also been valuable in assessing how microbial abundance and diversity in the rumen can affect performance traits that are of interest economically (Mizrahi & Jami, 2018). For example, species from the phylum Firmicutes are involved with starch hydrolysis and their abundance has a moderate level of heritability, suggesting a role for host genetics (Li & Guan, 2019). However while useful, taxonomic studies in general leave alone unanswered the question of why such changes to the microbiome alter traits of interest. A key reason for the lacuna is that taxonomic information alone does not fully capture the possible interactions between the microbes themselves and between the microbes and their host: indeed both types of interaction are still poorly understood.

Studies in mice have shown that diet can induce profound changes in the gut microbial community structure and also in the host. For instance, a study by Turnbaugh et al. demonstrated that transplanting microbes from obese mice into a germ-free mice resulted in increased ability to harvest energy from the diet in those animals, resulting in

increased body fat (Turnbaugh et al., 2006). Another mice study by Ley et al. showed that obesity was caused by increased proportional change of Bacteroidetes (Ley et al., 2005). In addition to their intrinsic interest, these studies lay the groundwork for using the experimental manipulation of the host diet to explore interactions between the rumen microbes and their host.

Initially, studies to characterize rumen microbes were conducted with culture-based techniques and molecular methods. Traditional culturing methods rely on bacteria being grown on culture media to identify microbes in the rumen (Koch, 1982). This approach has significantly underestimated the rumen microbial diversity because it is unable to quantify the many unculturable organisms: the organisms identified account for less than 1% of the total number of rumen microbial species (McCabe et al., 2015). Moreover, bacterial species inferred to have large populations after culturing in such media may actually be proportionally much less abundant in natural ecosystems, suggesting that culture-independent methods will provide a better estimate of species richness (species diversity) in these ecosystems (Kisand & Wikner, 2003). The field is also complicated by the difficulties in defining taxonomic groups among primarily asexually reproducing single-celled organisms, and by the fact that the phenotypic characterizations used to do so differ as such is made by the discovering scientist (Staley, 2006). A slightly *ad hoc* solution to this problem was to make genotypic classifications assuming that a 70% similarity in DNA hybridization would indicate a classification within the same bacterial species (Wayne, 1988). Further complexity is added by the facts that many bacterial species have not been grown in pure cultures and hence are

known only from molecular methods (Hugenholtz, Goebel, & Pace, 1998) and that the history of these lineages is replete with horizontal gene transfer (Staley, 2006).

The advent of next-generation sequencing technologies has advanced prediction of gut microbiota function through the availability of abundant 16s rRNA sequencing data (Shendure & Ji, 2008). It has been estimated that approximately 99% of prokaryotes cannot be cultured in isolation, which limits our understanding of the microbial interactions occurring in the ecosystem (Schloss & Handelsman, 2005). PCR amplification of 16s rRNA short-reads are used to generate OTU sequences that are used to infer taxonomy (Johnson et al., 2019). The 16s rRNA gene was selected for bacterial taxa identification because it is found in almost all bacteria (and archaeans), and its function has not changed over time. As a result it represents a “clock-like” gene that is useful as a phylogenetic marker (Janda & Abbott, 2007). The PCR amplification of rRNA genes is limited in assessing diversity in communities because it has difficulties discriminating between closely related bacterial species (Konstantinidis, Ramette, & Tiedje, 2006). Part of the reason for these issues is the fact that microbes in the gut go through genetic transformations such as mutations, lateral gene transfer, and chromosomal rearrangements making it more difficult to characterize them (Ley, Turnbaugh, Klein, & Gordon, 2006).

With the early use of Sanger-sequencing, scientists were able to determine the microbiota richness in a wide range of environments across the biosphere (Ley, Lozupone, Hamady, Knight, & Gordon, 2008). This sequencing approach advanced the field of metagenomics by allowing for the detection of unculturable prokaryotes directly from microbial DNA extracted from an environmental sample. For example, Tringe et al.

used shotgun sequencing to characterize metabolic capabilities and discover specific genes in different marine and terrestrial environments (Tringe et al., 2005). Also, Tyson et al. used shotgun sequencing to sequence full complement of organisms in a biofilm to determine pathways for carbon fixation and energy production as possible survival mechanisms (Tyson et al., 2004).

Many of the early metagenomic studies of ruminants were taxonomic in their focus, yielding valuable insights. However, they were necessarily limited in the functional information they recovered. Recent studies have focused on gaining a better understanding of the ecosystem on studying the rumen microbiome as an interwoven network of species metabolism by using systems biology approaches (Greenblum, Turnbaugh, & Borenstein, 2012). To discover and characterize the interactions between the host and its microbiome, scientists have mapped metagenomic reads to a metabolic pathway database such as KEGG or MetaCyc to predict metabolic pathways from the annotation of an organism's genome (Caspi et al., 2014; Kanehisa & Goto, 2000). This approach of predicting a metabolic pathway relies on the accuracy of the reference metabolic pathways in the database (Altman, Travers, Kothari, Caspi, & Karp, 2013). The increasingly large and complex metagenomic datasets that are being obtained from next-generation DNA sequencing is underscoring the importance of developing new computational tools and models to better analyze these datasets.

The objectives of our studies were to provide a greater understanding of rumen metabolism by using shotgun metagenomic data to construct a metabolic network of all the reactions present in this ecosystem (Zhu et al., 2007). A metabolic network represents a set of biochemical reactions and metabolites of a particular genome or

environment (Ma & Zeng, 2003). The nodes in the network are reactions and are connected by edges when they share metabolites. One of the first studies of metabolic networks constructed was that of *Saccharomyces cerevisiae*, which increased the knowledge of yeast metabolism (Duarte, Herrgård, & Palsson, 2004). Previously, Greenblum et al. employed a similar approach to mapping reads to an enzyme database to study the microbial metabolism in the human gut of individuals who were obese and had inflammatory bowel disease (Greenblum et al., 2012). Their study revealed differences in the interactions between human and microbial metabolism when comparing lean and obese individuals, differences that were particularly evident when considering the interface of the microbiome and the host's metabolism. For instance, it showed that obese microbiomes have a lower centrality value indicating that lean microbiomes were more thoroughly metabolically connected than were obese ones (Hahn & Kern, 2005).

Previously, our lab group used a similar strategy of mapping reads to the MetaCyc database to explore the complex gut microbiome of ruminant animals such as cattle and sheep (Taxis et al., 2015 and Wolff et al., 2017). We analyzed the community metabolic networks of the ruminant microbes using shotgun metagenomic data. In this approach, we enhanced our understanding of rumen metabolism by verifying the interface metabolites that are shared between the microbial and host animal's metabolic network. We created three sets of metabolites that are presumed to be exchanged between the microbe and host network defined as VFA, VFA + AA, and ALL (Taxis et al., 2015). The first set of interface metabolites consists of acetate, butyrate, and propionate which are three most common volatile fatty acids in the rumen (VFAs) (Bergman, 1990). The second set adds 20 universal amino acids to the previous VFA set (VFA + AA). The

third set contains the VFA + AA with additional metabolites that are absorbed by various human cells which were obtained from the reconstruction of the human metabolic network (Duarte et al., 2007). This final set of interface metabolites (ALL) includes a total of 204 compounds excluding DNA/RNA, nucleotides, carbon dioxide, and NAD⁺ and NADP⁺. By employing a community metabolic approach to study two steer, it was shown that ruminal ecosystems had relatively different taxonomic profiles while their metabolic networks, and especially their interface with the host network were conserved (Taxis et al., 2015). We later used this pipeline to explore how changes in sheep diets altered the structure of their microbial metabolic network. We found that sheep fed a concentrate-based diet (grain-based diet) had substrates with shorter carbon structures and were mapped to layers closer in proximity to the host's metabolism compared with sheep fed a forage-based diet (composed primarily of alfalfa) (Wolff et al., 2017). Furthermore, it was determined that there were a significant number of metabolic reactions differing between the sheep's diets.

In the three studies encompassing my dissertation, we wanted to employ the previous computational pipeline to gain a greater understanding of the rumen ecosystem for traits of economic value in ruminant animals. We again used a microbial metabolic community network, which was inferred from shotgun metagenomic sequences to assess changes in the metabolic system when comparing feed efficiency, feed additives, and breed differences in ruminants. When comparing the metagenomes of ovine individuals that differed in feed efficiency, we found that there were no individual enzymes that were significantly different in their abundance between low and high RFI animals. However by analyzing the complete metabolic network, microbes from low RFI animals were seen

to more commonly use enzymes closer in proximity to the host's reactions compared with high RFI animals (Patil et al., 2018). In another study, we compared the metabolic networks of sheep fed antibiotic-supplemented diet to those of animals fed a non-supplemented diet. Sheep fed antibiotics had metabolic reactions that were closer to the host's network space compared with sheep fed non-antibiotics. Finally, we employed our network framework to evaluate how the microbiomes of different bovine breeds (Angus and Charolais) differed in their microbial ecology. We found that Charolais animals had microbial ecosystems that used reactions closer to the host's reactions than did the ecosystems found in Angus animals.

The use of metabolic networks can be used to improve animal and human health and welfare in a variety of ways. For example, researchers have discovered biomass-degrading genes in the cow rumen that could be utilized in the conversion of cellulose biomass to biofuels (Hess et al., 2011). Using the metabolic networks approach, we are able to gain a better understanding of how diet and feed efficiency can affect the metabolic capacities in the rumen ecosystem, which is not seen only examining the taxonomy (Patil et al., 2018; Wolff et al., 2017). The information obtained from these studies could also be an initial step toward approaches that improve economics and sustainability of animal agriculture.

REFERENCES

- Altman, T., Travers, M., Kothari, A., Caspi, R., & Karp, P. D. (2013). A systematic comparison of the MetaCyc and KEGG pathway databases. *BMC Bioinformatics*, *14*, 112. doi:10.1186/1471-2105-14-112
- Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., . . . Karp, P. D. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res*, *42*(Database issue), D459-471. doi:10.1093/nar/gkt1103
- Duarte, N. C., Herrgård, M. J., & Palsson, B. (2004). Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res*, *14*(7), 1298-1309. doi:10.1101/gr.2250904
- Greenblum, S., Turnbaugh, P. J., & Borenstein, E. (2012). Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc Natl Acad Sci U S A*, *109*(2), 594-599. doi:10.1073/pnas.1116053109
- Hahn, M. W. and A. D. Kern. "Comparative Genomics of Centrality and Essentiality in Three Eukaryotic Protein-Interaction Networks." *Mol Biol Evol*, vol. 22, no. 4, 2005, pp. 803-806, doi:10.1093/molbev/msi072.

- Hess, M., Sczyrba, A., Egan, R., Kim, T. W., Chokhawala, H., Schroth, G., . . . Rubin, E. M. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science*, 331(6016), 463-467. doi:10.1126/science.1200387
- Hugenholtz, P., Goebel, B. M., & Pace, N. R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol*, 180(18), 4765-4774. doi:10.1128/JB.180.18.4765-4774.1998
- Janda, J. M., & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol*, 45(9), 2761-2764. doi:10.1128/JCM.01228-07
- Johnson, J. S., Spakowicz, D. J., Hong, B. Y., Petersen, L. M., Demkowicz, P., Chen, L., . . . Weinstock, G. M. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun*, 10(1), 5029. doi:10.1038/s41467-019-13036-1
- Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1), 27-30. doi:10.1093/nar/28.1.27
- Kisand, V., & Wikner, J. (2003). Combining culture-dependent and -independent methodologies for estimation of richness of estuarine bacterioplankton consuming riverine dissolved organic matter. *Appl Environ Microbiol*, 69(6), 3607-3616. doi:10.1128/aem.69.6.3607-3616.2003
- Koch, R. (1982). Classics in infectious diseases. The etiology of tuberculosis: Robert Koch. Berlin, Germany 1882. *Rev Infect Dis*, 4(6), 1270-1274.

- Konstantinidis, K. T., Ramette, A., & Tiedje, J. M. (2006). The bacterial species definition in the genomic era. *Philos Trans R Soc Lond B Biol Sci*, 361(1475), 1929-1940. doi:10.1098/rstb.2006.1920
- Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A*, 102(31), 11070-11075. doi:10.1073/pnas.0504978102
- Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R., & Gordon, J. I. (2008). Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol*, 6(10), 776-788. doi:10.1038/nrmicro1978
- Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Microbial ecology: human gut microbes associated with obesity. *Nature*, 444(7122), 1022-1023. doi:10.1038/4441022a
- Li, F., Li, C., Chen, Y., Liu, J., Zhang, C., Irving, B., . . . Guan, L. L. (2019). Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. *Microbiome*, 7(1), 92. doi:10.1186/s40168-019-0699-1
- Ma, H., & Zeng, A. P. (2003). Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*, 19(2), 270-277. doi:10.1093/bioinformatics/19.2.270
- Marchesi, J. R., Adams, D. H., Fava, F., Hermes, G. D., Hirschfield, G. M., Hold, G., . . . Hart, A. (2016). The gut microbiota and host health: a new clinical frontier. *Gut*, 65(2), 330-339. doi:10.1136/gutjnl-2015-309990

- McCabe, M. S., Cormican, P., Keogh, K., O'Connor, A., O'Hara, E., Palladino, R. A., . . . Waters, S. M. (2015). Illumina MiSeq Phylogenetic Amplicon Sequencing Shows a Large Reduction of an Uncharacterised Succinivibrionaceae and an Increase of the Methanobrevibacter gottschalkii Clade in Feed Restricted Cattle. *PLoS One*, 10(7), e0133234. doi:10.1371/journal.pone.0133234
- Mizrahi, I., & Jami, E. (2018). Review: The compositional variation of the rumen microbiome and its effect on host performance and methane emission. *Animal*, 12(s2), s220-s232. doi:10.1017/S1751731118001957
- Patil, R. D., Ellison, M. J., Wolff, S. M., Shearer, C., Wright, A. M., Cockrum, R. R., . . . Conant, G. C. (2018). Poor feed efficiency in sheep is associated with several structural abnormalities in the community metabolic network of their ruminal microbes. *J Anim Sci*, 96(6), 2113-2124. doi:10.1093/jas/sky096
- Schloss, P. D., & Handelsman, J. (2005). Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. *Genome Biol*, 6(8), 229. doi:10.1186/gb-2005-6-8-229
- Shendure, J., & Ji, H. "Next-Generation Dna Sequencing." *Nat Biotechnol*, vol. 26, no. 10, 2008, pp. 1135-1145, doi:10.1038/nbt1486.
- Staley, J. T. (2006). The bacterial species dilemma and the genomic-phylogenetic species concept. *Philos Trans R Soc Lond B Biol Sci*, 361(1475), 1899-1909. doi:10.1098/rstb.2006.1914

- Stevens, C. E., & Hume, I. D. (1998). Contributions of microbes in vertebrate gastrointestinal tract to production and conservation of nutrients. *Physiol Rev*, 78(2), 393-427.
- Taxis, T. M., Wolff, S., Gregg, S. J., Minton, N. O., Zhang, C., Dai, J., . . . Conant, G. C. (2015). The players may change but the game remains: network analyses of ruminal microbiomes suggest taxonomic differences mask functional similarity. *Nucleic Acids Res*, 43(20), 9600-9612. doi:10.1093/nar/gkv973
- Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., . . . Rubin, E. M. (2005). Comparative metagenomics of microbial communities. *Science*, 308(5721), 554-557. doi:10.1126/science.1107851
- Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, 444(7122), 1027-1031. doi:10.1038/nature05414
- Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., . . . Banfield, J. F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, 428(6978), 37-43. doi:10.1038/nature02340
- Wayne, L. G. (1988). International Committee on Systematic Bacteriology: announcement of the report of the ad hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Zentralbl Bakteriol Mikrobiol Hyg A*, 268(4), 433-434. doi:10.1016/s0176-6724(88)80120-2

Wolff, S. M., Ellison, M. J., Hao, Y., Cockrum, R. R., Austin, K. J., Baraboo, M., . . .

Conant, G. C. (2017). Diet shifts provoke complex and variable changes in the metabolic networks of the ruminal microbiome. *Microbiome*, 5(1), 60.

doi:10.1186/s40168-017-0274-6

Zhu, X., Gerstein, M., & Snyder, M. "Getting Connected: Analysis and Principles of

Biological Networks." *Genes Dev*, vol. 21, no. 9, 2007, pp. 1010-1024,

doi:10.1101/gad.1528707.

Poor feed efficiency in sheep is associated with several structural abnormalities in the community metabolic network of their ruminal microbes¹

Rocky D. Patil,* Melinda J. Ellison,† Sara M. Wolff,* Courtney Shearer,‡ Anna M. Wright,|| Rebecca R. Cockrum,§ Kathy J. Austin,# William R. Lamberson,* Kristi M. Cammack,¶ and Gavin C. Conant*¹ *†‡¶§²

*Division of Animal Sciences, University of Missouri-Columbia, Columbia, MO, 65202

†Animal and Veterinary Science, Nancy M. Cummings Research, Extension, and Education Center, University of Idaho, Carmen, ID, 83462

‡School of Computing, Clemson University, Clemson, SC, 29634

||Department of Psychological Sciences, University of Missouri-Columbia, Columbia, MO, 65211

§Department of Dairy Science, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24061

#Department of Animal Science, University of Wyoming, Laramie, WY, 82071

¶Department of Animal Science, South Dakota State University, Brookings, SD, 57007

*†Program in Genetics, North Carolina State University, Raleigh, NC, 27695

*‡Bioinformatics Research Center, North Carolina State University, Raleigh, NC, 27695

*§Department of Biological Sciences, North Carolina State University, Raleigh, NC, 27695

¹Acknowledgements

This project was supported USDA National Research Initiative (NRI) grant # 2011-68006-30185 (MJE, WRL, KMC, GCC) and National Science Foundation grant NSF-DBI-1358997 (CS, AMW, GCC).

²Correspondence: G. Conant, gconant@ncsu.edu

This chapter has been published in the Journal of Animal Sciences.

ABSTRACT

Ruminant animals have a symbiotic relationship with the microorganisms in their rumens. In this relationship, rumen microbes efficiently degrade complex plant-derived compounds into smaller digestible compounds, a process that is very likely associated with host animal feed efficiency. The resulting simpler metabolites can then be absorbed by the host and converted into other compounds by host enzymes. We used a microbial community metabolic network inferred from shotgun metagenomics data to assess how this metabolic system differs between animals that are able to turn ingested feedstuffs into body mass with high efficiency and those that are not. We conducted shotgun sequencing of microbial DNA from the rumen contents of sixteen sheep that differed in their residual feed intake (RFI), a measure of feed efficiency. Metagenomic reads from each sheep were mapped onto a database-derived microbial metabolic network, which was linked to the sheep metabolic network by interface metabolites (metabolites transferred from microbes to host). No single enzyme was identified as being significantly different in abundance between the low and high RFI animals ($P > 0.05$, Wilcoxon Test). However, when we analyzed the metabolic network as a whole, we found several differences between efficient and inefficient animals. Microbes from low RFI (efficient) animals use a suite of enzymes closer in network space to the host's reactions than those of the high RFI (inefficient) animals. Similarly, low RFI animals have microbial metabolic networks that, on average, contain reactions using shorter carbon chains than do those of high RFI animals, potentially allowing the host animals to extract metabolites more efficiently. Finally, the efficient animals possess community networks with greater Shannon diversity among their enzymes than do inefficient ones.

