



OPEN

# Influence of host genetics in shaping the rumen bacterial community in beef cattle

Waseem Abbas<sup>1,4</sup>, Jeremy T. Howard<sup>1,4</sup>, Henry A. Paz<sup>1,2</sup>, Kristin E. Hales<sup>3</sup>, James E. Wells<sup>3</sup>, Larry A. Kuehn<sup>3</sup>, Galen E. Erickson<sup>1</sup>, Matthew L. Spangler<sup>1</sup>✉ & Samodha C. Fernando<sup>1</sup>✉

In light of recent host-microbial association studies, a consensus is evolving that species composition of the gastrointestinal microbiota is a polygenic trait governed by interactions between host genetic factors and the environment. Here, we investigated the effect of host genetic factors in shaping the bacterial species composition in the rumen by performing a genome-wide association study. Using a common set of 61,974 single-nucleotide polymorphisms found in cattle genomes ( $n = 586$ ) and corresponding rumen bacterial community composition, we identified operational taxonomic units (OTUs), Families and Phyla with high heritability. The top associations (1-Mb windows) were located on 7 chromosomes. These regions were associated with the rumen microbiota in multiple ways; some (chromosome 19; position 3.0–4.0 Mb) are associated with closely related taxa (*Prevotellaceae*, *Paraprevotellaceae*, and *RF16*), some (chromosome 27; position 3.0–4.0 Mb) are associated with distantly related taxa (*Prevotellaceae*, *Fibrobacteraceae*, *RF16*, *RFP12*, *S24-7*, *Lentisphaerae*, and *Tenericutes*) and others (chromosome 23; position 0.0–1.0) associated with both related and unrelated taxa. The annotated genes associated with identified genomic regions suggest the associations observed are directed toward selective absorption of volatile fatty acids from the rumen to increase energy availability to the host. This study demonstrates that host genetics affects rumen bacterial community composition.

Complex and diverse microbial communities facilitate the degradation of nutrients within ruminants<sup>1</sup>. The composition of this complex microbial community is shaped by the highly dynamic physical, chemical, and predatory conditions within the rumen, and potentially by genetic factors of the host<sup>1–6</sup>. In turn, the microbial community contributes to the environmental conditions within the rumen and the nutrient availability to the host<sup>2</sup>. This complex microbial community can convert otherwise unusable organic matter into useable protein and energy and can provide up to 70% of the animal's protein and energy needs<sup>7</sup> in the form of microbial cell protein (MCP) or volatile fatty acids (VFAs) for host metabolism. Therefore, ruminal microbial diversity and abundance critically influences nutrient cycling and when inorganic nutrients and carbon are made available for host consumption. As such, differences in the microbial community can change the energy profiles available to the ruminant host and its subsequent performance.

In ruminants, the microbial population is established by successive waves where convergence of microbial populations are seen reaching a more stable population structure<sup>8–11</sup>. The establishment of a gut microbial community is influenced by multiple factors, including diet, environment, and host genotype<sup>2–6,12,13</sup>. Among these contributors, the influence of diet on gut microbial population structure is well established<sup>14–17</sup>. Rumen bacterial species' and functional composition of the rumen microbiota is an important factor that contributes towards animal performance and efficiency<sup>18</sup>. Additionally, studies in small ruminants have demonstrated the host mucosal innate immune function to affect rumen microbial community composition<sup>19</sup>. With studies demonstrating host genetics to influence host immunity<sup>20</sup>, it is tempting to speculate in ruminants, host genetics may directly or indirectly affect rumen microbial community composition. However, our understanding of how a stable rumen microbial community assembles and what factors affect rumen microbial community composition

<sup>1</sup>Department of Animal Science, University of Nebraska-Lincoln, Lincoln, NE, USA. <sup>2</sup>Department of Animal and Dairy Science, Mississippi State University, Starkville, MS, USA. <sup>3</sup>Meat Animal Research Center, Clay Center, NE 68933, USA. <sup>4</sup>These authors contributed equally: Waseem Abbas and Jeremy T. Howard. ✉email: mspangler2@unl.edu; samodha@unl.edu

and function is limited. One such process that is still largely unknown is how the ruminant host genotype effect rumen microbial community assembly. Recent studies have demonstrated that the rumen microbiota is influenced by host genetic factors<sup>5,6,21</sup>. In mammals, Benson et al. demonstrated genome-wide linkages of bacterial taxa abundance in the murine gut using Quantitative Trait Locus (QTL) mapping. While breeds did not have broad representation, a recent study suggested that microbial phylotypes can be influenced by the sire breed and can impact host metabolism<sup>22</sup>.

In light of these recent host-microbial association studies, a consensus is evolving that species' composition of the gastrointestinal microbiota is a polygenic trait governed by interactions between host genetic factors and the environment. As such, genome-wide association studies (GWAS) can be used to identify host chromosomal regions and a subset of single nucleotide polymorphisms (SNP) that influence microbiome composition and function in the rumen. Here we evaluated the effect of host genetic factors in shaping the bacterial species composition in the rumen and the role of such associations on host metabolism by performing a genome wide association study using different cohorts of beef cattle totaling 586 animals. All animals were genotyped using medium to high density single nucleotide polymorphisms (SNP) chips (770 K, 150 K, or 80 K) and the underlying SNPs were used for analysis of genome wide associations using the bacterial community composition as the phenotype. The rumen microbial community of all 586 animals were phenotyped by sampling and characterizing the V4 region of the 16S rRNA gene and was used as a collection of individual traits to perform GWAS to identify host chromosomal regions and subset of SNP that influence bacterial community composition and assembly within the rumen. This study provides detailed and novel insight into how host genotype in ruminants can influence rumen microbial assembly. We also quantify the genomic heritability for bacterial taxonomic traits thus identifying the degree to which they are controlled by host genetic background. This study also demonstrates the potential to utilize associations between rumen microbiota and genetic markers for use in genomic selection and marker-assisted management that can be used to improve feed efficiency, animal health and microbiome manipulation mediated by selecting for favorable microbial taxa within the rumen.