Thus, our systems approach to the ruminal microbiome identified differences attributable to feed efficiency in the structure of the microbes' community metabolic network that were undetected at the level of individual microbial taxa or reactions.

INTRODUCTION

Ruminant animals have a symbiotic relationship with their ruminal microbiota, where microbial enzymes degrade complex polysaccharides indigestible to most animals (Stevens and Hume, 1998). Several studies have found a link between the presence of some bacterial lineages and differences in feed efficiency between individual ruminants (Guan et al., 2008; Hernandez-Sanabria et al., 2010; Rius et al., 2012; McCann et al., 2014). Greater feed efficiency would reduce animal agriculture's environmental impact while at the same time improving its economic efficiency (Herrero et al., 2013). Hence, modifying the ruminal microbiome represents a potentially powerful mechanism for improving producer outcomes. Unfortunately, as with many other features of microbial ecosystems, making a link between the structure of the microbiome and feed efficiency is difficult (Hernandez-Sanabria et al., 2012). In part, this difficulty is due to the fact that the techniques used to date have been taxonomically focused, even though many microbial ecosystems can display considerable taxonomic variation while preserving their underlying functions (Moore et al., 2009; Doolittle and Booth, 2017).

Using a metabolic network approach (Lacroix et al., 2008), we compared the enzymes found in the ruminal microbiomes of feed-efficient and feed-inefficient ewes. Strikingly, no single metabolic reaction was significantly different in its abundance between the two groups. However, by analyzing their community metabolic networks,

we were able to show that, as we hypothesized, the low RFI (efficient) animals used reactions that were located in closer proximity to the host network than were those of the high RFI animals, presumably allowing for the more efficient extraction of metabolites in the former. Furthermore, low RFI animals had metabolic networks with greater mean Shannon entropy, an increased network complexity that might enable them to exploit a greater proportion of the metabolites they consume.

MATERIALS AND METHODS

Feeding trial and rumen sample collection

All animal procedures were approved by the University of Wyoming Animal Care and Use Committee. Sixteen Targhee ewe lambs showing extremes of residual feed intake (selected from a surveyed group of 59 animals with initial body weight, BW, = 45.8 ± 2.5 kg) were selected for metagenomic sequencing (Ellison et al., 2017). All lambs were fed a forage-based diet primarily consisting of alfalfa and wheat middlings, and their individual feed intakes were measured with the GrowSafe System for each day of the 70-day trial (Airdrie, Alberta, Canada). The average daily gain (ADG) was computed using the difference between initial and final BW divided by the 70-day period. The metabolic mid-weight (MMWT) was calculated from the mid BW which was the body weight measured on the middle day of trial. For this dataset, residual feed intake (RFI) was calculated as the difference between actual feed intake and the expected feed intake. The expected feed intake for the animal's production was determined by regressing actual feed intake on ADG and MMWT (Cammack et al., 2005). The sheep were then ranked on their computed RFI values. We selected the 15% of animals

showing the most extreme low and high values of RFI for our metagenomics analyses. Thus, the eight sheep having the lowest RFI (the most feed efficient) were chosen as our low RFI group while the eight animals with the largest RFI were our high RFI group.

Rumen fluid from these 16 animals (≥ 30.0 mL) was suctioned through a tygon tube (length: 1 m, diameter: 1.5 cm) using a dosing syringe (400 mL) placed through the mouth down into the lamb's esophagus and running into the rumen. Rumen samples were collected after the feeding trial before ewes were returned back to private producer's breeding flock and immediately stored on dry ice until the samples were placed in long-term storage at -80°C .

DNA extraction and Illumina library preparation

Extraction of particulate matter from the fluid was achieved by centrifugation (Ellison et al., 2014). One ml of lysis buffer, silicon (0.1 g of 0.5 mm), and zirconia (0.3 g of 0.1 mm) beads were combined with thawed rumen samples to isolate DNA. The lysis buffer consisted of 500 mM NaCl, 50 mM Tris-CL, 50 mM EDTA, and 4% SDS. The mixture was homogenized with a Mini-Beadbeater-8 set at maximum speed for centrifugation of 16,000 g for 3 min followed immediately by incubation at 70°C for 15 min with gentle mixing every 5 min and centrifuged for 5 min at 4°C . After homogenization, the supernatant was transferred into a new tube followed by the addition of lysis buffer (300 μL). This procedure was repeated for a second time, and the supernatants combined. The QIAamp DNA Stool Mini Kit (Qiagen, Santa Clarita, CA) was used in the removal of proteins, RNA, and the precipitation of nucleic acids. Using a NanoDrop spectrophotometer (NanoDrop, Wilmington, DE), we found that the quality of

the DNA from the rumen samples was within acceptable ranges, with an A260/280 absorbance ratio ≥ 1.8 .

Genomic libraries were constructed for the 16 rumen samples based on the recommendations described in the Illumina's TruSeq DNA sample prep kit (Ellison et al., 2014). Diagenode BioRuptor was used to generate fragments of appropriately 350 base pairs by the shearing of genomic DNA. The fragmented DNA was made blunt-ended through an end repair reaction that removed 3' and 5' overhangs. At the 3' end of the blunt DNA molecule, a single adenosine was added with the ligation of Illumina adapters. The Agilent BioAnalyzer High Sensitivity DNA assay determined the fragment size while the Qubit assay quantified each of the sixteen purified libraries. Next, these libraries were diluted, multiplexed into four samples per lane and sequenced using an Illumina HiSeq 2500, producing 100 bp paired-end shotgun reads.

Sequence quality filtering and OTU analysis

We quality filtered the resulting reads by removing all bases in each read after the first three successive bases having a phred quality score of less than 15 (Ewing et al., 1998). Any read pair with one member less than 85 bases in length (after the above filtering) or having an average phred quality score of less than 25 was completely removed from the analysis (Table 1). We compared the filtered reads to 16S rDNA sequences from the Ribosomal Database Project (RDP, release 10.24) using Bowtie: read pairs both having sequence matches with greater than 97% identity to one or more database sequences were retained (Cole et al., 2009). Operational taxonomic units (OTUs) were defined using the sequences in the RDP itself: sequences in the database were connected by edges if their sequence identities were greater than 97% and OTUs

defined as connected components in this network (Ellison et al., 2014). We considered any of the above read pair hits that matched to sequences from one and only one such database-generated OTU to be an instance of that OTU.

MetaCyc

MetaCyc is a comprehensive enzyme and metabolic pathway database which contains more than 2,491 metabolic pathways from 2,816 organisms, primarily bacteria and archaeans, but also including some vertebrates (Caspi et al., 2014). The database describes enzymes, metabolites and reactions derived from the literature. It thus includes a set of enzyme sequences, each of which is associated with one or more biochemical reactions that it is believed to catalyze. In our previous work, we described how the contents of this database were used to infer a metabolic network, where nodes represent metabolic reactions and edges connect nodes that share a metabolite. Reactions with identical metabolites were merged into a single reaction (Wolff et al., 2017). The result was a set of 6,140 reaction nodes that could be potentially observed in a community metabolic network. Isolated reactions were omitted from this network.

In this study, two metabolic networks were constructed: one for the ovine host and one for the ruminal microbes. The MetaCyc database does not include an ovine metabolic network: instead we inferred one from those of humans and cattle. To do so, we used orthology inference to determine which of the enzymes for the 2,863 annotated reactions of the human metabolic network and the 1,404 reactions of the bovine network we could identify orthologs for in the ovine genome (Jiang et al., 2014). There were 15,562 human-sheep and 15,479 bovine-sheep ortholog pairs found with our synteny-based orthology program (Bekaert and Conant, 2011, 2014), which uses annotations from

Ensembl release 75 (Flicek et al., 2014). The ovine metabolic network was constructed from these ortholog pairs under the assumption that the presence of an orthologous enzyme in sheep indicates the presence of the corresponding reaction. Because MetaCyc has no vertebrate reaction allowing the assimilation of butyrate, a known nutrient for ruminants (Masson and Phillipson, 1951), we added a pseudoreaction converting butyrate to butyryl-CoA to the sheep metabolic network. In total, we inferred 1,989 nodes in this ovine metabolic network: 1,834 from the human network, 154 from the bovine one, and the butyrate to butyryl-CoA reaction just described.

To prevent the presence of ubiquitous metabolites (often called “currency metabolites”) from creating a nearly fully connected network, we excluded the most common metabolites when constructing network edges, using three thresholds to do so: namely metabolites that occurred in a total of more than 25, more than 50, or more than 100 reactions. We will hereafter refer to the resulting networks as N_{25} , N_{50} , and N_{100} (Pérez-Bercoff et al., 2011). For N_{25} , N_{50} , and N_{100} , there were 261, 206, and 174 currency metabolites removed, respectively.

Matching translated metagenomic reads to the microbial network

The Illumina data from the sixteen samples were translated to amino acid sequences in all six possible open reading frames (ORFs). Only paired amino acid sequences each having a translated ORF longer than 29 residues were retained and searched against the MetaCyc protein database. The sequence search against MetaCyc was made using a search tool based on the SeqAn library (Döring et al., 2008). We required two seven residue identical word matches to seed a Smith-Waterman local alignment between a database sequence and the translated read (Smith and Waterman,

1981). A match between the read pair and database sequences was confirmed if both reads produced alignments having greater than 80% amino acid identity over 80% of the metagenomic ORF. We assigned such a read pair to one or more metabolic network nodes so long as all of the enzyme sequences of the pair matched to belonged to the same overlapping set of nodes. Translated read pairs were not mapped if they matched two or more reactions with non-overlapping annotated enzymes: see (Wolff et al., 2017) for further details.

Inference of the interface between the host and microbial metabolic networks

To model the exchange of metabolites between microbes and the host, we defined sets of interface metabolites that were likely to be absorbed from the rumen by the host animal. Because the details of this transfer are not fully known, we employed three different metabolite sets. The first set, hereafter “VFA” for volatile fatty acids, is composed of the most common VFAs in the rumen: acetate, butyrate, and propionate (Bergman, 1990). Other volatile fatty acids are unfortunately not annotated in MetaCyc. The second interface set (VFA + AA) was defined by adding the 20 universal amino acids to the VFA set (for a total of 23 metabolites). The third interface set was composed of all the metabolites in the VFA + AA set with the addition of a subset of 234 metabolites which are known to be absorbed in the human gut and which were taken from the global reconstruction of the human metabolic network (Duarte et al., 2007). These compounds were matched to MetaCyc, and inorganic compounds, the DNA/RNA nucleotides, carbon dioxide, and NAD or NADP were removed. The remaining 204 metabolites were added to the VFAs and amino acids to yield the 227 metabolites in the “ALL” set. The complete list of these metabolites is available as supplemental

information from our previous paper (Taxis et al., 2015). The host and microbial networks were linked by adding edges between pairs of nodes, one from each network, that each employs the same interface metabolite (Fig. 1).

Metabolic network layers

The creation of this single metabolic network allows for the calculation of distances between reactions in the host and microbial metabolic networks. Dijkstra's algorithm was used to determine the shortest path between every pair of nodes in the merged metabolic network (Dijkstra, 1959). A node's distance to the other subnetwork then defined as its distance to the closest node in that subnetwork. Hence, nodes using interface metabolites are at the minimal distance of 0, with other nodes being successively more distant. The result of this distance definition is to decompose the merged metabolic network into the layers shown in Fig. 1.

Node read density analysis

We used a two-sample Wilcoxon test to assess if there were significant differences in the normalized read counts between the high and low RFI animals for the nodes in the microbial network. Only nodes with at least one mapped read pair in each of the two RFI groups were analyzed. The *P*-values from the Wilcoxon test were subjected to a 5% false-discovery rate correction (Benjamini, 1995).

Network structure analysis

The mean layer network for mapped reads were calculated separately for the low and high RFI status animals for all possible combinations of the different networks (N₂₅, N₅₀, and N₁₀₀) and the interface metabolites (3 x 3 = 9 comparisons). To determine whether there were significant differences between the mean layers of the two RFI

classes, we used a randomization approach. First, the mapped reads for the low and high RFI groups were pooled together and then the same number of reads as were originally mapped to the high RFI animals were randomly selected and assigned to a new “pseudo-high” set, with the remaining reads assigned to “pseudo-low”. We then compared the difference in mean layer number for the two classes for each of 1,000 of these randomized datasets to the difference in mean layer observed in the actual data. If the real difference was larger than all but 5% of the randomized datasets, we considered this evidence that the RFI classes differed in mean layer.

The association of RFI status and metabolic network structure was also assessed using several network statistics: “carbon sum”, betweenness-centrality, node degree, and clustering coefficient. We define carbon sum to be the sum of the number of carbon atoms present in a reaction’s suite of products and reactants. Betweenness-centrality is defined as the number of shortest paths between all pairs of nodes in a network that pass through the specified node (Freeman et al., 1991). The clustering coefficient measures the proportion of a node’s neighbors that are also each other’s neighbors (Watts and Strogatz, 1998), and node degree is simply the total edge count for a node (Watts and Strogatz, 1998). For each node and RFI class, the four statistics were weighted by the normalized number of mapped reads for that class and node, and the mean of this weighted value computed over all nodes. The weighted differences in these statistics between the two RFI classes were then compared to weighted differences from the randomized networks just described. If 5% or fewer of the randomized networks displayed a difference in the statistic in question as large as or larger than that observed in the actual data, we concluded that the RFI classes differed for that statistic (Table 2).

Principal component analysis

Principal component analysis (PCA) was performed with the R princomp() (Team, 2008) function to explore the patterns of variation seen using the normalized read or OTU counts from the sixteen animals.

Shannon entropy of OTU and node distributions

The mean Shannon entropies (Shannon, 1948) of the node and OTU distributions were calculated for the high and low RFI animals. To assess the differences between the RFI classes in these entropy measures, the differences between classes were again compared to the differences seen for networks where reads were randomly assigned to animals without reference to their RFI status.

RESULTS

Metagenomic sequencing and community metabolic network interference.

Using an Illumina HiSeq 2500, we shotgun sequenced microbial DNA from rumen fluids sampled from eight animals judged to be efficient in their feed usage (low RFI) and eight judged to be inefficient (high RFI). After quality filtering, we obtained approximately 1.75 billion 100 base pair, paired-end, reads (Table 1).

From the MetaCyc database (Caspi et al., 2014), we inferred pair of reaction-centric metabolic networks that model the flow of metabolites between the microbes and the host. Interface metabolites, defined as compounds produced by the microbes and potentially absorbed by the host as nutrients, define the interface between these two networks (*Materials and Methods*). We also employed three different stringencies for removing currency metabolites like ATP: from high to low stringency, these networks are

defined as N_{25} , N_{50} , and N_{100} , respectively. Starting from MetaCyc's database of enzyme sequences, we used sequence similarity searches against the metagenomic read data to infer which enzymes (and hence metabolic reactions) were present in each rumen sample. For each such sample, metagenomic reads were mapped to the microbial metabolic network, which was connected to the inferred sheep metabolic network by one of three pre-defined sets of interface metabolites, denoted as VFA, VFA + AA, and ALL (*Materials and Methods*). Network N_{50} with the inference set VFA + AA is illustrated in Fig.1.

Nodes do not differ by RFI in their proportion of mapped reads.

We explored whether differences in residual feed intake were associated with differences in enzyme abundances between the two efficiency groups. In Fig 1., the upper half of each microbial node (right-hand network) is colored by relative read abundance in the low RFI animals and the lower half by abundance in high RFI animals. Visually, there are very few nodes with any apparent difference in relative enzyme-coding gene abundance between the two groups. To formally test whether such differences were present, we applied a Wilcoxon test. There were no metabolic reactions with differing read abundance between the low and high RFI animals ($P > 0.05$). Differences in the microbiomes of low and high RFI animals are hence not apparent at the level of single reactions.

Low RFI animals employ reactions nearer to the host network and involving shorter carbon chains.

We next tested four network-level statistics for differences between the two RFI groups. The average number of carbon atoms involved in a reaction was significantly

greater in the high RFI animals compared to the low RFI animals (Table 2; $P < 0.05$: see also Supplemental Fig. 1). On the other hand, the two diet classes did not differ on average in their between-centrality or clustering coefficients (Table 2; $P > 0.05$), with a weak difference in node degree seen only for network N₁₀₀.

However, while little difference was seen between the low and high RFI animals for individual nodes and for most network statistics, the gross structure of the microbial network does differ between the groups. In particular, the reads from the microbes found in the low RFI animals map to reactions that are significantly closer to the host metabolic network than do the reads from the high RFI (inefficient) animals (Table 3 and Supplemental Fig. 1). This result is in accord with our hypothesis that low RFI animals have community metabolic networks that are able to interface more easily with that of the host animal because they use reactions that are nearer to it in network space.

Principal component analysis

We used PCA to visualize the differences between animals in the pattern of OTUs and nodes mapped in their microbiomes. For the taxonomic analysis, it generally appears that low RFI animals cluster together while the high RFI animals separate into two groups (Fig. 2A). However, there was no clear difference between the low and high RFI animals when the normalized read count per metabolic network node was used as the underlying data for the PCA (Fig. 2B).

Shannon Entropy

We compared the Shannon entropies (Shannon, 1948) of the distributions of reads onto OTUs and onto reactions for the low and for the high RFI animals. For the metabolic network nodes, the efficient animals showed significantly greater mean

entropy that the inefficient ones ($P < 0.001$; Fig. 3B). For the taxonomic data, no significant difference in mean entropy between the two classes was seen ($P = 0.073$; Fig. 3B). For both the taxonomic and metabolic network measures, however, the mean entropy of the samples was smaller than that seen when reads were randomized within an RFI group ($P < 0.001$; Fig. 3A and C, see also Supplemental Fig. 2).

DISCUSSION

Understanding the degree to which feed efficiency differences are causally associated with the observed microbial differences between efficient and inefficient animals (Guan et al., 2008; Hernandez-Sanabria et al., 2010; Rius et al., 2012; McCann et al., 2014) is obviously of considerable importance. Unfortunately, making such links can be difficult (Hernandez-Sanabria et al., 2012). For instance, it appears efficient cattle produce less methane than do inefficient ones (Hegarty et al., 2007). As methane production diverts the flow of carbon atoms toward methane and away from other reduced organic compounds that ruminants absorb for nutrition, a simple hypothesis would be that inefficient animals simply possess more methanogens (Pinares-Patiño et al., 2013). However, the actual situation is considerably more complex and not mechanistically understood (Zhou et al., 2009). Shi et al., found that while sheep differing in methane production showed no statistical differences in the number genes from methane-production pathways present between animals with low and high methane production, the expression of those genes was higher in individuals with high methane production (Shi et al., 2014). We suggest that the complexity of these results suggest the need for systems-level studies of the ruminal ecosystem.

A second reason such approaches are needed is the limited association between taxonomic and functional variation in the microbiome (Moore et al., 2009; Doolittle and Booth, 2017). Indeed, it is possible that this disconnect between taxonomy and function may contribute to the difficulties seen when trying to use the taxonomy of the ruminal microbiome to predict feed efficiency (Hernandez-Sanabria et al., 2010; Rius et al., 2012). Thus, in our own work, we have shown that functional approaches employing metabolic networks uncover features of microbial ecosystems that are not apparent from taxonomic analyses (Taxis et al., 2015). We have also used this metabolic network approach to show that a diet shift in a group of sheep initially fed a forage-like diet induced complex changes in the community metabolic networks of their rumens. In particular, when switched to a grain-based diet, the set of microbial reactions present shifted to favor those that are closer in network space to the reactions of the host sheep (Wolff et al., 2017). Likewise, we found differences in almost every aspect of the networks induced by the two diets, from the abundances of reads mapped to individual nodes to the overall network structure. These results induced us to conduct the work just described, exploring whether network analyses of the microbiome could uncover new insights as to the linkage between the microbiome and feed efficiency.

Given the importance of feed efficiency, it would be very desirable to find equally dramatic differences between high and low RFI animals. Unfortunately, such is not the case: when we considered standard comparisons of node abundances and network statistics, we generally found no significant difference between the RFI classes. Indeed, for no single metabolic network node was there a statistically significant difference between the RFI classes after multiple test corrections. Likewise, our PCA analysis did

not suggest that RFI status was a major differentiating factor between the animals (Fig 2B).

However, when we used a set of analyses that were more explicitly based on hypotheses regarding *why* inefficient and efficient animals might differ in their ruminal microbiomes, the picture was strikingly different. Ruminants obtain nutritional energy from a relatively small number of compounds they absorb from their symbiotic microbes. When we structured the metabolic network to account for this fact, we saw that reads from the low RFI animals mapped nearer (on average) to the host metabolic network than did those from high RFI animals. When coupled to the fact that low RFI animals also on average used reactions with shorter carbon chains than did high RFI animals, a picture emerges of RFI being partly a function of how well the microbial ecosystem is structured to allow the host animal to extract metabolites from it. This proposal is further supported by the analyses of the diversity (Shannon entropy) of the metabolic networks found in the two groups where we found that low RFI ewes had greater mean node entropy than did high RFI ones. This observation is actually the converse of the pattern observed in obese humans, whereby the individuals who gain weight more easily have reduced taxonomic diversity (Turnbaugh et al., 2009). Several studies have also shown a decrease in microbial diversity have been linked to digestive diseases like Crohn's disease (Sha et al., 2013) and irritable bowel syndrome (Carroll et al., 2012) in humans. The increased application of antibiotics in animal production has resulted in decreased resilience in the gut microbiota due to reduced biodiversity (Mosca et al., 2016). However, microbiomes with reduced diversity are associated with poor health more generally, suggesting low diversity may indeed be the pathological case (Cotillard et al., 2013). These observations

are contradicted to a degree, in the specific case of cattle fed grain-based diets, by the observation that low taxonomic diversity is associated with higher feed efficiency (Shabat et al., 2016; Li and Guan, 2017). However, we suspect that this example is the exception rather than the rule, both because the diet in question is much less complex than the evolved bovine diet and because it focuses on taxonomy alone, whereas our results in Fig. 3 speak to functional diversity. We suggest that studies of the microbiome would benefit from critically considering the ecological research suggesting that complex, diverse ecosystems are more efficient and stable.(Cardinale et al., 2002; Cotillard et al., 2013)

It is important to note that nothing in our data speak to causality {itself a challenging concept in complex systems like these (Wagner, 1999)} with respect to feed efficiency. Host genetics could very well drive the microbiome differences we have detected, as could environmental differences or developmental trajectories (Li and Guan, 2017). Likewise, the structure of the microbiome will interact with host physiology in complex ways (such as changes in ruminal pH), and the causes and effects in these relationships may not be yet understood. The importance of studying the microbiome lies in improving our mechanistic understanding of the processes of ruminal fermentation and feed efficiency, regardless of the ultimate origins of these differences.

A few other caveats are also in order. Our analyses of the microbial metabolic networks were limited to known and identifiable enzymes, and it is possible the story would grow in complexity with a more complete view of microbial metabolism. Likewise, the linkage between ecosystem diversity and efficiency is still debated (Fridley, 2001). However, we note that several factors increase our confidence in our

results. In all three cases, the significant differences between the two RFI classes were in the direction we predicted *a priori*: namely efficient animals should use shorter carbon chain metabolites that are near to the host network and should show increased diversity in their networks. Likewise, the mean entropies of the actual samples are lower than that seen when reads were randomized within RFI class, as would be expected if there are further rules structuring these microbial communities such as patterns of required taxa co-occurrence (Carr and Borenstein, 2014; Wolff et al., 2017).

Synthesizing from our observations, we propose that it may be unhelpful to view inefficiency in feed usage as a unitary trait. Figure 3 and the two supplemental figures suggest that inefficient animals show a good deal of animal-to-animal variation. From a systems perspective, this fact may not be surprising: constructing an ecosystem that allows efficient growth is likely subject to errors in many dimensions: perhaps some inefficient animals have ecosystems that are both non-diverse and far from the host network (e.g., animals 15 and 16 in Supplemental Fig. 2), while others are well-positioned but employ longer carbon chains than would be optimal (animal 7 in Supplemental Fig. 1). To gain a better understanding of the gut microbiome, we argue that a systems perspective is invaluable (Consortium, 2012; Franzosa et al., 2014), both as a means of generating hypotheses and as a starting point for developing interventions. Such perspectives will also allow for the development of predictive models providing mechanistic insights into the sources of inefficiency in feed use (Karlsson et al., 2011).

Table 1: RFI animal read statistics

Class	Animal ID	Total Reads After Quality Filtering ^a	Number of Reads Matching MetaCyc ^b	Number of Reads Matching OTU ^c
Low RFI	1	119833042	310921 (0.26%)	1192
	2	115699275	307898 (0.27%)	1252
	3	17712966	253800 (1.43%)	849
	4	16336193	183859 (1.13%)	645
	9	136044846	439771 (0.32%)	1562
	10	129561726	323016 (0.25%)	1033
	11	126647671	242727 (0.19%)	828
	12	83677521	220034 (0.26%)	707
High RFI	5	119789299	320006 (0.27%)	1050
	6	126030561	316015 (0.25%)	1133
	7	131776487	375399 (0.28%)	1713
	8	113135369	336808 (0.30%)	1149
	13	202547203	706605 (0.35%)	2528
	14	100366944	197717 (0.20%)	686
	15	101143440	528514 (0.52%)	889
	16	109260821	313093 (0.29%)	1101

a: Total reads sequenced after quality filtering

b: Number and percent of reads that were mapped to nodes from MetaCyc

c: Number of specific OTUs identified in the sequences

Table 2: RFI network structure

Statistic ^a	Network ^b	Low ^c	High ^c	Real Difference (Low-High) ^d	Mean Random Difference ^e	<i>P</i> ^f
Carbon Sum	N ₂₅	33.07	33.15	0.08	0.03	0.021
	N ₅₀	33.07	33.15	0.08	0.03	0.022
	N ₁₀₀	33.07	33.15	0.08	0.03	0.020
Betweenness Centrality	N ₂₅	8,938.85	8,970.73	31.88	19.22	0.196
	N ₅₀	18,090.94	18,093.78	2.84	34.03	0.954
	N ₁₀₀	18,136.57	18,115.52	21.05	31.32	0.591
Degree	N ₂₅	3.52	3.53	0.01	0.004	0.187
	N ₅₀	9.17	9.19	0.02	0.01	0.119
	N ₁₀₀	12.08	12.14	0.06	0.01	< 0.001
Clustering Coefficient	N ₂₅	0.88	0.88	< 0.001	< 0.001	0.075
	N ₅₀	0.86	0.86	< 0.001	< 0.001	0.622
	N ₁₀₀	0.85	0.85	< 0.001	< 0.001	0.919

a: Network statistic compared between the Low and High RFI animals. Carbon sum: Total number of carbon atoms in a reaction. Betweenness-centrality: Number of shortest paths crossing a node. Degree: Number of edges of a node. Clustering coefficient: Number of fully connected triangles a node participates in relative to its total number of edges.

b: Network defined by currency metabolite removal (*Materials and Methods*).

c: Mean value for selected network statistic mapped for Low or High RFI animals.

d: Difference between the mean statistic for Low and High RFI animals.

e: Mean random difference between Low and High RFI groups when reads are randomly assigned.

f: *P*-value for the test of the null hypothesis that there is no difference in the mean statistic. *P* was computed by randomly assigning reads to RFI groups and comparing the difference in the two mean values for 1,000 such randomizations to the actual data (see *Materials and Methods*). *P*-values less than 0.05 were judged significant and are shown in bold.

Table 3: RFI group and network position

Group ^a	Currency Cutoff ^b	Mean layer: Low RFI ^c	Mean layer: High RFI ^c	Real Difference: mean layers ^d	Max Random Difference ^e	<i>P</i> ^f
VFA	N ₂₅	1.976	1.974	0.003	0.009	0.304
	N ₅₀	1.869	1.882	0.013	0.008	< 0.001
	N ₁₀₀	1.772	1.787	0.016	0.006	< 0.001
VFA + AA	N ₂₅	0.798	0.805	0.007	0.004	< 0.001
	N ₅₀	0.863	0.872	0.009	0.004	< 0.001
	N ₁₀₀	0.899	0.910	0.011	0.004	< 0.001
ALL	N ₂₅	0.504	0.507	0.003	0.003	< 0.001
	N ₅₀	0.456	0.457	0.001	0.002	0.313
	N ₁₀₀	0.500	0.502	0.002	0.002	< 0.001

a: Interface metabolite set

b: Network (see Table 2)

c: Mean layer number for reads mapped from Low or High RFI animals

d: Difference between the mean layer for Low and High RFI

e: Maximum random difference between the mean layer for Low and High RFI

f: *P*-value for the test of the null hypothesis that there is no difference in the mean layer of the two RFI groups. Reads were randomly assigned to RFI status and mean layers were computed 1,000 times and the difference in these random networks compared to the real network (*Materials and Methods*). *P*-values less than 0.05 were judged significant and are shown in bold.

Fig. 1: The interaction between the host (purple, left) and microbial metabolic networks (right) for efficient and inefficient sheep. Each node (circle) represents an enzyme-catalyzed reaction: edges (lines) connect pairs of reactions that share metabolites. The colored lines in the center are interface metabolites that are potentially transferred from

the microbes to the host animal and hence connect the two metabolic networks (*Materials and Methods*). For the illustrated network, these metabolites consisted of the 20 universal amino acids and three volatile fatty acids that are the primary nutrients of the host (VFA + AA: *Materials and Methods*). The ten interface metabolites with microbial reactions with the largest number of mapped reads are shown with individual colors (upper left). The upper half of each microbial node is colored based on the relative abundance reads mapping to that node for the low RFI animals while the lower half of the node similarly colored for the high RFI animals: see the scale diagram at the lower left. There were no metabolic reactions that showed significantly different abundance between the two RFI groups ($P > 0.05$, Wilcoxon Test).

Fig. 2: First two principal components (x and y -axes, respectively) from PCAs of the OTUs: **A)** and normalized node abundances (**B**). The low RFI animals are shown in green while the high RFI animals are in blue.

Fig. 3: Shannon entropies of the distribution of microbial reads onto OTUs (x -axis) and onto metabolic network nodes (y -axis) for the low (green) and high (blue) RFI animals.

A) Distribution of randomized node-wise entropies for low (green) and high (blue) RFI

animals (*Materials and Methods*). **B)** Shannon entropies for OTUs and metabolic network nodes for each sampled animal. The mean Shannon entropies for the two classes are marked by the open squares. The mean node-wise entropy for the low RFI animals is significantly greater than that for the high RFI animals ($P < 0.001$). There is not a significant difference between the mean taxonomic entropy of the two classes ($P = 0.073$). In all cases, the mean entropy for the two classes is smaller than the mean entropies seen when reads were randomized among individuals within a class (histograms in **A** and **C**; $P < 0.001$, *Materials and Methods*). **C)** As for **A** but with taxonomic values.

REFERENCES

- Bekaert, M., and G. C. Conant. 2011. Copy number alterations among mammalian enzymes cluster in the metabolic network. *Mol Biol Evol* 28(2):1111-1121. doi: 10.1093/molbev/msq296
- Bekaert, M., and G. C. Conant. 2014. Gene duplication and phenotypic changes in the evolution of mammalian metabolic networks. *PLoS One* 9(1):e87115. doi: 10.1371/journal.pone.0087115
- Benjamini, Y. a. H., Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. No. 57. p 289-300, *Journal of the Royal Statistical Society. Series B (Methodological)*.
- Bergman, E. N. 1990. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol Rev* 70(2):567-590.
- Cammack, K. M., K. A. Leymaster, T. G. Jenkins, and M. K. Nielsen. 2005. Estimates of genetic parameters for feed intake, feeding behavior, and daily gain in composite ram lambs. *J Anim Sci* 83(4):777-785.
- Cardinale, B. J., M. A. Palmer, and S. L. Collins. 2002. Species diversity enhances ecosystem functioning through interspecific facilitation. *Nature* 415(6870):426-429. doi: 10.1038/415426a
- Carr, R., and E. Borenstein. 2014. Comparative analysis of functional metagenomic annotation and the mappability of short reads. *PLoS One* 9(8):e105776. doi: 10.1371/journal.pone.0105776
- Carroll, I. M., T. Ringel-Kulka, J. P. Siddle, and Y. Ringel. 2012. Alterations in composition and diversity of the intestinal microbiota in patients with diarrhea-

- predominant irritable bowel syndrome. *Neurogastroenterol Motil* 24(6):521-530, e248. doi: 10.1111/j.1365-2982.2012.01891.x
- Caspi, R., T. Altman, R. Billington, K. Dreher, H. Foerster, C. A. Fulcher, T. A. Holland, I. M. Keseler, A. Kothari, A. Kubo, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, D. Weerasinghe, P. Zhang, and P. D. Karp. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* 42(Database issue):D459-471. doi: 10.1093/nar/gkt1103
- Cole, J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity, and J. M. Tiedje. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 37(Database issue):D141-145. doi: 10.1093/nar/gkn879
- Consortium, H. M. P. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486(7402):207-214. doi: 10.1038/nature11234
- Cotillard, A., S. P. Kennedy, L. C. Kong, E. Prifti, N. Pons, E. Le Chatelier, M. Almeida, B. Quinquis, F. Levenez, N. Galleron, S. Gougis, S. Rizkalla, J. M. Batto, P. Renault, J. Doré, J. D. Zucker, K. Clément, S. D. Ehrlich, and A. M. consortium. 2013. Dietary intervention impact on gut microbial gene richness. *Nature* 500(7464):585-588. doi: 10.1038/nature12480
- Dijkstra, E. W. 1959. A note on two problems in connexion with graphs. No. 1. p 269-271, *Numerische Mathematik*

- Doolittle, W., and A. Booth. 2017. It's the song, not the singer: an exploration of holobiosis and evolutionary theory. *Biology and Philosophy* 32(1):5-24.
- Duarte, N. C., S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas, and B. Palsson. 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A* 104(6):1777-1782. doi: 10.1073/pnas.0610772104
- Döring, A., D. Weese, T. Rausch, and K. Reinert. 2008. SeqAn an efficient, generic C++ library for sequence analysis. *BMC Bioinformatics* 9:11. doi: 10.1186/1471-2105-9-11
- Ellison, M., G. Conant, W. Lamberson, R. Cockrum, K. Austin, D. Rule, and K. Cammack. 2017. Diet and feed efficiency status affect rumen microbial profiles of sheep. *Small Ruminant Research*
- Ellison, M. J., G. C. Conant, R. R. Cockrum, K. J. Austin, H. Truong, M. Becchi, W. R. Lamberson, and K. M. Cammack. 2014. Diet alters both the structure and taxonomy of the ovine gut microbial ecosystem. *DNA Res* 21(2):115-125. doi: 10.1093/dnares/dst044
- Ewing, B., L. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8(3):175-185.
- Flicek, P., M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C. G. Girón, L. Gordon, T. Hourlier, S. Hunt, N. Johnson, T. Juettemann, A. K. Kähäri, S. Keenan, E. Kulesha, F. J. Martin, T. Maurel, W. M. McLaren, D. N. Murphy, R. Nag, B. Overduin, M. Pignatelli, B. Pritchard, E. Pritchard, H. S. Riat, M. Ruffier, D. Sheppard, K.

- Taylor, A. Thormann, S. J. Trevanion, A. Vullo, S. P. Wilder, M. Wilson, A. Zadissa, B. L. Aken, E. Birney, F. Cunningham, J. Harrow, J. Herrero, T. J. Hubbard, R. Kinsella, M. Muffato, A. Parker, G. Spudich, A. Yates, D. R. Zerbino, and S. M. Searle. 2014. Ensembl 2014. *Nucleic Acids Res* 42(Database issue):D749-755. doi: 10.1093/nar/gkt1196
- Franzosa, E. A., X. C. Morgan, N. Segata, L. Waldron, J. Reyes, A. M. Earl, G. Giannoukos, M. R. Boylan, D. Ciulla, D. Gevers, J. Izard, W. S. Garrett, A. T. Chan, and C. Huttenhower. 2014. Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A* 111(22):E2329-2338. doi: 10.1073/pnas.1319284111
- Freeman, L. C., S. P. Borgatti, and D. R. White. 1991. Centrality in valued graphs: A measure of betweenness based on network flow No. 13 (2). p 141-154, *Social Networks*.
- Fridley, J. D. 2001. The influence of species diversity on ecosystem productivity: how, where, and why? *Oikos* 93:514-526.
- Guan, L. L., J. D. Nkrumah, J. A. Basarab, and S. S. Moore. 2008. Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. *FEMS Microbiol Lett* 288(1):85-91. doi: FML1343 [pii]
- 10.1111/j.1574-6968.2008.01343.x
- Hegarty, R. S., J. P. Goopy, R. M. Herd, and B. McCorkell. 2007. Cattle selected for lower residual feed intake have reduced daily methane production. *J Anim Sci* 85(6):1479-1486. doi: 10.2527/jas.2006-236

- Hernandez-Sanabria, E., L. A. Goonewardene, Z. Wang, O. N. Durunna, S. S. Moore, and L. L. Guan. 2012. Impact of feed efficiency and diet on adaptive variations in the bacterial community in the rumen fluid of cattle. *Appl Environ Microbiol* 78(4):1203-1214. doi: 10.1128/AEM.05114-11
- Hernandez-Sanabria, E., L. L. Guan, L. A. Goonewardene, M. Li, D. F. Mujibi, P. Stothard, S. S. Moore, and M. C. Leon-Quintero. 2010. Correlation of particular bacterial PCR-denaturing gradient gel electrophoresis patterns with bovine ruminal fermentation parameters and feed efficiency traits. *Appl Environ Microbiol* 76(19):6338-6350. doi: 10.1128/AEM.01052-10
- Herrero, M., P. Havlík, H. Valin, A. Notenbaert, M. C. Rufino, P. K. Thornton, M. Blümmel, F. Weiss, D. Grace, and M. Obersteiner. 2013. Biomass use, production, feed efficiencies, and greenhouse gas emissions from global livestock systems. *Proc Natl Acad Sci U S A* 110(52):20888-20893. doi: 10.1073/pnas.1308149110
- Jiang, Y., M. Xie, W. Chen, R. Talbot, J. F. Maddox, T. Faraut, C. Wu, D. M. Muzny, Y. Li, W. Zhang, J. A. Stanton, R. Brauning, W. C. Barris, T. Hourlier, B. L. Aken, S. M. Searle, D. L. Adelson, C. Bian, G. R. Cam, Y. Chen, S. Cheng, U. DeSilva, K. Dixen, Y. Dong, G. Fan, I. R. Franklin, S. Fu, P. Fuentes-Utrilla, R. Guan, M. A. Highland, M. E. Holder, G. Huang, A. B. Ingham, S. N. Jhangiani, D. Kalra, C. L. Kovar, S. L. Lee, W. Liu, X. Liu, C. Lu, T. Lv, T. Mathew, S. McWilliam, M. Menzies, S. Pan, D. Robelin, B. Servin, D. Townley, W. Wang, B. Wei, S. N. White, X. Yang, C. Ye, Y. Yue, P. Zeng, Q. Zhou, J. B. Hansen, K. Kristiansen, R. A. Gibbs, P. Flicek, C. C. Warkup, H. E. Jones, V. H. Oddy, F. W. Nicholas, J.