## Results

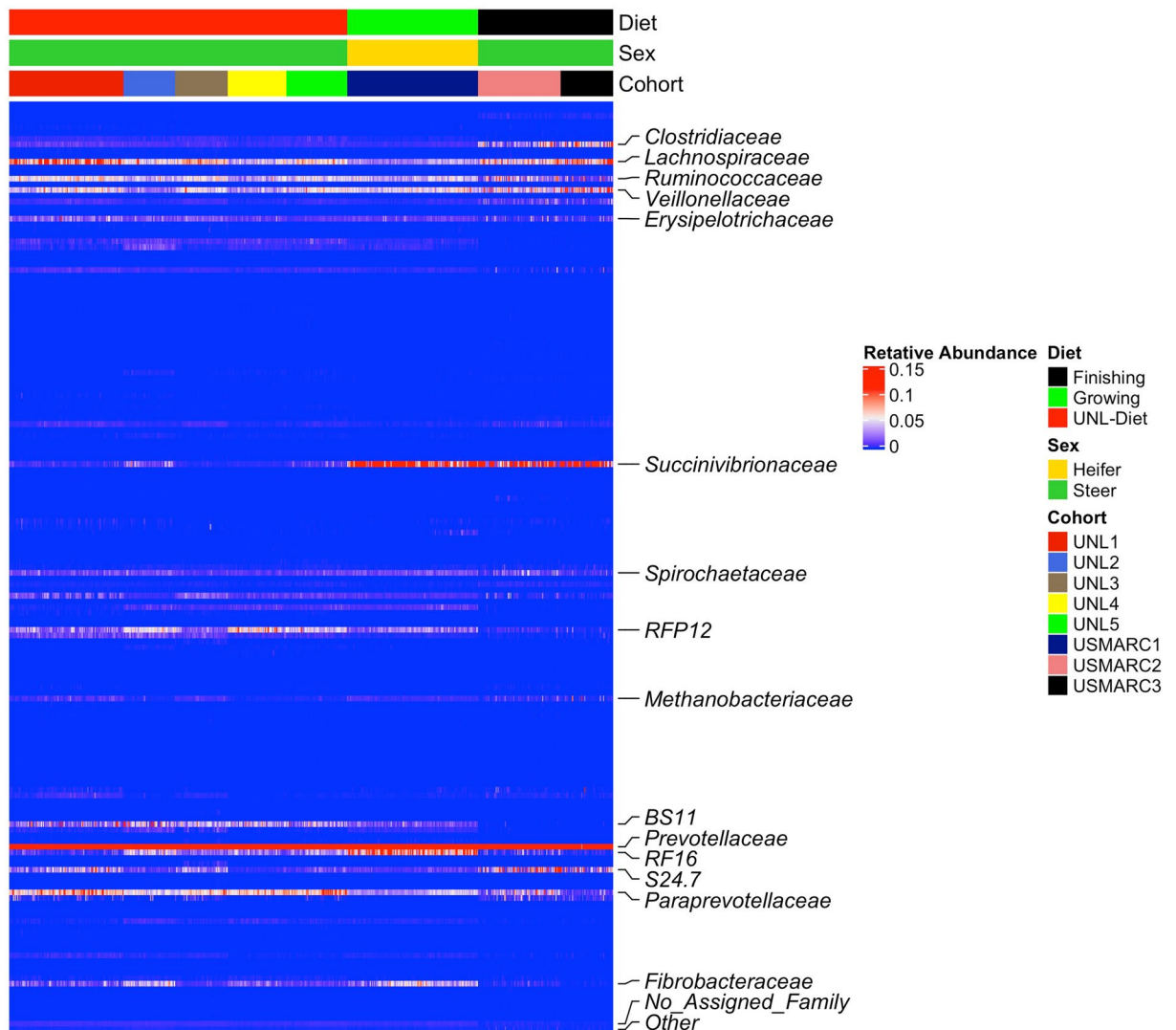
**The bacterial populations within the cattle rumen.** The bacterial community within the rumen microbiome of 8 cohorts were phenotyped by sequencing the V4 region of the 16S rDNA gene. The sequencing resulted in 18,992,394 quality filtered reads which included 9,755,502 reads from the USMARC cattle cohorts and 9,236,892 reads from the UNL cohorts. The rarefaction curves and goods coverage tests displayed adequate sampling depth to provide a detailed and quantitative estimate of the rumen bacterial community composition within each animal. The taxonomic classification of normalized OTUs at phylum, family and OTU level were performed using Naive Bayes classifier<sup>23</sup> using the greengenes database<sup>24</sup>. This analysis detected 439 genera, 237 families, 131 orders, 68 classes and 32 phyla from 7,228 OTUs identified from the 586 animals sequenced. The distribution of taxa across all cohorts were similar with the exception of family Succinivibrionaceae which were greater in the USMARC cohorts (Fig. 1). The relative abundance of the major phyla included Bacteroidetes 48.85%; Firmicutes 24.95%; Proteobacteria 13.3%; Verrucomicrobia 3.14% and Tenericutes 2.24%.

The distribution of OTUs across the cohorts displayed wide animal-to-animal variation, therefore to better characterize taxa that are more conserved across animals and variable in abundance, a core measurable microbiome (CMM) was identified from the total dataset. To ensure robust repeatability of bacterial phenotyping only OTUs present in at least 1% of the animals and was also part of the CMM were used for subsequent analysis. The CMM contained 99.94% of the total reads generated and therefore represent a major portion of the rumen bacterial population. A principal coordinate analysis (PCoA) (supplementary figure S3) and PERMANOVA analysis were performed to identify combined effects of diet, sex, location on microbial community. As all factors (diet, sex, location) cannot be untangled from each other, the PERMANOVA analysis was performed using management type to reflect the collective effect of all above factors. PERMANOVA analysis displayed a significant effect of management type ( $p < 0.001$ ) on microbial community composition. Therefore, in the model used for GWAS analysis, we included contemporary group (concatenation of management type and year) as a fixed effect to adjust for the variation in the microbial community composition resulting from management type.