- C. McEwan, J. W. Kijas, J. Wang, K. C. Worley, A. L. Archibald, N. Cockett, X. Xu, and B. P. Dalrymple. 2014. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344(6188):1168-1173. doi: 10.1126/science.1252806
- Karlsson, F. H., I. Nookaew, D. Petranovic, and J. Nielsen. 2011. Prospects for systems biology and modeling of the gut microbiome. *Trends Biotechnol* 29(6):251-258. doi: 10.1016/j.tibtech.2011.01.009
- Lacroix, V., L. Cottret, P. Thébault, and M. F. Sagot. 2008. An introduction to metabolic networks and their structural analysis. *IEEE/ACM Trans Comput Biol Bioinform* 5(4):594-617. doi: 10.1109/TCBB.2008.79
- Li, F., and L. L. Guan. 2017. Metatranscriptomic Profiling Reveals Linkages between the Active Rumen Microbiome and Feed Efficiency in Beef Cattle. *Appl Environ Microbiol* 83(9)doi: 10.1128/AEM.00061-17
- Masson, M. J., and A. T. Phillipson. 1951. The absorption of acetate, propionate and butyrate from the rumen of sheep. *J Physiol* 113(2-3):189-206.
- McCann, J. C., L. M. Wiley, T. D. Forbes, F. M. Rouquette, and L. O. Tedeschi. 2014. Relationship between the rumen microbiome and residual feed intake-efficiency of Brahman bulls stocked on bermudagrass pastures. *PLoS One* 9(3):e91864. doi: 10.1371/journal.pone.0091864
- Moore, S. S., F. D. Mujibi, and E. L. Sherman. 2009. Molecular basis for residual feed intake in beef cattle. *J Anim Sci* 87(14 Suppl):E41-47. doi: 10.2527/jas.2008-1418

- Mosca, A., M. Leclerc, and J. P. Hugot. 2016. Gut Microbiota Diversity and Human Diseases: Should We Reintroduce Key Predators in Our Ecosystem? *Front Microbiol* 7:455. doi: 10.3389/fmicb.2016.00455
- Pinares-Patiño, C. S., S. M. Hickey, E. A. Young, K. G. Dodds, S. MacLean, G. Molano, E. Sandoval, H. Kjestrup, R. Harland, C. Hunt, N. K. Pickering, and J. C. McEwan. 2013. Heritability estimates of methane emissions from sheep. *Animal* 7 Suppl 2:316-321. doi: 10.1017/S1751731113000864
- Pérez-Bercoff, Å., A. McLysaght, and G. C. Conant. 2011. Patterns of indirect protein interactions suggest a spatial organization to metabolism. *Mol Biosyst* 7(11):3056-3064. doi: 10.1039/c1mb05168g
- Rius, A. G., S. Kittelmann, K. A. Macdonald, G. C. Waghorn, P. H. Janssen, and E. Sikkema. 2012. Nitrogen metabolism and rumen microbial enumeration in lactating cows with divergent residual feed intake fed high-digestibility pasture. *J Dairy Sci* 95(9):5024-5034. doi: 10.3168/jds.2012-5392
- Sha, S., B. Xu, X. Wang, Y. Zhang, H. Wang, X. Kong, H. Zhu, and K. Wu. 2013. The biodiversity and composition of the dominant fecal microbiota in patients with inflammatory bowel disease. *Diagn Microbiol Infect Dis* 75(3):245-251. doi: 10.1016/j.diagmicrobio.2012.11.022
- Shabat, S. K., G. Sasson, A. Doron-Faigenboim, T. Durman, S. Yaacoby, M. E. Berg Miller, B. A. White, N. Shterzer, and I. Mizrahi. 2016. Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *ISME J* 10(12):2958-2972. doi: 10.1038/ismej.2016.62

- Shannon, C. 1948. A mathematical theory of communication. AT&T Tech J.:379-423 and 623-656.
- Shi, W., C. D. Moon, S. C. Leahy, D. Kang, J. Froula, S. Kittelmann, C. Fan, S. Deutsch, D. Gagic, H. Seedorf, W. J. Kelly, R. Atua, C. Sang, P. Soni, D. Li, C. S. Pinares-Patiño, J. C. McEwan, P. H. Janssen, F. Chen, A. Visel, Z. Wang, G. T. Attwood, and E. M. Rubin. 2014. Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res* 24(9):1517-1525. doi: 10.1101/gr.168245.113
- Smith, T. F., and M. S. Waterman. 1981. Identification of common molecular subsequences. *J Mol Biol* 147(1):195-197.
- Stevens, C. E., and I. D. Hume. 1998. Contributions of microbes in vertebrate gastrointestinal tract to production and conservation of nutrients. *Physiol Rev* 78(2):393-427.
- Taxis, T. M., S. Wolff, S. J. Gregg, N. O. Minton, C. Zhang, J. Dai, R. D. Schnabel, J. F. Taylor, M. S. Kerley, J. C. Pires, W. R. Lamberson, and G. C. Conant. 2015. The players may change but the game remains: network analyses of ruminal microbiomes suggest taxonomic differences mask functional similarity. *Nucleic Acids Res* 43(20):9600-9612. doi: 10.1093/nar/gkv973
- Team, R. D. C. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
- Turnbaugh, P. J., M. Hamady, T. Yatsunenko, B. L. Cantarel, A. Duncan, R. E. Ley, M. L. Sogin, W. J. Jones, B. A. Roe, J. P. Affourtit, M. Egholm, B. Henrissat, A. C.

- Heath, R. Knight, and J. I. Gordon. 2009. A core gut microbiome in obese and lean twins. *Nature* 457(7228):480-484. doi: 10.1038/nature07540
- Wagner, A. 1999. Causality in Complex Systems. *Biology and Philosophy* (14):83-101.
- Watts, D. J., and S. H. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature* 393(6684):440-442. doi: 10.1038/30918
- Wolff, S. M., M. J. Ellison, Y. Hao, R. R. Cockrum, K. J. Austin, M. Baraboo, K. Burch, H. J. Lee, T. Maurer, R. Patil, A. Ravelo, T. M. Taxis, H. Truong, W. R. Lamberson, K. M. Cammack, and G. C. Conant. 2017. Diet shifts provoke complex and variable changes in the metabolic networks of the ruminal microbiome. *Microbiome* 5(1):60. doi: 10.1186/s40168-017-0274-6
- Zhou, M., E. Hernandez-Sanabria, and L. L. Guan. 2009. Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies. *Appl Environ Microbiol* 75(20):6524-6533. doi: AEM.02815-08 [pii]
- 10.1128/AEM.02815-08

A Metagenomic Analysis of the Effect of Antibiotic Feed Additives on the Ovine Rumen Metabolism

Rocky D. Patil,^a Melinda J. Ellison,^b Kathy J. Austin,^c William R. Lamberson,^a Kristi M. Cammack,^d and Gavin C. Conant^{a, e, f, g}

^aDivision of Animal Sciences, University of Missouri-Columbia, Columbia, MO, 65202, U.S.A. ^bAnimal and Veterinary Science, Nancy M. Cummings Research, Extension, and Education Center, University of Idaho, Carmen, ID, 83462, U.S.A.

^cDepartment of Animal Science, University of Wyoming, Laramie, WY, 82071, U.S.A.

^dDepartment of Animal Science, South Dakota State University, Brookings, SD, 57007, U.S.A.

^eProgram in Genetics, North Carolina State University, Raleigh, NC, 27695, U.S.A.

^fBioinformatics Research Center, North Carolina State University, Raleigh, NC, 27695, U.S.A.

^gDepartment of Biological Sciences, North Carolina State University, Raleigh, NC, 27695, U.S.A.

This chapter has been submitted for publication.

ABSTRACT

Antibiotics are used as feed additives for domesticated ruminants as they promote growth and improve feed efficiency while simultaneously reducing disease. At the same time, ruminants maintain a symbiotic relationship with the microorganisms in their rumen whereby those microbes degrade plant materials into simpler metabolites used by the host. The effects of antibiotics in feed on this symbiotic relationship are still not completely understood. In this study, we used shotgun metagenomics to compare the microbial communities from 16 sheep fed an antibiotic-supplemented diet to those of 16 sheep fed an unsupplemented diet. Metagenomic reads from each sheep were mapped to a microbial metabolic network which was linked to the host sheep metabolic network by sets of metabolites that are assumed to be absorbed by that host. Several differences in the global structure of this network were evident when comparing sheep that were versus were not fed antibiotics. There were 546 microbial enzymes that were significantly different in their relative abundance between the two groups. On a broader scale, the animals fed antibiotics showed a set of microbial enzymes that were closer in network space to the host metabolism than were the enzymes of the microbes of animals not fed antibiotics. There were also differences in the microbial taxa present in the two groups of animals, with members of the genus *Prevotella* representing approximately 68% of microbial individuals in sheep not fed antibiotics, but only 46% in antibiotic fed sheep. We found no difference in the metabolic diversity (Shannon entropies) of the two groups, but greater ($P = 0.01$) microbial taxonomic diversity in sheep fed antibiotics. Collectively, this research indicates that the addition of antibiotics to ruminant feed alters

both the taxonomic and functional structure of their ruminal microbial communities, even if the overall effects of these changes are not fully understood.

INTRODUCTION

The use of antimicrobial additives in ruminant feed has become a common agricultural practice as it improves feed efficiency and potential animal health for instance, antimicrobial feed can reduce the incidence of helminthiasis in sheep (Larsen et al., 2009). Recently, several studies have suggested that this use of antibiotics as a supplement has increased the prevalence of bacteria with antimicrobial resistance (AMR) genes in the environment (Cameron & McAllister, 2016). The appearance of such resistance raises the concern that the AMR genes could be horizontally transferred to human pathogens through the ingestion of contaminated food (Muloi et al., 2018). At the same time, it is also relevant to ask how antibiotics effect the domesticated animals themselves. Ruminants are unusually dependent on their gut microbiome, as it is these organisms that break down nutritionally inaccessible plant compounds such as lignin and cellulose into metabolites that host animal can use. The changes in the ruminal microbiome brought about by changes in diet or other alterations can influence the host animals' propensity for diseases and its general health (Thomas et al., 2017). Previous studies have indicated that dramatic changes in the composition of the gut microbiota can be brought about through changes in diet (Tajima et al., 2001; Wolff et al., 2017). It has also been demonstrated that antibiotics alter the taxonomic structure of the rumen microbiome (Looft et al., 2012), but how antibiotics may alter the functional

characteristics of the microbiome and the effects of those changes on the host are still not understood.

We have developed a community metabolic network approach for analyzing vertebrate-gut associated microbiomes (Patil et al., 2018; Taxis et al., 2015; Wolff et al., 2017) that compares the relative abundance of known microbial enzymes across host animals differing in their feed or other functional characteristics. Using this approach, we compared the community metabolic networks of the ruminant microbiota for sheep fed an antibiotic-supplemented (ANT) and those fed an unsupplemented (NON) diet. We observed that antibiotics reduced the prevalence of microbes from the genus *Prevotella* in the gut. We also found that there were 546 microbial metabolic reactions that were significantly different in their abundance between the two groups. The overall community metabolic network structure also differed between the groups, with ANT fed animals having microbial reactions that were closer in network space to the host network than were the reactions of the microbes in the NON-group. While ANT animals had ruminal microbiomes with greater alpha taxonomic diversity than did NON-animals, the community metabolic networks for the two groups were comparable in their diversity.

MATERIALS AND METHODS

Sheep feeding trial and rumen sample collection

Rumen fluid was collected from the two groups of lambs. One was a previously described group of Targhee ewe lambs (n = 16; final BW, = 76.9 ± 1.6 kg) fed a forage-based diet (NON: non-antibiotic feed) (Patil et al., 2018). The second group of Targhee ewes (n = 16; final BW, = 151.5 ± 3.6 kg) were fed a forage-based diet supplemented

with chlortetracycline (ANT: antibiotic feed) for 70 days. Rumen fluids were sampled from each lamb's rumen by using a dosing syringe (400 mL) to syphon fluid through a tygon tube (length: 1 m, diameter: 1.5 cm) that reached the rumen via the esophagus. Rumen samples were immediately gathered after the feeding trial and stored on dry ice until the samples were frozen at -80°C for long-term storage.

DNA extraction and Illumina sequencing and quality filtering

Particulate matter from the ewe's rumen fluid was obtained by low speed centrifugation. Thawed rumen samples were combined with one ml of lysis buffer, silicon (0.1 g of 0.5 mm), and zirconia (0.3 g of 0.1 mm) beads to isolate DNA. The samples were homogenized by using a Mini-Beadbeater-8 for 3 minutes at the maximum speed followed by incubation at 70°C for 15 min and centrifugation at 4°C for 5 min. The remaining supernatant was transferred into new tubes followed by the addition of lysis buffer. This step was repeated for a second time, and the supernatants were pooled together. The QIAamp DNA Stool Mini Kit (Qiagen, Santa Clarita, CA) was administered to the supernatant to remove RNA and proteins and precipitate nucleic acids. The quality of the DNA was confirmed by using a NanoDrop spectrophotometer (NanoDrop, Wilmington, DE) having an A260/280 absorbance ration ≥ 1.8 .

Genomic libraries for the 32 rumen samples were constructed based on the guidelines set in Illumina's TruSeq DNA sample prep kit (Ellison et al., 2014). Genomic DNA was sheared by Diagenode BioRuptor to obtain approximately 350 base pairs. This fragmented DNA had 3' and 5' overhangs removed by an end pair reaction followed by the addition of adenosine nucleotide to the 3' ends by ligation of Illumina adapters. Each of the 32 purified libraries was quantified by the Qubit assay while the fragment size was

confirmed by Agilent BioAnalyzer High Sensitivity DNA assay. These genomic libraries were diluted into four samples per lane and sequenced with an Illumina HiSeq 2500 producing 100 base-pair, paired-end, reads. These paired-end reads were quality filtered by removing all bases in the read after the first three successive bases have a phred quality score of less than 15 (Ewing et al., 1998). From the reads that were quality trimmed in this manner, any read pairs where either member was less than 85 bases long or had an average phred quality score of less than 25 was excluded from further analyses.

OTU analysis

The filtered reads were compared to the 16S rDNA sequences from the Ribosomal Database Project (RDP) (Cole et al., 2009) which had 27,290 sequences after our filtering (Ellison et al., 2014). Operational taxonomic units (OTUs) were inferred from this database by first computing all possible pairwise Needleman-Wunsch sequence alignments (Needleman & Wunsch, 1970) among its sequences using a GPU-based pairwise alignment tool (Truong et al., 2014). OTUs were then defined as connected components in a graph of these sequences using single-linkage clustering with a cutoff of 97% percent identity for pairs of database sequences. Metagenomic read pairs were matched to this database sequences using Bowtie (Langmead et al., 2009): a pair was considered to have matched an OTU if the forward and reverse reads both aligned only to sequences from that OTU at $\geq 97\%$ sequence identity.

Creation of the metabolic network from MetaCyc

The MetaCyc database is a comprehensive collection of 2,666 metabolic pathways from 2,900 organisms, primarily bacteria and archaeans (Caspi et al., 2014). It comprises enzymes and their amino acid sequences annotated to their associated

biochemical reactions (as well as the substrates and products of those reactions). As previously described, the contents of the MetaCyc database were used to infer a metabolic network where nodes represent metabolic reactions and edges connect nodes sharing a metabolite (Taxis et al., 2015). Abundant metabolites, or “currency metabolites,” were excluded to avoid having an overly connected network. Three different thresholds were used in the removal of these currency metabolites, resulting in networks we refer to as N₂₅, N₅₀, and N₁₀₀. Specifically, these networks were constructed by removing metabolites occurring in more than 25, 50, or 100 reactions, respectively (Hudson & Conant, 2011; Pérez-Bercoff et al., 2011).

Ovine host and microbial metabolic network construction

The MetaCyc database was used to infer two metabolic networks; one was for the ovine host and the other for the ruminal microbes (Patil et al., 2018). The microbial network consisted of a set of 6,140 reaction nodes (Taxis et al., 2015; Wolff et al., 2017). Because there is no ovine metabolic network in the MetaCyc database, we generated this network by inferring the ovine orthologs of genes coding for the enzymes catalyzing the 2,863 reactions of the human and 1,404 reactions of the bovine metabolic networks in MetaCyc (Jiang et al., 2014). We assumed that the orthologous enzymes from sheep catalyzed the same reactions known in the other two mammals, allowing us to infer a set of ovine network nodes. We identified these (ovine/human or ovine/bovine) orthologs using our synteny-based orthology program and annotations from Ensembl release 75 (Flicek et al., 2014). In this manner, we identified 15,562 human-ovine and 15,479 bovine-ovine ortholog pairs (Bekaert & Conant, 2011). The resulting ovine metabolic network had 1,989 nodes: 1,834 inferred from the human network, 154 from the bovine

one, and one butyrate to butyryl-CoA reaction added because MetaCyc does not have a reaction allowing for the assimilation of butyrate into vertebrate metabolic networks.

Matching translated reads to the microbial metabolic network

In order to infer the MetaCyc reactions present in our metagenomic Illumina data, the quality-trimmed read pairs from each of the 32 samples were computationally translated into amino acid sequences in each of the six possible reading frames, yielding sets of open reading frames (ORFs). Only paired reads each having an ORF greater than 29 residues were retained and searched against the MetaCyc protein database, which was done using a tool based on SeqAn library (Döring et al., 2008). A match between a translated read and a database sequence was spawned if the two sequences shared two seven residue identical word matches. Each such match was then aligned using the Smith-Waterman local alignment algorithm (Smith & Waterman, 1981) and the match was retained if there was greater than 80% amino acid identity over the local alignment and if that alignment spanned 80% or more of the metagenomic ORF. All such reads that matched enzyme sequences from a consistent set of nodes were then assigned to those nodes (Wolff et al., 2017).

Network interface between the ovine and microbial metabolic networks

To model the exchange of metabolites between the host and microbial network, we posited three sets of compounds that are most likely to be absorbed in the ovine rumen. The first set consisted of the most common volatile fatty acids (VFAs) in the rumen: acetate, butyrate, and propionate (Bergman, 1990). The second set added to these VFAs the 20 universal amino acids (VFA + AA). Finally, the third interface set consists of VFA + AA with an additional 234 metabolites that are known to be absorbed in the

human gut and were taken from the global reconstruction of the human metabolic network (Duarte et al., 2007). All these compounds were matched to the MetaCyc database and the following compounds were removed: inorganic compounds, DNA/RNA, nucleotides, carbon dioxide, and NAD⁺ and NADP⁺. After these removals, the result was a set (“ALL”) with an additional 204 metabolites beyond the 23 in VFA+AA. The host and microbial networks were linked to each other on the basis of pairs of nodes that shared an interface metabolite.