**Rumen bacterial community behaves as a “polygenic trait”.** To determine the extent of and to correct for population structure, a principle component analysis (PCA) on a genomic relationship matrix (G) was utilized. The first two principle components annotated by location (UNL and USMARC) are illustrated in Fig. 2. The degree of population differentiation across the two feeding locations was minimal as the first principal component only described 3% of the variation in population structure. We utilized genome wide association analysis using SNP to assess the contribution of host genotype to the variation of bacterial taxa within the rumen. Normalized abundance of each OTU, family and phylum were used as traits to test for associations with 61,974 SNP.

The posterior mean heritability estimates determined using a Bayesian genomic best linear unbiased prediction (GBLUP) model across classes within OTU, family, and phylum level were 0.161, 0.150, and 0.194, respectively. Although posterior mean heritability estimates were low, some taxonomic groups displayed high heritability estimates. The maximum heritability estimates of OTU, family, and phylum level were, 0.820, 0.722, and 0.722, respectively (Fig. 3). Taxa belonging to phyla *Euryarchaeota*, *TM6* and *Proteobacteria* and families, *Methanobacteriaceae*, *Sphaerochaetaceae*, and *Succinivibrionaceae* had heritability estimates greater than 0.5. Additionally, 364 OTUs had heritability estimates higher than 0.5 (supplementary Tables S2, S3, and S4). The effective sample size and summary statistics of heritability estimates are shown in supplementary figure S1.

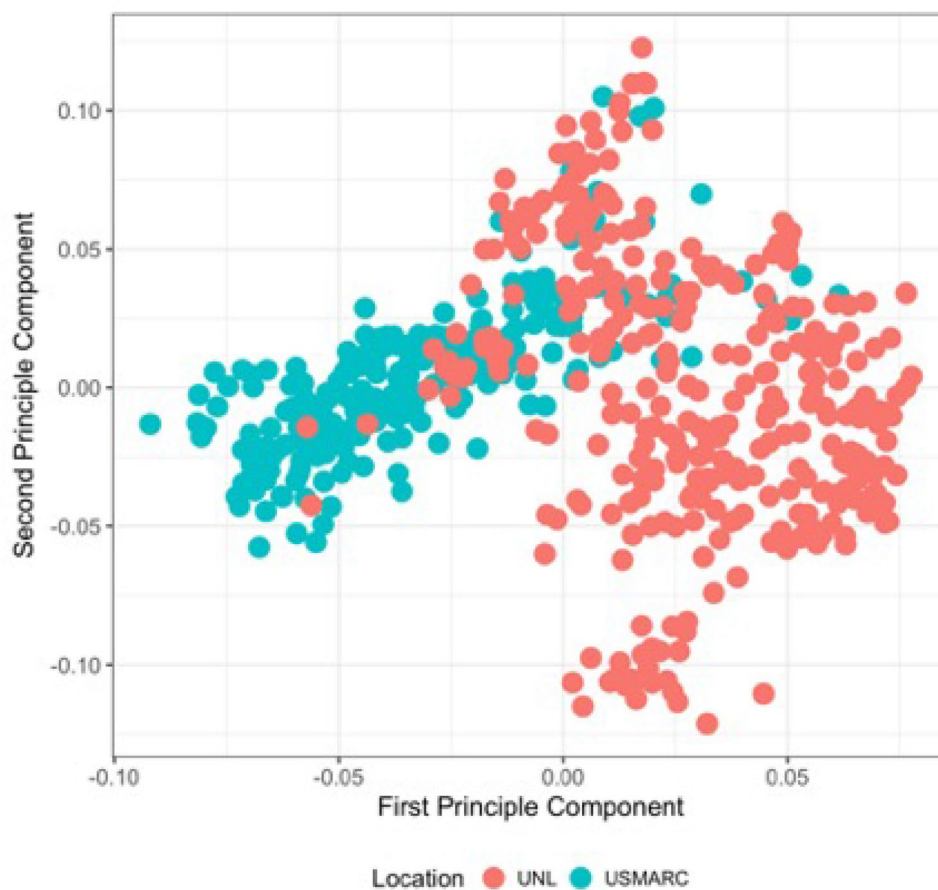
For genome wide association analysis, the estimated additive genetic effects across animals were decomposed into marker effects across OTU, family, and phylum within each group. Chromosomal regions with the largest WGBV for the selected OTU, family, and phylum are shown in Table 1. The identified SNP in the bovine



**Figure 1.** Distribution of taxa across all cohorts. The heatmap above displays the abundance and distribution of all taxa identified in at least 1% of the total animals ( $n = 586$ ). The relative abundances of 237 families present within the CMM is shown above. The columns represent the samples and rows represent the relative abundance of each family. Top 17 families which accounts for nearly 90% abundance of all the families are labeled in the heatmap.

genome and the phylogenetic relationship of the associated taxa are shown in Fig. 4. The 1-Mb chromosomal regions explained 0.32–3.24% (supplementary Table S5) of the phenotypic variation, where the largest variation was associated with OTU4 and family *Prevotellaceae*. Genomic associations were observed across diverse phylogenetic groups where associations were detected across 7 different phyla including the 4 major phyla that represent more than 90% of the rumen bacterial population (Fig. 4). The SNP associated with bacterial taxa were distributed across 7 autosomal chromosomes with chromosome 9 and 27 demonstrating associations at phylum, family and OTU level. Additionally, some taxa at OTU, family and phylum level were associated with 2 different SNP located on different chromosomes. As described previously<sup>21</sup>, this suggests the rumen microbiota is a heritable, polygenic trait.