Metabolic network layers & Node read density

We defined node’s distance from one subnetwork to be its distance to the closest node in the other subnetwork (Dijkstra, 1959). This definition breaks the microbial network into the layers shown in Figure 3. For example, nodes sharing an interface metabolite between the different networks would have a distance of 0. A two-sample Wilcoxon test was used to test for significant differences in the nodes’ read count between the non-antibiotic (NON) and antibiotic (ANT) diet. Only nodes with at least one mapped read pair were analyzed. The *P*-values from the Wilcoxon test had a 5% false-discovery rate correction applied (Benjamini, 1995).

Network structure analysis

Weighting each node by its number of mapped read pairs for each group, we calculated the mean network layers for the NON and ANT groups across all possible combinations of interface metabolites and currency metabolite comparisons (3 interface sets by 3 currency cutoffs = 9 comparisons). We used a randomization approach to assess whether this mean layer differed between the groups. To construct these random networks, we pooled reads for the ANT and NON-diets, and then selected at random

from that pool the same number of reads as originally found in the ANT diets to yield a “pseudo-ANT,” read set, with the remaining reads being defined as “pseudo-NON”. We then compared the difference in mean layer between ANT and NON in the real dataset against 1,000 such randomized datasets. Employing a two-sided statistical analysis, we concluded there was a significant difference in mean layer between the groups if the real difference in mean layer (positive or negative) was larger than all but 2.5% of the randomized datasets.

We also analyzed the effect of NON and ANT diets by comparing the values of four other network statistics between the CMNs of the two groups: “carbon sum”, betweenness-centrality, node-degree, and clustering coefficient. The carbon sum is the total number of carbon atoms among all the metabolites participating in a node’s reaction. A node’s betweenness-centrality gives the number of all the network’s shortest path that pass through that node (Freeman et al., 1991). The node degree is the node’s total edge count, and the clustering coefficient measures the degree to which connected nodes are also likely to be connected to each other’s neighbors (Watts & Strogatz, 1998). For each node, its value of a given network statistic was weighted by the normalized number of reads mapped to that node. We then calculated the difference in the mean of these weighted values between the ANT and the NON-animals. To assess statistical significance, we compared the observed differences in these weighted values to the differences seen for networks with randomly assigned reads, computed as above.

Shannon entropy of OTU and node distributions

The antibiotic and non-antibiotic diets were compared in their mean Shannon entropies (Shannon, 1948) both with respect to the metabolic network nodes and the OTU distributions:

$$E = - \sum_{i=1}^n P(r_i) \log_2(P(r_i))$$

In (1), the read count vector r_i represents the number of reads mapped to the i^{th} node or OTU and $P(r_i)$ represents the frequency reads mapped to that node/OTU relative to the total number of mapped reads.

Testing the significance of correlations between NON and ANT groups

A correlation coefficient analysis was conducted for NON and ANT body weights between OTUs, nodes, and mean layer using R to determine the Pearson's and Spearman's correlation coefficient (Team, 2008). Fisher's r to z transformations were performed to determine the statistical significance by comparing it to the z test statistic.

RESULTS

Metagenomic Sequencing of Sheep Rumen Fluid

We extracted microbial DNA from the rumen fluids of 32 sheep that were divided into two groups: one fed a diet consisting of antibiotic supplemented feed (ANT) and the other given standard feed (NON). Microbial DNA from rumen samples of these animals was shotgun sequenced on an Illumina HiSeq 2500, resulting in 2.78 billion 100 bp paired-end reads after quality-filtering (Table 1, Materials and Methods). We identified 376 bacterial and archaeal operational taxonomic units (OTUs) present in at least one of

these animals. Of these OTUs, 22 OTUs were statistically significantly differentially abundant between the ANT and NON-groups ($P < 0.05$, Supplemental Table 1).

Metabolic Network Interference

By using the MetaCyc database (Caspi et al., 2014), we were able to define a reaction-centric metabolic network that modeled the movement of metabolites between the microbes and host animals. Nodes in this community network represented enzyme-catalyzed reactions with one or more associated protein sequences in the MetaCyc database. We established two subnetworks: one for the microbes and one for the ovine host. This ovine network was inferred using the ovine orthologs of enzyme-coding genes from the human and bovine metabolic networks in MetaCyc (see Materials and Methods). We used three different stringencies for removing currency metabolites such as water and protons from these metabolic networks: we refer to the resulting networks, in order of decreasing stringency, as N_{25} , N_{50} , and N_{100} (Materials and Methods). Computationally translated paired reads from the shotgun metagenomic data were mapped to nodes in the microbial subnetwork using local alignment and filtered on sequence identity (Table 1, Materials and Methods).

To connect these two networks, we employed the concept of “interface metabolites:” compounds that are produced by microbes and potentially absorbed by the host as nutrients. The only connections between the host and microbial subnetworks are pairs of reactions, one from each subnetwork that each employ one of these interface metabolites: these pairs of nodes form the inner two layers of the network in Figure 3. By linking the host and microbial community metabolic network, this approach allows us to measure the distances between nodes in these two subnetworks (Materials and

Methods). We defined three sets of interface metabolites known as VFA, VFA + AA, and ALL. The VFA set consists of acetate, propionate, and butyrate which are the primary source of energy in the ruminant metabolism. The VFA + AA includes the VFA set while adding an additional twenty amino acids. Finally, the ALL group consists of the VFA + AA set while including a large set of metabolites necessary for human cellular metabolism (Materials and Methods).

Diet Differences Observed in the Proportion of Mapped Reads

The two groups of sheep (ANT and NON) showed many differences in the composition of their microbial metabolic network. Of the 2944 nodes with at least one mapped read among the 32 animals, there were 546 nodes that differed in relative abundance between the groups at a (false-discovery rate-corrected) P -value of (0.05 or less) and 346 nodes where $P \leq 0.01$ (Wilcoxon test with FDR correction, see Materials and Methods).

ANT Fed Sheep have Reactions Closer to the Host Network Compared to NON-Fed Sheep

Reads from ANT-fed animals map to reactions that are significantly closer to the host metabolic network compared with NON-fed animals for all combinations of currency and interface metabolites ($P < 0.001$, Table 2, randomization test, see Materials and Methods).

We tested four network statistics for differences between ANT and NON-fed sheep. The average number of carbon atoms involved in the reactions did not differ ($P > 0.05$, Table 3) on average between the two groups. The between centrality and clustering coefficients of the microbial networks did not show differences ($P > 0.05$, Table 3),

between the groups for N_{50} and N_{100} , but did show a weak difference in the N_{25} network. Node degree was greater ($P < 0.05$, Table 3) for NON than ANT for N_{25} and N_{50} .

Shannon entropy

We used Shannon entropy as a measure of the diversity and evenness of the observed OTU and metabolic reaction distributions. The metabolic network nodes showed no difference ($P = 0.16$) in entropy between the ANT and NON-groups. However, the ANT group had a significantly higher ($P = 0.0124$) OTU alpha diversity compared with the NON-group.

Lack of correlation between ANT and NON body weights with OTUs, nodes, and mean layer

Because the ANT animals had greater body mass ($P < 0.05$) compared to the NON animals, we examined the potential for a relationship between body size and the observed distribution of OTUs, metabolic nodes, and mean network layer individually for the ANT and for the NON-fed sheep. We tested the prediction that increased body mass within each feed group (ANT and NON) had a predictable influence on the OTU, metabolic reaction, or mean layer distributions. However, we found no statistically significant ($P > 0.05$, Table 4) association between body size and any of these three predictors in either of two groups. We concluded that these differences in mass are not driving the other network and OTU associations described.

DISCUSSION

As would be expected from prior work, administering antibiotics in the feed clearly alters the microbiome of the sheep rumen at both taxonomically and functionally.

These differences are evident at the level of individual microbial taxa (OTUs) and metabolic reactions. However, they are perhaps more interpretable from a more global perspective. The antibiotic-fed animals showed greater alpha taxonomic diversity, most likely because the dominance of *Prevotella* is reduced, giving a more even taxonomic profile. At the metabolic level, the ANT group showed a microbial community metabolic network that is closer in network space to the host network than seen for the non-antibiotic fed animals.

While these differences in the two groups are strongly statistically supported, their functional implications are less certain. It is difficult to properly assess the degree of association between taxonomic differences in the ruminant gut with metabolic functions for sheep fed ANT and NON-diets, respectively (Doolittle & Booth, 2017). Comparing these results to previous studies conducted in our group, we noted that a shift from a grass to a grain diet also shifted the microbial network closer to the host animal (Wolff et al., 2017), while animals that were efficient in their use of feedstuffs for growth had microbial networks closer to the host than did inefficient animals (Patil et al., 2018). Hence, this shift in the microbial network to be nearer the host can have different meanings in differing experimental contexts: it marked a deterioration the diet shift experiment but improved feed efficiency in the growth study. It is not clear which, if either, of these two interpretations applies to this study. However, when we consider that the ANT group also showed a shift away from a dominance of *Prevotella* species, it is at least plausible to argue that some of the complex plant compound degradation capacity that is provided by these microbes was also lost, a hypothesis supported by the lower mean network degree for the ANT animals CMNs. Since those complex compounds are

distant from the host network, the antibiotic treatment may have shifted the microbiome from one more like those we previously observed for forage diets to one more similar to grain-type diets (Wolff et al., 2017). Likewise, this loss of a dominant taxonomic group is likely the driver of the increased alpha diversity of the ANT microbiomes. On the other hand, we observed a proportion *Fibrobacter* abundance of 8.0% in the ANT animals compared to 6.5% in the NON-sheep. Several studies have argued that the presence of *Fibrobacter* facilitates cellulose degradation and enhances ruminant digestion (Kobayashi, Shinkai, & Koike, 2008). There were similarity higher levels of *Ruminococcus* in the ANT animals (10.8% verse 6.3%). These observations would tend to suggest that ANT-fed sheep might be more efficient in the degradation of cellulose products and cell wall breakdown compared with NON-fed sheep. However, both of these hypotheses are predicated on the understanding of OTU distributions one group at a time. Unfortunately, our approaches do not measure absolute abundances, only relative ones: it is possible that the decrease in the proportion of *Prevotella* is the only signal in these data, with the observed increases in other groups being merely the mathematical result of the change (Gloor et al., 2017).

The use of antibiotics as feed additives is considered a standard agricultural practice: it increases weight gain and prevents the spread of infectious diseases. Recently, use of antibiotics has been reconsidered due to the emergence of antibiotic resistance bacteria (Ventola, 2015). Our analysis with the use of antibiotics in feed raises the question of how this benefit accrues, given that antibiotics reduced the presence of bacteria such as *Prevotella* in the gut that may be essential in the absorption of nutrients by the host.

One caveat of our approach is that our analysis of the microbial metabolic networks is limited by the identifiable enzymes currently found in the MetaCyc database, and our conclusions could change with increasing knowledge of microbial metabolism (Caspi et al., 2014). Currently, there is still an intense debate as to whether more ecologically diverse microbiomes are generally indicators of good health (Cotillard et al., 2013). Several studies have suggested that reduced microbial diversity in the digestive system of humans have led to Crohn's disease (Sha et al., 2013) and irritable bowel syndrome (Carroll et al., 2012). The supplementation of antibiotics into these sheep's feed drastically reduced the abundance of *Prevotella*, which is essential for metabolic capacities in the gut. However, the converse side of this effect was to increase the alpha diversity of the microbiome. In the same vein, the closer proximity of the metabolic networks of the ANT-fed sheep to the host reactions could suggest that these animals are better able to extract essential metabolites from their microbes than are NON-fed sheep. The resolution of the apparent paradox and a general theory of the effect of antibiotics and on the structure of the microbiome will depend not on just additional data but new system approaches to modeling these systems such as community-scale flux models (Henson, 2015; Karlsson et al., 2011). Such large-scale approaches can both assess the differences introduced by diet and potentially direct beneficial alterations to these important ecosystems.

Acknowledgements

This project was supported by an USDA Agriculture and Food Research Initiative (AFRI) grant no. 2016-67016-24903 (R.D.P., W.R.L., K.M.C.) and National Science Foundation grant no. NSF-CCF-1421765 (G.C.C.).

Table 1: ANT/NON animal read statistics

Class	Animal ID	Total Paired-Reads After Quality Filtering^a	Number of Paired-Reads Matching MetaCyc^b	Number of Paired-Reads Matching OTU^c
Antibiotic	1	65,204,914	410,878 (0.63%)	1,670
	2	55,808,192	352,775 (0.63%)	1,339
	3	67,849,728	523,716 (0.77%)	2,492
	4	74,489,684	365,604 (0.49%)	1,419
	5	61,870,354	477,270 (0.77%)	1,971
	6	56,964,636	386,128 (0.68%)	1,618
	7	60,328,818	492,325 (0.82%)	2,098
	8	64,343,318	285,088 (0.44%)	1,058
	9	64,319,660	365,598 (0.57%)	1,409
	10	58,369,728	397,396 (0.68%)	1,699
	11	76,514,826	220,188 (0.29%)	741
	12	65,489,840	437,031 (0.67%)	1,396
	13	68,523,034	337,579 (0.49%)	1,132
	14	54,909,336	390,215 (0.71%)	1,436
	15	71,660,440	526,259 (0.73%)	2,404
	16	64,874,606	417,827 (0.64%)	1,701
Non-Antibiotic	1	119,833,042	310,921 (0.26%)	1,188
	2	115,699,275	307,898 (0.27%)	1,241
	3	17,712,966	253,800 (1.43%)	842
	4	16,336,193	183,859 (1.13%)	644
	5	119,789,299	320,006 (0.27%)	1,048
	6	126,030,561	316,015 (0.25%)	1,124
	7	131,776,487	375,399 (0.28%)	1,699
	8	113,135,369	336,808 (0.30%)	1,145
	9	136,044,846	439,771 (0.32%)	1,553

10	129,561,726	323,016 (0.25%)	1,031
11	126,647,671	242,727 (0.19%)	819
12	83,677,521	220,034 (0.26%)	701
13	202,547,203	706,605 (0.35%)	2,517
14	100,366,944	197,717 (0.20%)	683
15	101,143,440	528,514 (0.52%)	2,017
16	109,260,821	313,093 (0.29%)	1,091

a: Total paired-reads sequenced after quality filtering

b: Number and percent of paired-reads that were mapped to nodes from MetaCyc

c: Number of specific OTUs identified in the sequences

Table 2: Differences in network layer and antibiotic treatment

Group ^a	Currency Cutoff ^b	Mean layer: NON ^c	Mean layer: ANT ^c	Real Difference: mean layers ^d	Max Random Difference ^e	P ^f
VFA	N ₂₅	1.975	1.958	0.016	0.005	< 0.001
	N ₅₀	1.876	1.833	0.043	0.005	< 0.001
	N ₁₀₀	1.781	1.737	0.044	0.004	< 0.001
VFA_AA	N ₂₅	0.921	0.912	0.009	0.004	< 0.001
	N ₅₀	0.962	0.933	0.029	0.003	< 0.001
	N ₁₀₀	0.905	0.876	0.029	0.003	< 0.001
ALL	N ₂₅	0.663	0.652	0.011	0.002	< 0.001
	N ₅₀	0.624	0.614	0.010	0.002	< 0.001
	N ₁₀₀	0.501	0.491	0.010	0.001	< 0.001

a: Interface metabolite set (*Materials and Methods*).

b: Network after currency metabolic removal (*Materials and Methods*).

c: Mean layer number for reads mapped from NON or ANT animals (*Materials and Methods*).

d: Difference between the mean layer for NON and ANT animals

e: Maximum random difference between the mean layer for NON and ANT (*Materials and Methods*).

f: P-value tests for the null hypothesis that the mean layer of the NON was greater than ANT. Using the P-value test, reads were randomly assigned to diets and mean layers were computed 2000 times by using a two-sided test. P-values less than 0.05 are statistically significant which are shown in bold.

Table 3: ANT/NON network structure

Statistic ^a	Network ^b	RFI ^c	ANT ^c	Real Difference (Low-High) ^d	Mean Random Difference ^e	P ^f
Carbon Sum	N ₂₅	33.12	33.13	-0.01	-0.001	0.999
	N ₅₀	33.12	33.13	-0.01	-0.001	1
	N ₁₀₀	33.12	33.13	-0.01	0.001	1
Betweenness Centrality	N ₂₅	8,957.2	8,915.5	41.7	1.4	<0.001
	N ₅₀	18,092.6	18,308.4	-215.8	-0.2	1
	N ₁₀₀	18,124.4	18,696.5	-572.1	-1.5	1
Degree	N ₂₅	3.52	3.50	0.02	<0.001	<0.001
	N ₅₀	9.18	9.17	0.01	-0.001	<0.001
	N ₁₀₀	12.11	12.18	-0.06	-0.001	1
Clustering Coefficient	N ₂₅	0.88	0.88	- 0.001	-9.0x10 ⁻⁶	<0.001
	N ₅₀	0.86	0.86	-0.001	-4.0x10 ⁻⁶	1
	N ₁₀₀	0.85	0.85	-3.0x10 ⁻⁶	-5.0x10 ⁻⁵	0.24

a: Network statistic compared between the RFI and ANT animals. Carbon sum: Total number of carbon atoms in a reaction. Betweenness-centrality: Number of shortest paths crossing a node. Degree: Number of edges of a node. Clustering coefficient: Number of fully connected triangles a node participates in relative to its total number of edges (*Materials and Methods*).

b: Network defined by currency metabolite removal (*Materials and Methods*).

c: Mean value for selected network statistic mapped for RFI or ANT animals.

d: Difference between the mean statistic for RFI and ANT animals.

e: Mean random difference between RFI and ANT groups when reads are randomly assigned.

f: P-value for the test of the null hypothesis that there is no difference in the mean statistic. P was computed by randomly assigning reads to RFI groups and comparing the difference in the two mean values for 1,000 such randomizations to the actual data (see *Materials and Methods*). P-values less than 0.025 were judged significant and are shown in bold.

Table 4: Pearson's and Spearman's correlation test

Group^a	ANT r^b	NON r^c	P^d	ANT r_s^e	NON r_s^f	P^g
OTUs	0.21	0.13	0.42	0.22	0.08	0.36
Metabolic nodes	0.18	0.04	0.36	0.15	-0.04	0.31
VFA N ₅₀	-0.56	-0.26	0.17	-0.39	-0.25	0.35
VFA_AA N ₅₀	-0.48	-0.26	0.26	-0.42	-0.19	0.25

a: Response variables.

b: ANT animals Pearson's correlation.

c: NON animals Pearson's correlation.

d: Using the *P*-value test to compare Pearson's correlation between ANT and NON animals.

e: ANT animals Spearman's correlation.

f: NON animals Spearman's correlation.

g: Using the *P*-value test to compare Spearman's correlation between ANT and NON animals.

Figure 1: The distribution of Shannon entropies of the mapped OTUs (x-axis) and metabolic network nodes (y-axis) for ANT (red) and NON-fed animals (black). The ANT group had a higher ($P < 0.05$) mean OTU entropy compared with NON-group while there was no significant difference in node entropy between the two groups.

Figure 2: Microbial diversity between ANT (black) and NON-fed sheep (red). The x-axis is the species rank abundance where the most prevalent OTU is ranked one and every OTU thereafter after is ranked in descending order. The y-axis is the frequency of each ranked OTU.

Figure 3: The interaction between the host and microbial metabolic network for ANT and NON-fed sheep. A node represents an enzyme-catalyzed reaction were the edges connect reactions that share metabolites. The interface metabolites are depicted as the colored lines between the host organism and microbes showing the potential transfer of metabolites between each network. Microbial nodes are colored based on the normalized $\log_2$ difference of read counts between the two diets, respectively. There were 546 metabolic reactions that were different ($P < 0.05$, Wilcoxon test) in abundance between ANT and NON-fed diets.

REFERENCES

- Bekaert, M., and Conant, G. C. 2011. Copy number alterations among mammalian enzymes cluster in the metabolic network. *Mol Biol Evol*, 28(2), 1111-1121. <https://doi.org/10.1093/molbev/msq296>.
- Benjamini, Y. a. H., Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. In (Vol. 57, pp. 289-300). *Journal of the Royal Statistical Society. Series B (Methodological)*. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Bergman, E. N. 1990. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol Rev*, 70(2), 567-590. <https://doi.org/10.1152/physrev.1990.70.2.567>.
- Cameron, A., and McAllister, T. A. 2016. Antimicrobial usage and resistance in beef production. *J Anim Sci Biotechnol*, 7, 68. <https://doi.org/10.1186/s40104-016-0127-3>.
- Carroll, I. M., Ringel-Kulka, T., Siddle, J. P., and Ringel, Y. 2012. Alterations in composition and diversity of the intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome. *Neurogastroenterol Motil*, 24(6), 521-530, e248. <https://doi.org/j.1365-2982.2012.01891.x>.
- Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., Holland, T. A., Keseler, I. M., Kothari, A., Kubo, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D. S., Weerasinghe, D., Zhang, P., and Karp, P. D. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases.

- Nucleic Acids Res*, 42(Database issue), D459-471.
<https://doi.org/10.1093/nar/gkt1103>.
- Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., Marsh, T., Garrity, G. M., and Tiedje, J. M. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res*, 37(Database issue), D141-145.
<https://doi.org/10.1093/nar/gkn879>.
- Cotillard, A., Kennedy, S. P., Kong, L. C., Prifti, E., Pons, N., Le Chatelier, E., Almeida, M., Quinquis, B., Levenez, F., Galleron, N., Gougis, S., Rizkalla, S., Batto, J. M., Renault, P., ANR MicroObser consortium, Dore, J., Zucker, J. D., Clement, K., Ehrlich, S. D., and consortium, A. M. 2013. Dietary intervention impact on gut microbial gene richness. *Nature*, 500(7464), 585-588.
<https://doi.org/10.1038/nature12480>.
- Dijkstra, E. W. 1959. A note on two problems in connexion with graphs. In (Vol. 1, pp. 269-271). *Numerische Mathematik*. <https://doi.org/10.1007/BF01386390>.
- Doolittle, W., and Booth, A. 2017. It's the song, not the singer: an exploration of holobiosis and evolutionary theory. *Biology and Philosophy*, 32(1), 5-24.
<https://doi.org/10.1007/s10539-016-9542-2>.
- Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., Srivas, R., and Palsson, B. 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A*, 104(6), 1777-1782. <https://doi.org/10.1073/pnas.0610772104>.

- Döring, A., Weese, D., Rausch, T., and Reinert, K. 2008. SeqAn an efficient, generic C++ library for sequence analysis. *BMC Bioinformatics*, 9, 11.
<https://doi.org/10.1186/1471-2105-9-11>.
- Ellison, M. J., Conant, G. C., Cockrum, R. R., Austin, K. J., Truong, H., Becchi, M., Lamberson, W. R., and Cammack, K. M. 2014. Diet alters both the structure and taxonomy of the ovine gut microbial ecosystem. *DNA Res*, 21(2), 115-125.
<https://doi.org/10.1093/dnares/dst044>.
- Ewing, B., Hillier, L., Wendl, M. C., and Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res*, 8(3), 175-185. <https://doi.org/10.1101/gr.8.3.175>.
- Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Giron, C. G., Gordon, L., Hourlier, T., Hunt, S., Johnson, N., Juettemann, T., Kahari, A. K., Keenan, S., Kulesha, E., Martin, F. J., Maurel, T., McLaren, W. M., Murphy, D. N., Nag, R., Overduin, B., Pignatelli, M., Pritchard, B., Pritchard, E., Riat, H. S., Ruffier, M., Sheppard, D., Taylor, K., Thormann, A., Trevanion, S. J., Vullo, A., Wilder, S. P., Wilson, M., Zadissa, A., Aken, B. L., Birney, E., Cunningham, F., Harrow, J., Herrero, J., Hubbard, T. J. P., Kinsella, R., Muffato, M., Parker, A., Spudich, G., Yates, A., Zerbino, D. R., and Searle, S. M. 2014. Ensembl 2014. *Nucleic Acids Res*, 42(Database issue), D749-755. <https://doi.org/10.1093/nar/gkt1196>.
- Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front Microbiol*, 8, 2224. <https://doi.org/10.3389/fmicb.2017.02224>.

- Henson, M. A. 2015. Genome-scale modelling of microbial metabolism with temporal and spatial resolution. *Biochem Soc Trans*, 43(6), 1164-1171.
<https://doi.org/10.1042/BST20150146>.
- Hudson, C. M., and Conant, G. C. 2011. Expression level, cellular compartment and metabolic network position all influence the average selective constraint on mammalian enzymes. *BMC Evol Biol*, 11, 89. <https://doi.org/10.1186/1471-2148-11-89>.
- Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J. F., Faraut, T., Wu, C., Muzny, D. M., Li, Y., Zhang, W., Stanton, J. A., Brauning, R., Barris, W. C., Hourlier, T., Aken, B. L., Searle, S. M. J., Adelson, D. L., Bian, C., Cam, G. R., Chen, Y., Cheng, S., DeSilva, U., Dixen, K., Dong, Y., Fan, G., Franklin, I. R., Fu, S., Guan, R., Highland, M. A., Holder, M. E., Huang, G., Ingham, A. B., Jhangiani, S. N., Kalra, D., Kovar, C. L., Lee, S. L., Liu, W., Liu, X., Lu, C., Lv, T., Mathew, T., McWilliam, S., Menzies, M., Pan, S., Robelin, D., Servin, B., Townley, D., Wang, W., Wei, B., White, S. N., Yang, X., Ye, C., Yue, Y., Zeng, P., Zhou, Q., Hansen, J. B., Kristensen, K., Gibbs, R. A., Flicek, P., Warkup, C. C., Jones, H. E., Oddy, V. H., Nicholas, F. W., McEwan, J. C., Kijas, J., Wang, J., Worley, K. C., Archibald, A. L., Cockett, N., Xu, X., Wang, W., and Dalrymple, B. P. 2014. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*, 344(6188), 1168-1173.
<https://doi.org/10.1126/science.1252806>.

- Karlsson, F. H., Nookaew, I., Petranovic, D., and Nielsen, J. 2011. Prospects for systems biology and modeling of the gut microbiome. *Trends Biotechnol*, 29(6), 251-258. <https://doi.org/10.1016/j.tibtech.2011.01.009>.
- Kobayashi, Y., Shinkai, T., and Koike, S. 2008. Ecological and physiological characterization shows that *Fibrobacter succinogenes* is important in rumen fiber digestion - review. *Folia Microbiol (Praha)*, 53(3), 195-200. <https://doi.org/10.1007/s12223-008-0024-z>.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
- Larsen, J., Anderson, N., and Preshaw, A. 2009. Long-acting moxidectin for the control of trichostrongylid infections of sheep in south-eastern Australia. *Aust Vet J*, 87(4), 130-137. <https://doi.org/10.1111/j.1751-0813.2009.00395.x>.
- Looft, T., Johnson, T. A., Allen, H. K., Bayles, D. O., Alt, D. P., Stedtfeld, R. D., Sul, W. J., Stedtfeld T. M., Chai, B., Cole, J. R., Hashsham, S. A., Tiedje, J. M., and Stanton, T. B. 2012. In-feed antibiotic effects on the swine intestinal microbiome. *Proc Natl Acad Sci U S A*, 109(5), 1691-1696. <https://doi.org/10.1073/pnas.1120238109>.
- Muloi, D., Ward, M. J., Pedersen, A. B., Fèvre, E. M., Woolhouse, M. E. J., and van Bunnik, B. A. D. 2018. Are Food Animals Responsible for Transfer of Antimicrobial-Resistant *Escherichia coli* or Their Resistance Determinants to Human Populations? A Systematic Review. *Foodborne Pathog Dis*, 15(8), 467-474. <https://doi.org/10.1089/fpd.2017.2411>.

- Needleman, S. B., and Wunsch, C. D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*, 48(3), 443-453. [https://doi.org/10.1016/0022-2836\(70\)90057-4](https://doi.org/10.1016/0022-2836(70)90057-4).
- Patil, R. D., Ellison, M. J., Wolff, S. M., Shearer, C., Wright, A. M., Cockrum, R. R., Austin, K. J., Lamberson, W. R., and Conant, G. C. 2018. Poor feed efficiency in sheep is associated with several structural abnormalities in the community metabolic network of their ruminal microbes. *J Anim Sci*, 96(6), 2113-2124. <https://doi.org/10.1093/jas/sky096>.
- Pérez-Bercoff, Å., McLysaght, A., and Conant, G. C. 2011. Patterns of indirect protein interactions suggest a spatial organization to metabolism. *Mol Biosyst*, 7(11), 3056-3064. <https://doi.org/10.1039/C1MB05168G>.
- Sha, S., Xu, B., Wang, X., Zhang, Y., Wang, H., Kong, X., Zhu, H., and Wu, K. 2013. The biodiversity and composition of the dominant fecal microbiota in patients with inflammatory bowel disease. *Diagn Microbiol Infect Dis*, 75(3), 245-251. <https://doi.org/10.1016/j.diagmicrobio.2012.11.022>.
- Shannon, C. 1948. A mathematical theory of communication. *AT&T Tech J.*, 379-423 and 623-656. <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>.
- Smith, T. F., and Waterman, M. S. (1981). Identification of common molecular subsequences. *J Mol Biol*, 147(1), 195-197. [https://doi.org/10.1016/0022-2836\(81\)90087-5](https://doi.org/10.1016/0022-2836(81)90087-5).
- Tajima, K., Aminov, R. I., Nagamine, T., Matsui, H., Nakamura, M., and Benno, Y. 2001. Diet-dependent shifts in the bacterial population of the rumen revealed with

- real-time PCR. *Appl Environ Microbiol*, 67(6), 2766-2774.
<https://doi.org/10.1128/AEM.67.6.2766-2774.2001>.
- Taxis, T. M., Wolff, S., Gregg, S. J., Minton, N. O., Zhang, C., Dai, J., Schnabel, R. D., Taylor, J. F., Kerley, M. S., Pires, J. C., Lamberson, W. R., and Conant, G. C. 2015. The players may change but the game remains: network analyses of ruminal microbiomes suggest taxonomic differences mask functional similarity. *Nucleic Acids Res*, 43(20), 9600-9612. <https://doi.org/10.1093/nar/gkv973>.
- Team, R. D. C. 2008. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Thomas, M., Webb, M., Ghimire, S., Blair, A., Olson, K., Fenske, G. J., Fonder, A. T., Christopher-Hennings, J., Brake, D., and Scaria, J. 2017. Metagenomic characterization of the effect of feed additives on the gut microbiome and antibiotic resistome of feedlot cattle. *Sci Rep*, 7(1), 12257.
<https://doi.org/10.1038/s41598-017-12481-6>.
- Truong, H., Li, D., Sajjapongse, K., Conant, G. and Becchi, M. 2014. Large-scale pairwise alignments on GPU clusters: exploring the implementation of space. *Journal of Signal Processing Systems*, 77(1-2), 131-149.
<https://doi.org/10.1007/s11265-014-0883-2>.
- Ventola, C. L. 2015. The antibiotic resistance crisis: part 1: causes and threats. *P T*, 40(4), 277-283.
- Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684), 440-442. <https://doi.org/10.1038/30918>.

Wolff, S. M., Ellison, M. J., Hao, Y., Cockrum, R. R., Austin, K. J., Baraboo, M., Burch, K., Lee, H. J., Maurer, T., Patil, R., Ravelo, A., Taxis, T. M., Truong, H., Lamberson, W. R., Cammack, K. M., and Conant, G. C. 2017. Diet shifts provoke complex and variable changes in the metabolic networks of the ruminal microbiome. *Microbiome*, 5(1), 60. [https://doi.org/ 10.1186/s40168-017-0274-6](https://doi.org/10.1186/s40168-017-0274-6).

HOST GENETICS AND THE TAXONOMIC AND FUNCTIONAL STRUCTURE OF THE BOVINE RUMIAL MICROBIAL ECOSYSTEM

ABSTRACT

Ruminant animals rely on their symbiotic microbes in their gut to break down complex plant matter into metabolites that can be used by the host. We used a microbial community metabolic network from shotgun metagenomic data to assess how breed-level differences affect the bovine ruminal microbiome. Shotgun sequences from rumen microbial DNA was collected from individuals of the Angus (n = 36) and Charolais (n = 38) breeds of cattle. Using our previously developed approach for linking the microbial and host metabolic networks to each other through the use of exchanged "interface metabolites", the metagenomic reads were mapped to an inferred microbial metabolic network that was linked to the metabolic network of the bovine host. There was no taxonomic group or single biochemical reaction that was found to be statistically different in abundance between the two groups of animals. However, the microbes of Charolais animals had a microbial metabolic network that used reactions that were closer in network space to the reactions of the host animals than were the microbial reactions of the Angus animals. The use of this systems approach identified subtle differences in the structure of the rumen microbiome between cattle breeds that were not readily observable at the level of individual taxonomic units or metabolic reactions.

INTRODUCTION

Ruminant mammals maintain a symbiotic relationship with their rumen microbes: in return for sheltering these organisms and delivering nutrients to them, the microbes

benefit the host by degrading inaccessible plant substrates and converting them into usable volatile fatty acids and amino acids (Stevens & Hume, 1998). As such, these rumen microbes play an important nutritional role for domesticated ruminants such as cattle, with the microbes provide energy for animal growth and maintenance and nutritionally supporting the high milk production of dairy cattle. The rumen is predominantly occupied by strict anaerobes, and the resulting difficulties in recovering and culturing such organisms means that many of them have still not been classified or had their ecological roles identified (Kobayashi, Shinkai, & Koike, 2008). This problem is in fact a daunting one: the number of bacterial species present in the rumen is still relatively unknown but is estimated to be around several thousand species based on next-generation sequencing assays and the resulting collectors curves (Fouts et al., 2012). This taxonomic richness makes the ruminal ecosystem uniquely complex compared to other animal, terrestrial, or aquatic environments (Ley, Lozupone, Hamady, Knight, & Gordon, 2008).

Animal gut microbiomes have been extensively studied from functional and disease perspectives, exploring how diet and illness perturb this complex ecosystem (Ezenwa, Gerardo, Inouye, Medina, & Xavier, 2012; Ley et al., 2008; Qin et al., 2012). Recently, experiments on the gut microbiome have determined that obese individuals have lower abundance of Bacteroidetes compared with lean people, suggesting certain proportional bacterial abundances are essential for healthy individuals (Turnbaugh et al., 2009). Qin et al. identified specific gut microbial makers and decreased abundance of butyrate-producing bacteria in individuals with type 2 diabetes (Qin et al., 2012). Likewise, ruminants used in animal agriculture have had their microbiomes studied with

the goal of improving economically desirable traits, such as feed efficiency (Guan, Nkrumah, Basarab, & Moore, 2008; Hernandez-Sanabria et al., 2010; Herrero et al., 2013; McCann, Wiley, Forbes, Rouquette, & Tedeschi, 2014). However to date, only a few studies have explored how the genome of the host animal may alter the gut microbiome in ruminant animals (Clemmons, Voy, & Myer, 2019; Weimer, 2015). That such an effect exists is strongly suggested by other work, such as a study in mice that found a mutation in the leptin gene was associated with obesity and resulted in differences in the gut microbial composition relative to lean mice (Ley et al., 2005). Furthermore, a study has documented that cows exchanging ruminal contents with each other exhibited a degree of host specificity by having bacterial composition that was similar to its preexchange bacterial composition (Weimer, Stevenson, Mantovani, & Man, 2010). Other studies have found differences in the bacterial populations of different ruminant species: while these differences may be primarily attributable to differences in diet, an additional role for host animal genetics is also suspected (Ghorbani, Morgavi, Beauchemin, & Leedle, 2002; Weimer, 2015).

To gain a better understanding of the complexities of the ruminant microbiome, we wanted to examine breed-level differences in both the taxonomic structure of the ruminal microbiome and in the structure of the metabolic networks induced by those taxa. Taxonomic analysis has been an essential tool in providing scientists with the ability to compare biodiversity and understanding relationships of different organisms living in different ecosystems. As an example Guan et al. used such approaches to discover a correlation between feed efficiency in cattle and the presence of specific bacterial groups (Guan et al., 2008). Our analysis is of interest because previous studies have suggested a

casual link between the core rumen microbiota and its function such as the digestibility in sheep and milk production efficiency and methane emission in cattle (Morgavi et al., 2015; Wallace et al., 2019). It is important to compare both the taxonomy and the metabolic network structure between breeds because prior work has shown that even organisms having microbiomes of differing taxonomic structure can be metabolically similar (Taxis et al., 2015; Turnbaugh et al., 2009), meaning that breed-level differences in taxonomy are not necessarily of functional importance.

We have developed a community metabolic network approach for analyzing vertebrate gut microbiomes, an approach that is very appropriate for this problem. With this approach, we have assessed how the relative abundance of known microbial enzymes in host animals change as a result of differences in diet and feed efficiency (Patil et al., 2018; Wolff et al., 2017). Sheep fed concentrate diets had a smaller group of reactions that were in closer proximity to the network space to the host's reactions compared to sheep fed forage diets (Wolff et al., 2017). Also, we compared sheep with differing feed efficiencies with the same approach and discovered that controlled with changes in the diet, there were still significant, but smaller, differences in the taxonomic structure and metabolic networks of efficient versus inefficient animals (Patil et al., 2018). Our work is in keeping with the belief that a greater use of predictive modeling based on the integration of taxonomic data and community metabolic interactions will allow for a greater understanding of how the rumen's microbiota affects the host community structure (Greenblum, Turnbaugh, & Borenstein, 2012; Ley et al., 2005).

The use of sequencing data was been crucial in identifying new biomass-degrading genes and genomes of important functionality in the bovine's rumen (Hess et

al., 2011). By performing shotgun metagenomics, it allows for the detection of a full complement of microbial DNA in the rumen leading to the discovery of metabolic capabilities that couldn't be determined solely by taxa sequencing (Tringe et al., 2005). We are better able to understand the complexities of the rumen by understanding the relationship of the host ruminant and rumen microbes through the network interface to understand the functional structure. Using such sequence data, coupled to a systems biology approach, we sought to determine if there are any underlying structural differences in the microbe's community metabolic network that may be caused by breed differences in bovine animals.