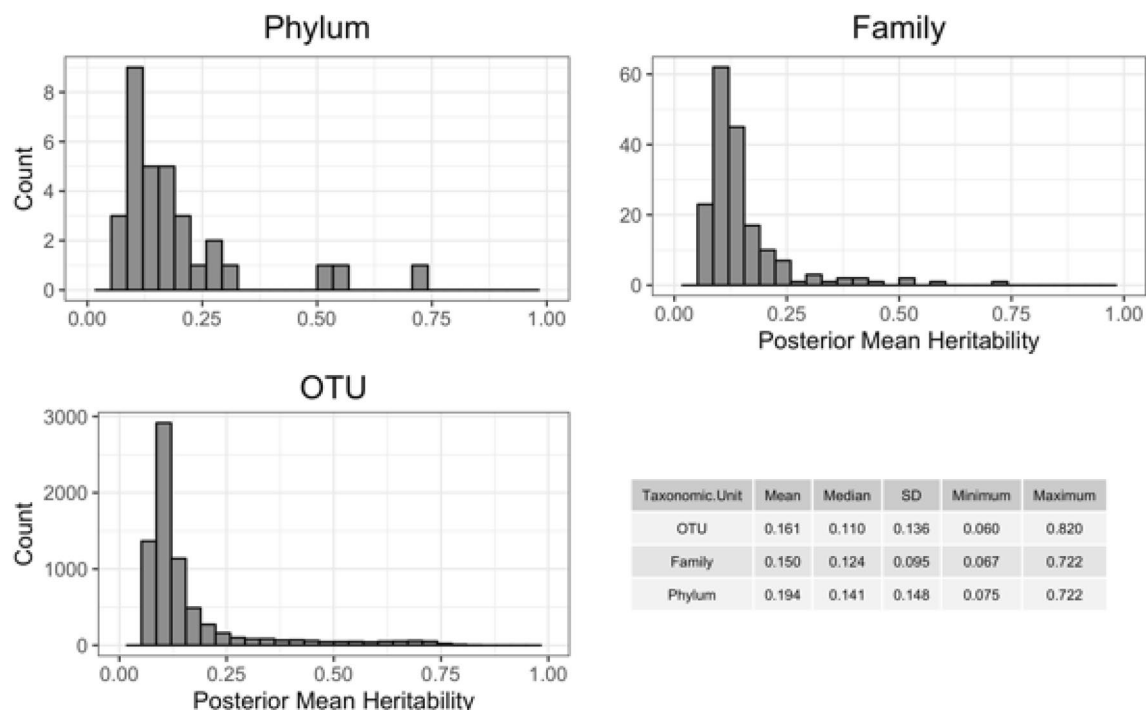
**Different taxonomic levels are under host genetic control.** The majority of the associations were identified at the OTU level. However, phylum and family level associations were also detected on chromosomes 9 and 27 in addition to OTU associations. As such, to determine if higher taxonomic level associations detected are a result of higher abundance of a single OTU, we performed correlation analysis between taxa on chromosomes 9 and 27 (Fig. 5). Correlation analysis demonstrated that some phylum level associations are driven by one family belonging to that phylum (phylum *Fibrobacter* and family *Fibrobacteriaceae*  $r = 1.0$ ; phylum *Proteobacteria* and family *Succinivibrionaceae*  $r = 0.999$ ; phylum *Verrucomicrobia* and family *RFP12*  $r = 0.955$ ). However, other phyla did not show such correlations with family level associations (*Firmicutes*, *Lentisphaerae*, and *Tenericutes*). A similar trend was seen between OTU level associations and family level associations where a few associations were highly correlated (OTU3 and family *RF16*  $r = 0.994$ , and OTU28 and family *Fibrobacteriaceae*



**Figure 2.** First and Second principle components of the genomic relationship matrix for USMARC and UNL animals demonstrating limited genomic variation in the two locations<sup>1</sup>. A genomic relationship matrix was constructed for all the animals sampled from USMARC and UNL using the SNP information. The principal component analysis (PCA) was run on genomic relationship matrix and first two principal components (PC1 and PC2) were plotted<sup>1</sup>. USMARC refers to animals that were fed at either the U.S. Meat Animal Research Center and UNL refers to animals fed at the University of Nebraska-Lincoln.

$r = 0.895$ ) (supplementary Figure S2). However, a majority of the associations detected at OTU level were unique and did not have correlated families.

**Host adapted bacterial species in the rumen.** Relatively large SNP associations were detected between genus *Methanobrevibacter*, *Succiniclacticum*, *Prevotella*, and *Fibrobacter*. Among the major genera associated with host genetics *Prevotella* was the most predominant genera. *Succiniclacticum* was significantly associated with both chromosome 1 and 2 similarly *Prevotella* was associated with chromosomes 2, 6, 9, 19, 23 and 27 (Table 1 and supplementary table S6). *Methanobrevibacter* was associated with chromosome 1 and *Fibrobacter* was associated with chromosome 27 (Table 1 and supplementary table S6). As described previously<sup>21</sup>, pleiotropic effects (having multiple effects from a single gene) were observed in many taxa. Including SNP that affect closely related taxa such as the associations observed on chromosome 9 for family *Veillonellaceae* and *Clostridiaceae*; and the association between *Prevotella* and *Paraprevotella* on chromosome 23. These observations suggest that genes in these 1 Mb regions can independently or in unity affect bacterial distribution and structure in the rumen. Additionally, the associated 1 Mb region on chromosome 27 is associated with the bacterial community composition of very diverse bacterial communities including phylum *Fibrobacteres*, *Lentisphaerae* and *Tenericutes*. As such, this genomic region can affect multiple diverse taxa within the rumen microbiome. Finally, we also detected genomic regions in multiple chromosomes that control the same bacterial genera. For example, OTU30 belonging to genus *Prevotella* was associated with 1 Mb regions in chromosomes 6, and 23. Additionally, OTU2337 belonging to genus *Succiniclacticum* was associated with 1 Mb regions in both chromosome 1 and 2. Similarly, OTU53 and OTU72 was associated with 1 Mb regions in chromosomes 6, 9, 23 and chromosomes 19, 23. OTU72 was associated with two different 1 Mb regions in chromosome 23.