MATERIAL AND METHODS

Breed selection and rumen sample collection

Rumen fluid was collected from two cattle breeds, Angus (n =36) and Charolais (n = 38). The Angus (ANG) and Charolais (CHA) animals were born and raised physically separated from each other to prevent cross-contamination due to cohabitation. All animals were fed a chopped forage diet for a 1-month period before transitioning to a total mixed ration (low energy/forage-based) for a 70-day feed intake trial. A dosing syringe was used to syphon fluid from the rumen: the resulting samples were stored in 200 mL containers (Jami, Israel, Kotser, & Mizrahi, 2013). The rumen samples were filtered using multiple layers of cheesecloth to remove any remaining feed particles and then immediately stored on dry ice until the samples were placed at -80°C for long-term storage.

DNA extraction and Illumina sequencing

We extracted DNA from the rumen samples for Illumina sequencing following the protocol established by Guan et al. (Guan et al., 2008). Low speed centrifugation was used to remove any particulate matter from the calf's rumen samples. Silicon (0.1 g of 0.5 mm) and zirconia (0.3 g of 0.1 mm) beads, together with one mL of lysis buffer was added to the thawed rumen samples, and the resulting mixture was homogenized using a Mini-Beadbeater-8 at its maximum speed for 3 minutes followed by incubation at 70°C for 15 min and centrifugation at 4°C for 5 min. Following this step, the supernatant was transferred to a new tube where further lysis buffer (300 µL) was added to the pelleted beads. The previous steps of homogenization, incubation, and centrifugation were then repeated, and the supernatants were pooled together. The QIAamp DNA Stool Mini Kit (Qiagen, Santa Clarita, CA) was used to remove any RNA, proteins, and nucleic acids remaining in the supernatant.

Genomic libraries were constructed from the DNA samples following the recommended protocol in Illumina's TruSeq DNA PCR-free Library Prep kit. The Diagenode BioRuptor was used to shear genomic DNA resulting in fragmented DNA. The 3' and 5' overhangs were removed by an end pair reaction while adenosine nucleotide was added to the 3' ends and Illumina adapters were ligated. A Qubit assay was used to quantify each of the purified libraries while the Agilent BioAnalyzer High Sensitivity DNA assay was used to confirm the fragment size. The genomic libraries were then diluted and sequenced with an Illumina HiSeq 2500, yielding 100 base-pair, paired-end reads.

Sequence quality filtering and OTU analysis

We quality filtered the paired-end reads by removing all bases immediately following three consecutive bases having a phred quality score of less than 15 (Ewing, Hillier, Wendl, & Green, 1998). Any read pair where either member was less than 85 bases in length or had a phred score of less than 25 was omitted from further analyses. The filtered reads were then compared to the 16S rDNA sequences from the Ribosomal Database Project (RDP) (Cole et al., 2009). Operational taxonomic units (OTUs) were computed from this database by computing all the possible pairwise Needleman-Wunsch sequence alignments between database sequences (Needleman & Wunsch, 1970) by using our GPU-based pairwise alignment tool (Truong, 2014). We used these pairwise alignments to create a network where a node was represented by a database sequence and edges connected pairs of sequences with greater than 97% sequence identity. OTUs were defined as connected components in this network (Ellison et al., 2014). Bowtie was then used to match metagenomic read pairs to the sequences in the database (Langmead, Trapnell, Pop, & Salzberg, 2009). Read pairs having a sequence match of greater than 97% identity to one or more database sequences in the same component/OTU were considered to be a member of that operational taxonomic unit (OTU) (Cole et al., 2009).

Inference of metabolic networks from the MetaCyc database

MetaCyc is a database that consists of inferred metabolic pathways and enzymes from all the domains of life, but primarily from bacteria and archaeans (Caspi et al., 2014). The MetaCyc database not only contains metabolic pathways but also has enzymes, metabolites, genes, and biochemical reactions (includes substrates and products

of these reactions) annotated to those pathways. We used this database to extract enzyme sequences that were linked to one or more metabolic reactions. This database was then used to infer the host and microbial metabolic networks where edges are connected to a pair of nodes sharing a common metabolite (Taxis et al., 2015). Reactions having identical metabolites were considered a single node in the network, which resulted in 6,140 reaction nodes in the community metabolic network (Wolff et al., 2017). Furthermore, abundant metabolites like water and hydrogen ions, referred to as “currency metabolites,” were removed prior to the construction of a metabolic network; their presence would have created an overly connected network and linkage between unrelated parts of the network. Three different stringencies were defined for currency metabolites: namely compounds occurred in more than 25, 50, or 100 MetaCyc reactions. The resulting metabolic networks will be referred to as N₂₅, N₅₀, and N₁₀₀ (Bekaert & Conant, 2011; Pérez-Bercoff, McLysaght, & Conant, 2011).

Construction of the bovine host and microbial metabolic network

In this study, we constructed two metabolic networks, one was for the bovine host and the other for the rumen microbes. Using comparative genomics, the human metabolic network was added to the MetaCyc bovine metabolic network because the bovine network has significantly fewer annotated reactions (1,404 in bovine compared to 2,863 for humans). Using our previously described tool that uses sequence similarity and syteny to infer orthology, we identified 2,553 bovine orthologs of the human enzymes found in the MetaCyc database by using Ensembl 75 (Flicek et al., 2014). These genes matched 1,850 network nodes, which were added to the bovine nodes that were not also

present in the human metabolic network (Bekaert & Conant, 2014). The final bovine host network thus had 2,126 nodes.

Matching translated metagenomic reads to microbial network

To identify the filtered reads that matched to MetaCyc sequences, reads from 74 samples from the two bovine groups were translated into their associated amino acid sequences in all six possible open reading frames (ORF). Paired reads each having translated ORF greater than 29 residues were selected and searched against the MetaCyc protein database using a search tool based on the SeqAn library (Döring, Weese, Rausch, & Reinert, 2008). A match was inferred between a translated read and a database sequence having alignments with greater than 80% amino acid identity over 80% of the paired metagenomic ORFs (Smith & Waterman, 1981). The threshold of 80% identity was selected as a cutoff based on our prior analyses (Wolff et al., 2017) which suggested that the proportion of reads having valid hits to the database leveled off at this cutoff, while there was an increase in the proportion of reads mapped to multiple reactions as the percent identity value dropped further. Moreover, as the percent identity threshold was increased, the proportion of reads mapping to incorrect nodes increased too. There are instances when enzymes in the MetaCyc database are annotated as catalyzing more than one reaction leading to an ORF pair having multiple matches in the database. When this happens, the multiple matches could map to the same set of reactions, subset of reactions, or non-overlapping set of reactions. If it maps to a non-overlapping set of reactions, the read pair is removed, as the reaction involved cannot be determined.

Interface between bovine and microbial metabolic networks

We compiled three sets of “interface metabolites” that may be absorbed by the host animal from the rumen microbes. Because the details of exchange of the metabolites between the host and microbes are not fully understood: we used these three different interface sets to assess whether this uncertainty was affecting our results. The first interface set of metabolites consists of three of the most common volatile fatty acids (VFAs) in the rumen: acetate, butyrate, and propionate (Bergman, 1990). The second set combines the previous VFAs and the 20 universal amino acids (VFA + AA). The third set contains the VFA + AA with additional metabolites that are known to be absorbed by various human cells which were obtained from the reconstruction of the human metabolic network (Duarte et al., 2007). This final set (ALL) includes a total of 204 compounds and excludes DNA/RNA, nucleotides, carbon dioxide, and NAD⁺ and NADP⁺. The bovine and microbial network are linked to each other by edges connecting a pair of nodes between each network indicating a shared interface metabolite.

Construction of merged metabolic network layers

Using these interface metabolites, we constructed a merged metabolic network including both the host and microbial networks. We used Dijkstra’s algorithm to determine the shortest path length between every pair of nodes in the network (Dijkstra, 1959). The distance between a node and the other subnetwork was defined to be the shortest path between it and a node in that subnetwork. Hence, a microbial node that shared an interface metabolite with a host node would have a minimal distance of 0. We defined the mean layer for a particular sample as weighted average of this distance, taken over all nodes and mapped reads.

Node read density analysis

We used a two-sample Wilcoxon test to determine if there were significant differences in the per-node normalized reads counts between the ANG and CHA breeds. Only nodes that had at least one mapped read pair between the two groups was analyzed with this test. The resulting P-values were corrected for multiple tests by applying a 5% false-discovery rate correction (Benjamini, 1995).

Network structure analysis

The mean network layers of the mapped reads for ANG and CHA groups were calculated for each possible combination of networks and interface metabolites (3 interface metabolite sets by 3 network currency cutoffs = 9 possible combinations). To determine if there were any significant differences between the mean layers between the 2 breed groups, we used a randomization method. This randomization consisted of first pooling the reads from ANG and CHA groups and then randomly selected the same number of reads that were found in ANG and labeling this set as “pseudo-ANG:” the remaining reads were labeled as “pseudo-CHA”. We compared the mean layer observed in the actual data with those computed from the 1,000 randomized data simulations to determine if there was a difference in the mean layer between the breeds. We concluded that there was a significant difference in the mean layer between the breeds if less than 5% of the randomized datasets.

Metabolic network structure statistics

We examined the effect of breed on the metabolic network structure by comparing the following network statistics between the breeds: carbon sum, node degree, betweenness centrality, and clustering coefficient. We defined the carbon sum as the

total number of carbon atoms that were present in the metabolites involved in a node's reaction. The node degree is the total edge count for a node, and the clustering coefficient measures the likelihood that nodes are connected to each other's neighboring nodes (Watts & Strogatz, 1998). The nodes betweenness-centrality is the total number of shortest paths between any pair of network nodes that pass through a particular node (Freeman, Borgatti, & White, 1991). For each node and breed, the network statistics were weighted by each node's normalized mapped read count and the absolute differences between those weighted values for the two breeds was computed and compared to the distribution of differences from the randomized networks. We inferred a significant difference between the breeds for a particular statistic if 5% or fewer of the randomized networks showed differences as large or larger than in the actual dataset.

RESULTS

Metagenomic Sequencing of Bovine Rumen Fluid

Microbial DNA was extracted from ruminal fluid of individuals from two different cattle breeds, Angus (n = 36) and Charolais (n = 38). These breeds were selected for metagenomic sequencing because they showed differences in feed efficiency presumably as a result of differing selective pressures in their recent domestication history (Snowder et al., 2005). We shotgun sequenced the microbial DNA of the rumen samples on an Illumina HiSeq 2500, obtaining approximately 2.1 billion 100 bp paired-end reads after quality filtering (Table 1). Across these 74 samples, we identified 610 bacterial and archaeal operational taxonomic units (OTUs) that were present in at least

one animal. No single OTU was statistically different in its abundance between the ANG and CHA groups.

Metabolic Network Interference

We used the MetaCyc database (Caspi et al., 2014) to define a reaction-centric metabolic network in which two subnetworks were constructed: one for the microbes and one for the bovine host. We assumed that a set of interface metabolites (see below) were produced by the microbes and then absorbed as nutrients in the host: this assumption allows us to define a movement of metabolites from the microbial network to that of the host. In potential microbial network is defined by a set of nodes representing enzyme-catalyzed reactions from MetaCyc that are also associated with one or more protein sequence matches from the MetaCyc enzyme database. The bovine network was inferred by using the MetaCyc bovine metabolic network with additional bovine orthologs from the human metabolic network added. In our analysis, we used three different stringencies to remove any overly abundant currency metabolites such as ATP or water from the metabolic network, defining three network cutoffs: N_{25} , N_{50} , and N_{100} (See Materials and Methods). Each rumen sample was mapped to the microbial metabolic network which were connected by these interface metabolites that were defined as VFA, VFA + AA, and ALL with the inferred bovine metabolic network. This allowed us to have a greater power of detecting differences in the interactions between the interface of rumen microbes and host by focusing on the individual animal's mean layer (Figure 1). It is clearly evident that Angus cattle have individual mean layers that are more equally distributed compared to Charolais cattle as observed for N_{50} .

We then sought to match sequences from our shotgun data to this network by running a fast-sequencing matching algorithm on the conceptually-translated metagenomic reads against the database of enzyme sequences from MetaCyc. This approach allowed us to infer which enzymes were present in each of the rumen samples and at what relative abundance, defining our microbial network.

Breed Differences Observed in Proportion of Mapped Reads

We explored if there were any differences between the cattle breeds by comparing the enzyme abundance between ANG and CHA groups using a Wilcoxon test assess if the abundance differences between the individual metabolic reactions were statistically significant. None of the nodes differed in the read abundance out of the 6,140 nodes that mapped to at least one read from the bovine dataset with a (false-discovery) rate-corrected $P < 0.05$.

CHA Bovine has Reactions Closer in Proximity to Host Network Compared to ANG Bovine

Employing this concept of interface metabolites to define a network distance, we found that the reads from the microbes of the CHA animals mapped to nodes that were closer to the host metabolic network compared with ANG animals ($P < 0.05$) (Table 2). We employed four network-level statistics to evaluate network structural differences between the bovine breeds (Table 3). The average number of carbon atoms involved in the reactions was significantly greater in CHA animals than the ANG animals ($P < 0.05$). The microbial reads from the CHA animals mapped to nodes with higher degree and higher betweenness centrality ($P < 0.05$) there was a significant, though numerically small difference in clustering coefficients between the breeds ($P < 0.05$).

DISCUSSION

In cattle production, nutritional and genetic strategies have been employed to increase feed efficiency and weight gain in animals. The nutritional status of ruminant animals is affected by several factors such as diet, animal management, host genetics, and the symbiotic microbes in the gut (Myer, 2019). In this study, we wanted to gain a better understanding of the effects of the host genome on the ruminant ecosystem, both functionally and taxonomically, by comparing different cattle breeds. Perhaps interestingly, no differences were seen between the breeds when comparing individual microbial taxa or enzyme frequency. We hence employed a systems approach to assess whether there were more subtle differences between the two breeds, since such approaches better understand model the interaction between the rumen microbiota and the host. We found that the CHA animals, on average, had microbial ecosystems whose reactions were closer in network space to the host metabolic network and employed reactions with longer carbon atoms compared than did the ANG animals. We might therefore hypothesize that this combination of closer reactions and the observed higher betweenness-centrality for the CHA animals might allow them to more efficiently transfer metabolites from the microbial network to that of the host.

There are only a few studies that have explored the effect of host genetics on the rumen microbiome. Differences in bacterial composition were observed when comparing Holstein and Jersey cows under similar dietary conditions (Paz, Anderson, Muller, Kononoff, & Fernando, 2016). Another study surveyed rumen samples from over 32 different ruminants and found that 30 of the most common rumen bacterial groups were present in 90% of the samples. This result suggested that there might be a core

microbiome that is conserved across ruminant animals (Henderson et al., 2015).

However, all of these results need to be considered with the caveat that diet appears to have a greater influence on the rumen microbial composition than do even differences in host species, not to speak of breeds within a species.

Recently, a few studies have documented the effects of breed on the rumen microbiome through metagenomic analysis. For instance, metatranscriptomic analyses revealed a linkage between rumen microbiota and feed efficiency at a functional level (Li & Guan, 2017; Li et al., 2019). A study comparing breed effects determined Angus and Charolais cattle had different bacterial composition in the gut microbiome (Li et al., 2019). These studies could be useful in future studies to link the rumen microbes to specific traits of economic interest in cattle. Previously, we compared the taxonomic and microbial structure of two steers by mapping reads to a microbial metabolic network (Taxis et al., 2015), finding that while the animals were relatively dissimilar in their taxonomic structure, the community metabolic networks of induced by those microbes were more similar. The use of this community microbial metabolic network was able to uncover aspects of the microbiome structure that would not have been detected solely by taxonomic analysis.

There has been debate as to whether comparing taxonomy and function with each other is a meaningful given that these methods are characterized by problems of scale and the ability to distinguish individuals (Inkpen, 2017). At the same time, studies have documented that function is more highly conserved compared with taxonomic composition (Consortium, 2012; Taxis et al., 2015). In our study, the lack of significant differences at the taxonomic level between the rumen microbiome of the two breeds

could have been attributed to their similar diet and environmental conditions. At the metabolic level, it remains to be determined what the effect of the closer proximity of the microbial networks of the CHA animals to the host's metabolic network is. In previous work (Patil et al. 2018), we found that such closer proximity was predictive of increased feed efficiency, which could also be the case here.

A few points need to be borne in mind with regard to our analyses. First the differences observed could be a result of some currently unknown difference in the environment for the two breeds, though the experiments were generally intended to avoid such issues. On a more positive note, our results suggest that future work may uncover further interesting differences in microbiome structure by focusing on larger functional associations rather than individual taxa or reactions. The analyses performed here were the largest sample of animals we have yet used without microbial metabolic network pipelines, and both the statistical power that resulted and the stability of our approach to larger samples are encouraging. We expect the power of approaches such as this one to only improve, and sequencing technologies and databases improve, potentially leading to true predictive models of the mass flow through community metabolic networks that will explain phenotypic differences that we believe are coupled to microbiome structure.