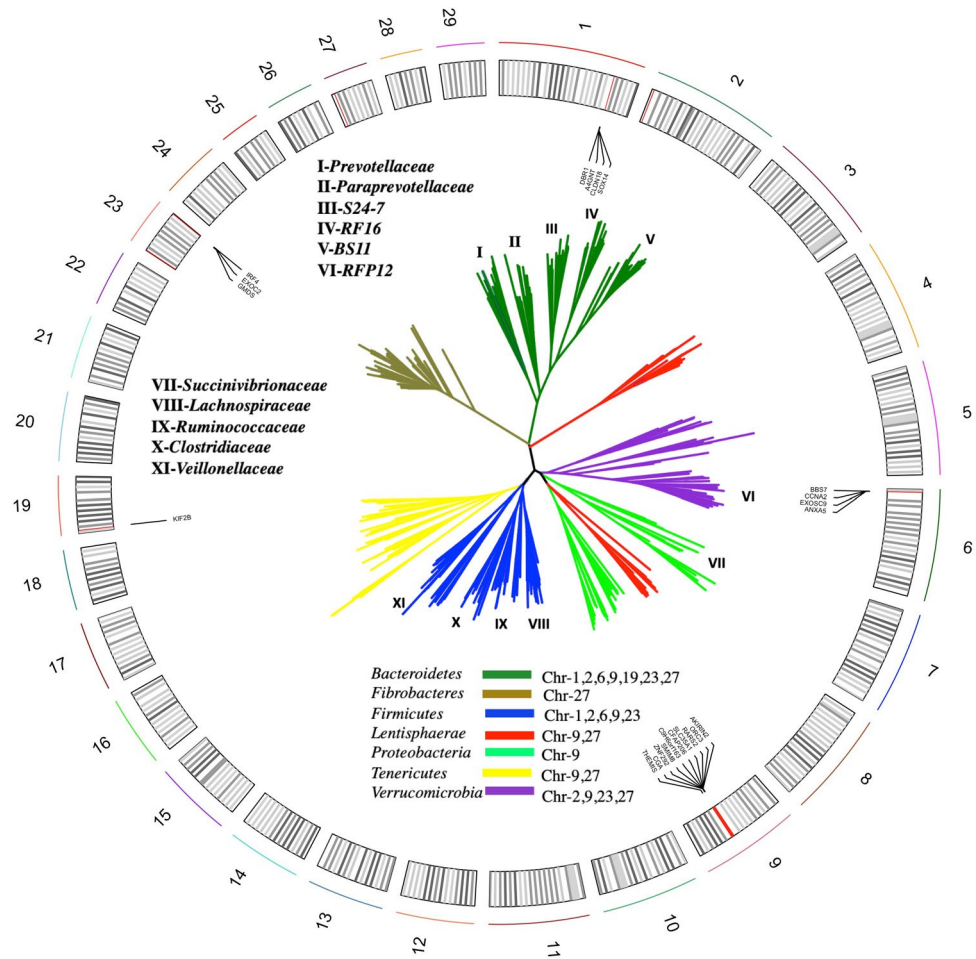
**Figure 3.** Posterior heritability estimates and summary statistics across taxa for phylum, family and operational taxonomic unit (OTU) categories. The relative abundances of Phyla, Families and OTUs were used to estimate the posterior heritability by using the GBLUP model (see “Methods”). The histograms represent the number of Phyla, Families and OTUs that fall into a specific range of posterior heritability. The table shows the mean, median, standard deviation and the minimum and maximum posterior heritability for different taxonomy levels.

Class	Chromosome	Position (Mb; start–end)	Names
OTU	1	132.0–133.0	41, 57, 2,337
	2	2.0–3.0	21, 2,337, 43
	6	3.0–4.0	27, 30, 53
	9	63.0–64.0	4, 12, 53
	19	3.0–4.0	2,737, 3, 72
	23	0.0–1.0	12, 27, 2,737, 30, 4, 53, 72
	23	51.0–52.5	21, 64, 72
	27	3.0–4.0	19, 28, 3
Family	6	3.0–4.0	<i>BS11</i> , <i>Ruminococcaceae</i> , <i>Succinivibrionaceae</i>
	9	63.0–64.0	<i>Lachnospiraceae</i> , <i>RFP12</i> , <i>Succinivibrionaceae</i> , <i>Veillonellaceae</i> , <i>Clostridiaceae</i>
	23	0.0–1.0	<i>Paraprevotellaceae</i> , <i>Prevotellaceae</i> , <i>RFP12</i> , <i>S24-7</i> , <i>Ruminococcaceae</i>
	27	3.0–4.0	<i>RF16</i> , <i>S24-7</i> , <i>Fibrobacteraceae</i> , <i>Clostridiaceae</i>
Phylum	9	63.0–67.0	<i>Firmicutes</i> , <i>Lentisphaerae</i> , <i>Proteobacteria</i> , <i>Verucomicrobia</i>
	27	3.0–4.0	<i>Fibrobacteres</i> , <i>Lentisphaerae</i> , <i>Tenericutes</i>

**Table 1.** Chromosomal regions with the largest Window Genomic Estimated Breeding Value (WGBEV) for the selected OTUs, families, and phyla.

## Discussion

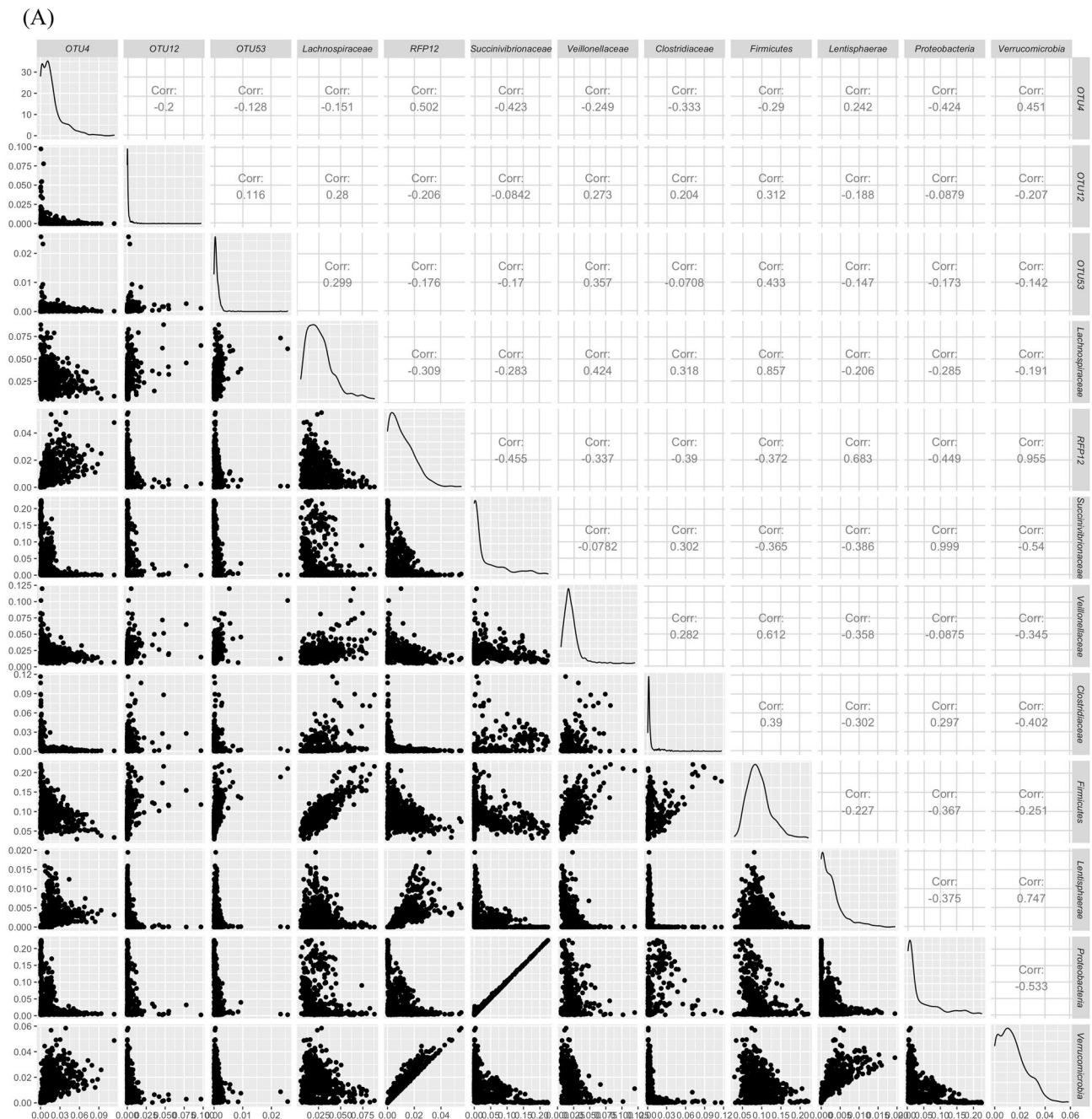
The rumen microbiome has recently begun to be explored<sup>1,25–27</sup>. Similar to the mammalian intestinal microbiota, the cattle rumen microbiota is dominated by *Bacteroidetes*, *Firmicutes*, and *Proteobacteria* (based on type of diet)<sup>14,15,25,26</sup>. However, cattle rumen microbiota is distinct from intestinal microbiota of monogastric animals because it has a considerable proportion of members of the Phylum *Fibrobacter*, which are involved in fiber digestion, and because the microbial communities in the rumen utilize the diet before the host digestive system.



**Figure 4.** SNP mapping of the rumen gut microbiota. The circular diagram depicts the 29 bovine autosomes drawn to scale. Each black line represents 3 Mb region of the chromosome that includes the position of the SNPs used for Genome Wide Association Study. Red lines represent 1 Mb regions (Table 1) that have associations with different bacterial taxa in the rumen. The list of identified genes when annotated in the 1 Mb window are listed close to the region. A representative phylogenetic tree was generated from the rumen bacterial reads using the Interactive tree of life (iTOL). Major phyla are color-coded and associated chromosome(s) are listed with each phyla. The roman numerals represent families as follows; I—*Prevotellaceae*, II—*Paraprevotellaceae*, III—*S24-7*, IV—*RF16*, V—*BS11*, VI—*RFP12*, VII—*Succinivibrionaceae*, VIII—*Lachnospiraceae*, IX—*Ruminococcaceae*, X—*Clostridiaceae*, XI—*Veillonellaceae*. The complete list of annotated genes and their position on each chromosome is listed in supplementary Table S1.