Table 1: Breed group animal read statistics

Breeds	Animal ID	Total Paired- Reads After Quality Filtering ^a	Number of Paired-Reads Matching MetaCyc ^b	Number of Paired- Reads Matching OTU ^c
Angus	847	28,073,481	358,739 (1.28%)	2,866
	664	31,732,768	152,329 (0.48%)	1,663
	635	30,084,914	269,186 (0.89%)	3,225
	700	28,061,407	208,205 (0.74%)	2,043
	692	25,447,749	262,360 (1.03%)	613
	733	36,040,510	514,473 (1.43%)	1,152
	776	27,687,597	149,845 (0.54%)	1,233
	765	26,613,844	139,053 (0.52%)	1,408
	738	27,277,296	42,401 (0.16%)	82
	710	28,666,814	36,008 (0.13%)	69
	662	29,008,162	21,611 (0.07%)	49
	701	38,761,562	85,155 (0.22%)	315
	811	34,179,490	62,486 (0.18%)	154
	678	36,961,873	57,457 (0.16%)	172
	688	30,541,061	42,686 (0.14%)	115
	637	29,281,960	60,629 (0.21%)	208
	715	34,181,634	266,851 (0.78%)	1,754
	641	43,530,185	298,086 (0.68%)	2,322
	691	34,720,033	300,761 (0.87%)	2,801
	634	19,660,243	107,114 (0.54%)	864
	667	17,521,082	166,128 (0.95%)	1,131
	214	22,053,445	239,410 (1.09%)	1,699
	221	23,523,770	170,148 (0.72%)	982
	227	16,331,022	172,279 (1.05%)	1,273
	237	27,788,294	51,033 (0.18%)	246
	763	31,047,340	48,851 (0.16%)	181
	758	28,021,519	43,180 (0.15%)	167
	689	24,555,834	41,445 (0.17%)	175
	730	23,202,533	32,538 (0.14%)	121
	628	20,743,488	28,755 (0.14%)	40
	813	26,218,902	44,960 (0.17%)	171
	737	26,846,193	37,194 (0.14%)	89
	841	26,881,082	49,378 (0.18%)	182
	693	28,402,359	59,588 (0.21%)	222
	752	28,639,426	54,158 (0.19%)	177
	723	30,068,618	52,954 (0.18%)	167

a: Total paired-reads sequenced after quality filtering

b: Number and percent of paired-reads that were mapped to nodes from MetaCyc

c: Number of specific OTUs identified in the sequences

Table 1: Breed group animal read statistics

Breeds	Animal ID	Total Paired-Reads After Quality Filtering ^a	Number of Paired-Reads Matching MetaCyc ^b	Number of Paired-Reads Matching OTU ^c
Charolais	757	26,651,556	105,939 (0.40%)	817
	844	21,452,023	71,288 (0.33%)	363
	785	35,749,440	514,681 (1.44%)	3,049
	770	23,296,595	204,923 (0.88%)	500
	666	32,427,749	194,099 (0.60%)	1,912
	609	24,847,516	347,134 (1.40%)	424
	773	21,057,581	59,829 (0.28%)	727
	822	30,830,913	287,714 (0.93%)	1,987
	845	38,473,265	77,609 (0.20%)	203
	636	35,329,016	56,548 (0.16%)	181
	820	32,774,531	45,344 (0.14%)	176
	836	43,947,319	83,948 (0.19%)	215
	646	37,386,333	53,685 (0.14%)	128
	772	35,308,754	79,190 (0.22%)	223
	699	40,667,321	55,858 (0.14%)	185
	744	37,054,952	58,743 (0.16%)	150
	783	39,158,230	302,387 (0.77%)	1,861
	696	38,924,626	300,761 (0.77%)	1,451
	616	18,920,712	106,018 (0.56%)	695
	711	24,630,593	161,121 (0.65%)	1,067
	617	17,624,499	50,433 (0.29%)	721
	718	19,218,547	127,508 (0.66%)	522
	759	14,731,351	47,941 (0.33%)	519
	690	18,692,666	98,210 (0.53%)	771
	606	20,151,359	30,991 (0.15%)	89
	833	23,974,453	23,030 (0.10%)	84
	681	20,350,101	5,831 (0.03%)	13
	726	25,908,354	20,585 (0.08%)	97
	792	27,021,558	60,911 (0.23%)	139
	829	24,323,496	29,770 (0.12%)	59
	729	26,171,101	44,164 (0.17%)	105
	817	26,040,455	25,282 (0.10%)	56
	768	24,971,090	46,754 (0.19%)	143
	749	23,789,294	55,505 (0.23%)	176
	612	28,257,851	63,162 (0.22%)	253
	673	30,343,126	67,703 (0.22%)	207
	698	38,406,503	74,234 (0.19%)	195
	716	31,133,397	63,024 (0.20%)	174

a: Total paired-reads sequenced after quality filtering

b: Number and percent of paired-reads that were mapped to nodes from MetaCyc

c: Number of specific OTUs identified in the sequences

Table 2: Breed group network position

Group ^a	Currency Cutoff ^b	Mean layer: ANG ^c	Mean layer: CHA ^c	Real Difference: mean layers ^d	Max Random Difference ^e	<i>P</i> ^f
VFA	N ₂₅	4.537	4.577	-0.040	0.009	<0.001
	N ₅₀	3.254	3.243	0.010	0.005	<0.001
	N ₁₀₀	2.805	2.796	0.009	0.004	<0.001
VFA_AA	N ₂₅	1.543	1.517	0.026	0.006	<0.001
	N ₅₀	1.326	1.302	0.024	0.004	<0.001
	N ₁₀₀	1.290	1.267	0.023	0.003	<0.001
ALL	N ₂₅	0.791	0.787	0.003	0.004	0.007
	N ₅₀	0.664	0.659	0.005	0.002	<0.001
	N ₁₀₀	0.681	0.678	0.003	0.002	<0.001

a: Interface metabolite set

b: Network

c: Mean layer number for reads mapped from ANG or CHA cattle

d: Difference between the mean layer for ANG and CHA cattle

e: Maximum random difference between the mean layer for ANG and CHA

f: *P*-value tests for the null hypothesis that the mean layer of the ANG was greater than CHA. Using the *P*-value test, reads were randomly assigned to breeds and mean layers were computed 2000 times by using a two-sided test. *P*-values less than 0.05 are statistically significant which are shown in bold.

Table 3: Breed group network structure

Statistic ^a	Network ^b	ANG ^c	CHA ^c	Absolute Difference (ANG-CHA) ^d	Mean Random Difference ^e	P ^f
Carbon Sum	N ₂₅	34.18	34.56	0.38	0.001	<0.001
	N ₅₀	34.18	34.56	0.38	0.003	<0.001
	N ₁₀₀	34.18	34.56	0.38	0.002	<0.001
Betweenness Centrality	N ₂₅	8,232.85	8,249.63	16.78	1.09	0.001
	N ₅₀	15,818.23	16,040.94	222.71	2.22	<0.001
	N ₁₀₀	16,631.22	17,010.17	378.94	0.83	<0.001
Degree	N ₂₅	3.35	3.38	0.03	0.0003	<0.001
	N ₅₀	7.72	7.82	0.10	0.001	<0.001
	N ₁₀₀	10.57	10.73	0.16	0.0003	<0.001
Clustering Coefficient	N ₂₅	0.89	0.89	0.0002	<0.0001	0.001
	N ₅₀	0.87	0.87	0.001	<0.0001	<0.001
	N ₁₀₀	0.86	0.86	0.001	<0.0001	<0.001

a: Network statistic compared between the ANG and CHA animals. Carbon sum: Total number of carbon atoms in a reaction. Betweenness-centrality: Number of shortest paths crossing a node. Degree: Number of edges of a node. Clustering coefficient: Number of fully connected triangles a node participates in relative to its total number of edges (*Materials and Methods*).

b: Network defined by currency metabolite removal (*Materials and Methods*).

c: Mean value for selected network statistic mapped for ANG or CHA animals.

d: Absolute difference between the mean statistic for ANG and CHA animals.

e: Mean random difference between ANG and CHA groups when reads are randomly assigned.

f: P-value for the test of the null hypothesis that there is no difference in the mean statistic. P was computed by randomly assigning reads to ANG groups and comparing the difference in the two mean values for 1,000 such randomizations to the actual data (see *Materials and Methods*). P-values less than 0.05 were judged significant and are shown in bold.

Figure 1: Mean layer of the individual animal samples for ANG and CHA for N_{50}

REFERENCES

- Bekaert, M., & Conant, G. C. (2011). Copy number alterations among mammalian enzymes cluster in the metabolic network. *Mol Biol Evol*, 28(2), 1111-1121. doi:10.1093/molbev/msq296
- Bekaert, M., & Conant, G. C. (2014). Gene duplication and phenotypic changes in the evolution of mammalian metabolic networks. *PLoS One*, 9(1), e87115. doi:10.1371/journal.pone.0087115
- Benjamini, Y. a. H., Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. In (Vol. 57, pp. 289-300). *Journal of the Royal Statistical Society. Series B (Methodological)*.
- Bergman, E. N. (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol Rev*, 70(2), 567-590.
- Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., . . . Karp, P. D. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res*, 42(Database issue), D459-471. doi:10.1093/nar/gkt1103
- Clemmons, B. A., Voy, B. H., & Myer, P. R. (2019). Altering the Gut Microbiome of Cattle: Considerations of Host-Microbiome Interactions for Persistent Microbiome Manipulation. *Microb Ecol*, 77(2), 523-536. doi:10.1007/s00248-018-1234-9

- Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., . . . Tiedje, J. M. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res*, 37(Database issue), D141-145. doi:10.1093/nar/gkn879
- Consortium, H. M. P. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), 207-214. doi:10.1038/nature11234
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. In (Vol. 1, pp. 269-271). *Numerische Mathematik*
- Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., . . . Palsson, B. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A*, 104(6), 1777-1782. doi:10.1073/pnas.0610772104
- Döring, A., Weese, D., Rausch, T., & Reinert, K. (2008). SeqAn an efficient, generic C++ library for sequence analysis. *BMC Bioinformatics*, 9, 11. doi:10.1186/1471-2105-9-11
- Ellison, M. J., Conant, G. C., Cockrum, R. R., Austin, K. J., Truong, H., Becchi, M., Lamberson, W. R., & Cammack, K. M. "Diet Alters Both the Structure and Taxonomy of the Ovine Gut Microbial Ecosystem." *DNA Res*, vol. 21, no. 2, 2014, pp. 115-125, doi:10.1093/dnares/dst044.

- Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res*, 8(3), 175-185.
- Ezenwa, V. O., Gerardo, N. M., Inouye, D. W., Medina, M., & Xavier, J. B. (2012). Microbiology. Animal behavior and the microbiome. *Science*, 338(6104), 198-199. doi:10.1126/science.1227412
- Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., . . . Searle, S. M. (2014). Ensembl 2014. *Nucleic Acids Res*, 42(Database issue), D749-755. doi:10.1093/nar/gkt1196
- Fouts, D. E., Szpakowski, S., Purushe, J., Manolito, T., Waterman, R. C., MacNeil, M. D., Alexander, L. J., & Nelson, K. E. "Next Generation Sequencing to Define Prokaryotic and Fungal Diversity in the Bovine Rumen." *PLoS One*, vol. 7, no. 11, 2012, p. e48289, doi:10.1371/journal.pone.0048289.
- Freeman, L. C., Borgatti, S. P., & White, D. R. (1991). Centrality in valued graphs: A measure of betweenness based on network flow. In (Vol. 13 (2), pp. 141-154). *Social Networks*.
- Ghorbani, G. R., Morgavi, D. P., Beauchemin, K. A., & Leedle, J. A. (2002). Effects of bacterial direct-fed microbials on ruminal fermentation, blood variables, and the microbial populations of feedlot cattle. *J Anim Sci*, 80(7), 1977-1985. doi:10.2527/2002.8071977x

- Greenblum, S., Turnbaugh, P. J., & Borenstein, E. (2012). Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc Natl Acad Sci U S A*, 109(2), 594-599. doi:10.1073/pnas.1116053109
- Guan, L. L., Nkrumah, J. D., Basarab, J. A., & Moore, S. S. (2008). Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. *FEMS Microbiol Lett*, 288(1), 85-91. doi:10.1111/j.1574-6968.2008.01343.x
- Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Janssen, P. H., & Collaborators, G. R. C. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci Rep*, 5, 14567. doi:10.1038/srep14567
- Hernandez-Sanabria, E., Guan, L. L., Goonewardene, L. A., Li, M., Mujibi, D. F., Stothard, P., . . . Leon-Quintero, M. C. (2010). Correlation of particular bacterial PCR-denaturing gradient gel electrophoresis patterns with bovine ruminal fermentation parameters and feed efficiency traits. *Appl Environ Microbiol*, 76(19), 6338-6350. doi:10.1128/AEM.01052-10
- Herrero, M., Havlík, P., Valin, H., Notenbaert, A., Rufino, M. C., Thornton, P. K., . . . Obersteiner, M. (2013). Biomass use, production, feed efficiencies, and greenhouse gas emissions from global livestock systems. *Proc Natl Acad Sci U S A*, 110(52), 20888-20893. doi:10.1073/pnas.1308149110

- Hess, M., Sczyrba, A., Egan, R., Kim, T. W., Chokhawala, H., Schroth, G., . . . Rubin, E. M. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science*, 331(6016), 463-467. doi:10.1126/science.1200387
- Inkpen, S. A., Douglas, G. M., Brunet, T. D. P., Leuschen, K., Doolittle, W. F., & Langille M. G. I. (2017). The coupling of taxonomy and function in micro biomes. In (pp. 1225-1243): Biology & Philosophy.
- Jami, E., Israel, A., Kotser, A., & Mizrahi, I. (2013). Exploring the bovine rumen bacterial community from birth to adulthood. *ISME J*, 7(6), 1069-1079. doi:10.1038/ismej.2013.2
- Kobayashi, Y., Shinkai, T., & Koike, S. (2008). Ecological and physiological characterization shows that *Fibrobacter succinogenes* is important in rumen fiber digestion - review. *Folia Microbiol (Praha)*, 53(3), 195-200. doi:10.1007/s12223-008-0024-z
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10(3), R25. doi:10.1186/gb-2009-10-3-r25
- Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A*, 102(31), 11070-11075. doi:10.1073/pnas.0504978102

- Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R., & Gordon, J. I. (2008). Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol*, 6(10), 776-788. doi:10.1038/nrmicro1978
- Li, F., & Guan, L. L. (2017). Metatranscriptomic Profiling Reveals Linkages between the Active Rumen Microbiome and Feed Efficiency in Beef Cattle. *Appl Environ Microbiol*, 83(9). doi:10.1128/AEM.00061-17
- Li, F., Li, C., Chen, Y., Liu, J., Zhang, C., Irving, B., . . . Guan, L. L. (2019). Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. *Microbiome*, 7(1), 92. doi:10.1186/s40168-019-0699-1
- McCann, J. C., Wiley, L. M., Forbes, T. D., Rouquette, F. M., & Tedeschi, L. O. (2014). Relationship between the rumen microbiome and residual feed intake-efficiency of Brahman bulls stocked on bermudagrass pastures. *PLoS One*, 9(3), e91864. doi:10.1371/journal.pone.0091864
- Morgavi, D. P., Rathahao-Paris, E., Popova, M., Boccard, J., Nielsen, K. F., & Boudra, H. (2015). Rumen microbial communities influence metabolic phenotypes in lambs. *Front Microbiol*, 6, 1060. doi:10.3389/fmicb.2015.01060
- Myer, P. R. (2019). Bovine Genome-Microbiome Interactions: Metagenomic Frontier for the Selection of Efficient Productivity in Cattle Systems. *mSystems*, 4(3). doi:10.1128/mSystems.00103-19

- Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*, 48(3), 443-453. doi:10.1016/0022-2836(70)90057-4
- Patil, R. D., Ellison, M. J., Wolff, S. M., Shearer, C., Wright, A. M., Cockrum, R. R., . . . Conant, G. C. (2018). Poor feed efficiency in sheep is associated with several structural abnormalities in the community metabolic network of their ruminal microbes. *J Anim Sci*, 96(6), 2113-2124. doi:10.1093/jas/sky096
- Paz, H. A., Anderson, C. L., Muller, M. J., Kononoff, P. J., & Fernando, S. C. (2016). Rumen Bacterial Community Composition in Holstein and Jersey Cows Is Different under Same Dietary Condition and Is Not Affected by Sampling Method. *Front Microbiol*, 7, 1206. doi:10.3389/fmicb.2016.01206
- Pérez-Bercoff, Å., McLysaght, A., & Conant, G. C. (2011). Patterns of indirect protein interactions suggest a spatial organization to metabolism. *Mol Biosyst*, 7(11), 3056-3064. doi:10.1039/c1mb05168g
- Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., . . . Kristiansen, K. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*, 490(7418), 55-60. doi:10.1038/nature11450
- Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular subsequences. *J Mol Biol*, 147(1), 195-197.
- Snowder, G. D., Van Vleck, L. D., Cundiff, L. V., & Bennett, G. L. (2005). Influence of breed, heterozygosity, and disease incidence on estimates of variance components

of respiratory disease in preweaned beef calves. *J Anim Sci*, 83(6), 1247-1261.
doi:10.2527/2005.8361247x

Stevens, C. E., & Hume, I. D. (1998). Contributions of microbes in vertebrate gastrointestinal tract to production and conservation of nutrients. *Physiol Rev*, 78(2), 393-427.

Taxis, T. M., Wolff, S., Gregg, S. J., Minton, N. O., Zhang, C., Dai, J., . . . Conant, G. C. (2015). The players may change but the game remains: network analyses of ruminal microbiomes suggest taxonomic differences mask functional similarity. *Nucleic Acids Res*, 43(20), 9600-9612. doi:10.1093/nar/gkv973

Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., . . . Rubin, E. M. (2005). Comparative metagenomics of microbial communities. *Science*, 308(5721), 554-557. doi:10.1126/science.1107851

Truong, H., Li, D., Sajjapongse, K., Conant, G. & Becchi, M. (2014). Large-scale pairwise alignments on GPU clusters: exploring the implementation of space. *Journal of Signal Processing Systems*, 77(1-2), 131-149.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., . . . Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature*, 457(7228), 480-484. doi:10.1038/nature07540

Wallace, R. J., Sasson, G., Garnsworthy, P. C., Tapio, I., Gregson, E., Bani, P., . . . Mizrahi, I. (2019). A heritable subset of the core rumen microbiome dictates dairy

cow productivity and emissions. *Sci Adv*, 5(7), eaav8391.

doi:10.1126/sciadv.aav8391

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks.

Nature, 393(6684), 440-442. doi:10.1038/30918

Weimer, P. J. (2015). Redundancy, resilience, and host specificity of the ruminal

microbiota: implications for engineering improved ruminal fermentations. *Front*

Microbiol, 6, 296. doi:10.3389/fmicb.2015.00296

Weimer, P. J., Stevenson, D. M., Mantovani, H. C., & Man, S. L. (2010). Host specificity

of the ruminal bacterial community in the dairy cow following near-total

exchange of ruminal contents. *J Dairy Sci*, 93(12), 5902-5912.

doi:10.3168/jds.2010-3500

Wolff, S. M., Ellison, M. J., Hao, Y., Cockrum, R. R., Austin, K. J., Baraboo, M., . . .

Conant, G. C. (2017). Diet shifts provoke complex and variable changes in the

metabolic networks of the ruminal microbiome. *Microbiome*, 5(1), 60.

doi:10.1186/s40168-017-0274-6

SUMMARY

Using a systems biology approach, we used a community metabolic network to study the changes that occur in the ruminant microbiome due to feed efficiency, feed additives, and breed differences in ruminant animals. This approach of creating a merged metabolic network between rumen microbes and host through the use of “interface metabolites” allowed us to gain a better understanding of the actual interactions that were occurring in the rumen ecosystem that could not be explained solely by taxonomic analyses. This approach of mapping reads to the MetaCyc database allowed for computational prediction of metabolic pathways in the rumen from metagenomic reads. This approach can be important in situations when there is lack of preexisting functional information or when organisms are unable to be genetically manipulated.

In our studies, it was clearly apparent that using a metabolic network aided in providing us with a better understanding of the rumen ecosystem that we were not able to observe by solely analyzing the individual taxonomic and metabolic reactions. Specifically in our feed efficiency and breed ruminant studies, we did not observe any single enzyme that was statistically over or under abundant when comparing each group, with a Wilcoxon test with FDR correction. Using a systems approach, we were able to show subtle differences in the microbial structure of the rumen that could not be observed by focusing on individual units. As more enzymes, metabolic pathways, and metabolites are experimentally validated and included in the MetaCyc database, we will be able to make accurate computational predictions of the interactions between the microbes of the rumen and host.

VITA

Rocky Patil was born on May 21, 1984. Throughout his childhood, he was involved in STEM organizations that peaked his curiosity in the sciences. During his high school years, Rocky particularly enjoyed his chemistry and biology courses as his teachers; Mr. Cole and Mr. Taylor were passionate educators who simultaneously made science informative, fun, and exciting. These STEM experiences led Rocky to attend the University of Nevada at Las Vegas (UNLV) on the prestigious Provost's scholarship where he graduated with a bachelors of science in biological sciences. After undergraduate, Rocky focused his research interests primarily on food security and animal health as he graduated with a master's of science in microbiology from Iowa State University and a master's of science in food science from Michigan State University. Rocky continued his graduate studies by pursuing his Ph.D. in animal sciences and a minor in statistics at the University of Missouri at Columbia under the supervision of Dr. Gavin Conant. During his time, Rocky has published his research in the Journal of Animal Sciences and presented his work a few times at the Plant and Animal Genome meeting.