It is well established that diet is a major contributor of rumen microbial community composition<sup>1,14,16,17</sup>. However, in the current study cattle originated from two locations, whereby 2 of the 3 diets were only represented at one location. Furthermore, heifers only originated from one location and were fed a diet that differed from that fed to steers. This data structure, and the fact that samples were generated across multiple years, makes parsing these effects impossible. As such, to identify the collective effect of these factors on the rumen bacterial community, we performed PCoA analysis (supplementary figure S3) and PERMANOVA analysis using management type to reflect the collective effect diet, location, and sex. PERMANOVA analysis displayed a significant effect of management type on microbial community composition suggesting that collectively these factors affected the rumen microbial community composition. As the objective of this study was to identify association between host genetics and the rumen bacterial community composition, we fitted contemporary group, the concatenation of management type and year, in the model used for GWAS to adjust for the variation in the microbial community composition resulting from management type as described in methods. As such, the association observed in this study has been corrected for variations in the rumen bacterial community that can result from diet, sex, location and year.

Studies investigating the effect of host genetics on shaping the microbial community composition are limited. Recent studies in the bovine rumen have demonstrated that the rumen microbiome is influenced by host genetic factors<sup>5,6,21</sup>. As such, the concept of core taxa within the rumen microbiome being controlled by host genetics is intriguing as it lends to the potential to utilize associations between rumen microbiota and genetic markers for



**Figure 5.** Correlations between OTU, Phyla and Family abundance associated with chromosome 9 (A) and chromosome 27 (B). A matrix was generated using  $\log(1+x)$  transformed relative abundances for OTUs, Phyla and Families associated with each chromosome and pairwise Pearson correlations were calculated and scatter plots and density plots were generated for each chromosome. The pairwise correlations among all the identified taxa on all chromosomes are shown in supplementary figure S3 and further taxonomic information of all identified OTUs can be found in Supplementary Table S1.

marker-assisted selection and management to improve feed efficiency, animal health and microbiome manipulation mediated by selecting for favorable microbial taxa within the rumen.

In this study, we demonstrate that the relative abundance of the most abundant taxa within the rumen microbiome is a “polygenic trait”<sup>12</sup> under host genetic control. The pleiotropic effect of host genetics on the rumen microbiome demonstrates effects at multiple taxonomic levels. Data in this study demonstrates that the influence of host genetics in shaping the rumen microbiome is more effective at lower taxonomic units and in most cases the effect on the microbiome can be very specific. This is clearly demonstrated by the OTU level associations identified in this study. The association identified with the genus *Methanobrevibacter* on chromosome 1 is the first report of an association identified between an archaeal species in the rumen and the bovine host. We also identified family *Succinivibrionaceae*, which has been previously identified to be linked with methane emission<sup>28</sup>,

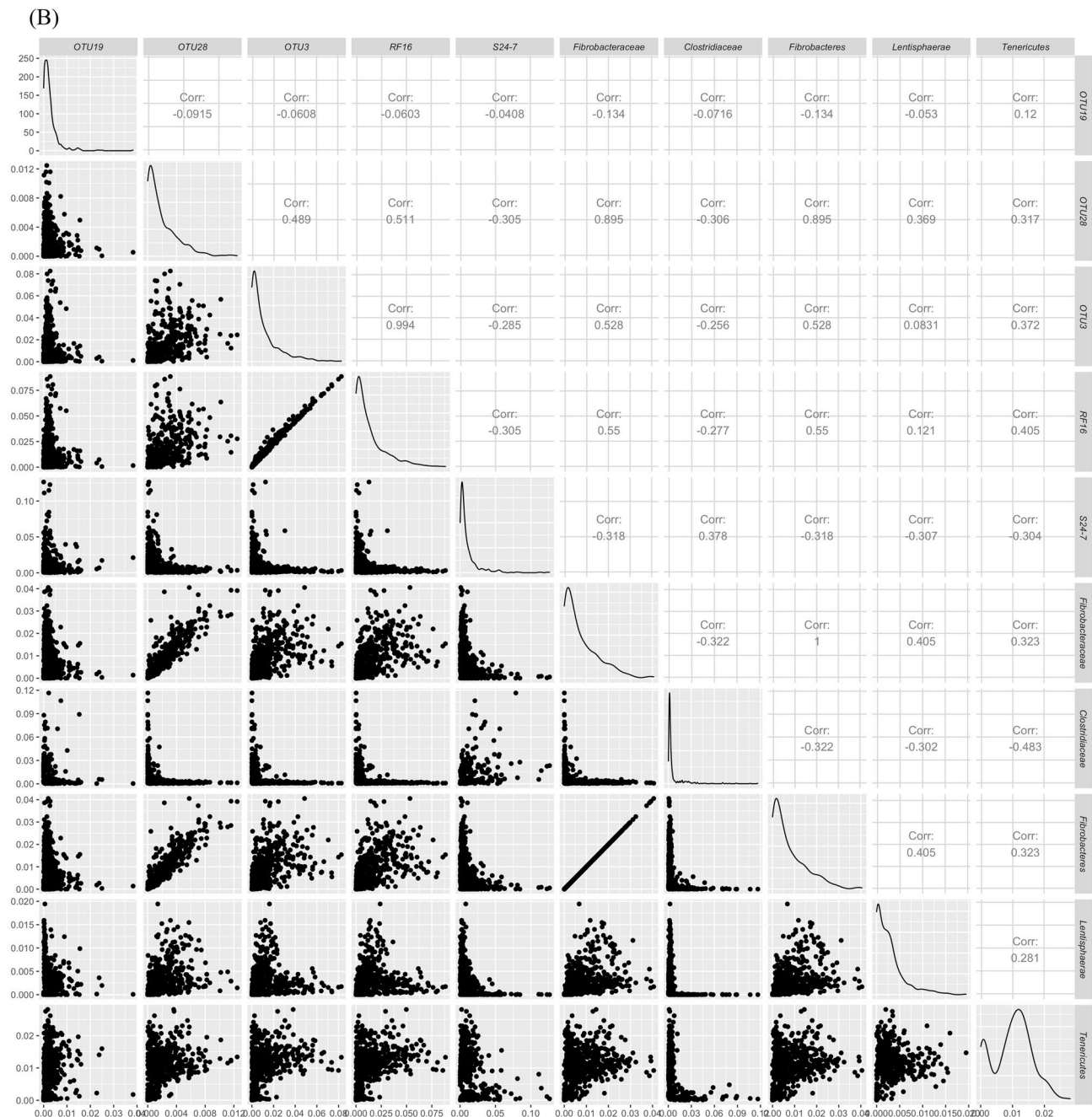


Figure 5. (continued)

associated with chromosome 9. A previous study by Wallace et al. identified *Succinovibrionaceae* to be heritable in dairy cows. Other studies have also reported host genetics and the microbiome to be associated with methane emission<sup>29</sup> and this association further supports this notion. In the rumen, archaea play a critical role in recycling NADH back to NAD for glycolysis<sup>2,30</sup>, so that pyruvate in the rumen is spared for VFA production which becomes an energy source for the host<sup>2,30</sup>. In the absence of methanogens, recycling of NADH is performed by conversion of pyruvate to lactate or ethanol<sup>2</sup>, as such the pyruvate available for VFA production is wasted to recycle NADH making the rumen ecosystem less efficient. Therefore, the association observed between the host and the methanogen is directed toward increased energy to the host animal.

We observed associations on chromosome 1 and 2 for genus *Succiniclasicum*. Only one species named *Succiniclasicum ruminis* has been described in this genus<sup>31</sup>. *Succiniclasicum ruminis* is a common inhabitant in the rumen specialized in its ability to convert succinate to propionate as its sole mechanism of energy production. In the rumen, succinate is not accumulated as it is rapidly converted to propionate, and *Succiniclasicum ruminis* is considered as the major organism that is involved in this process. As, propionate is the only gluconeogenic volatile fatty acid in the rumen and also provides more ATP to the host than any other VFA produced in the rumen, it is not surprising that the host animal would prefer selective enrichment of microbes that increase



energy supply to the host. As such the association between the host and the genus *Succiniclasticum* has important consequences towards host performance and well-being. The association observed for *Fibrobacter succinogenes* in chromosome 27 also provides evidence to support the notion that the host genome controls for microbes that help increase nutrient metabolism within the rumen. *Fibrobacter succinogenes* is characterized as one of the key cellulolytic microbes in the rumen<sup>32</sup>, that help breakdown cellulose in the rumen. Genome sequence of *Fibrobacter succinogenes* S85<sup>33</sup> has revealed that this organism contains 83 glycosyl hydrolases including 33 cellulases and 24 xylanases, 7 pectate lyases and 14 carbohydrate esterases<sup>33,34</sup>. As such *Fibrobacter succinogenes* is a key microbe in cellulose digestion in the rumen and the association with host genetics suggests that this organism is under genetic control to ensure efficient nutrient metabolism in the rumen and that the host animal's energy requirements can be met from metabolites of microbial fermentation. Wallace et al.<sup>21</sup> also reported the *Fibrobacter succinogenes* as one of the heritable bacteria in dairy cows.

We observed several loci on different chromosomes (2, 6, 9, 19, 23 and 27) to be associated with the genus *Prevotella*. *Prevotella* is a dominant genus found in the rumen and has been implicated in protein and energy metabolism in the rumen<sup>1,14</sup>. It is possible that each of the loci may be affecting the compositional changes of different species of *Prevotella* such as *P. ruminicola*, *P. bryanti*, *P. brevis*, and *P. albinensis* which are all been reported to be members of the rumen microbiome<sup>35</sup>. Although, we are unable to classify the OTUs identified at species level resolution, the fact that different OTUs belonging to genus *Prevotella* are associated with different loci suggests that different *Prevotella* species may be controlled by different loci. Recently, Li et al. reported 19 SNPs to be associated with 14 microbial taxa in the rumen<sup>5</sup>. This study also identified associations between *Paraprevotellaceae* and bovine chromosome 16, further supporting the notion of genus *Prevotella* may be controlled by many genetic loci. Similarly, family *Prevotellaceae* were reported as highly heritable among Nordic Red dairy cows<sup>21</sup>.

Describing factors that shape the rumen microbiome is important to improve animal health and performance and here we see a trend of the host selecting for bacterial species that help in nutrient metabolism in the rumen. To further investigate the associations between the taxa identified and the host, we performed "positional candidate gene analysis". Positional gene candidate analysis helped identify genes located within the 1 Mb regions associated with rumen bacterial species (supplementary table S1 and S5). Many of the 1 Mb chromosomal regions identified did not contain annotated genes within the bovine genome. The only well annotated region in the bovine genome was in chromosome 9 (supplementary table S1). The 1 Mb regions between 63 and 67 Mb in chromosome 9 demonstrated association at OTU (63–64 Mb), family (63–64 Mb) and phylum (63–67 Mb) level controlling the abundance of both gram-positive and negative bacteria in the rumen. Positional candidate gene analysis in this region identified genes required for innate immunity, specifically AKIRIN2, a downstream effector of the Toll-like receptor (TLR), TNF and IL-1 beta signaling pathways that results in IL-6 production<sup>36</sup>. As such, it is possible that the association of rumen bacteria to this region of the host chromosome may result in cytokine secretion leading to modulation of host immunity. A similar association was observed on chromosome 9, 66–67 Mb region with the Thymocyte-Expressed Molecule Involved In Selection (THEMIS) gene. This gene has been described to play a critical role in thymocyte development and maturation of T-cells<sup>37</sup> further implicating the interaction between the host genetics and the rumen microbiome in immune modulation.

Additionally, the rumen epithelium is characterized as a stratified squamous epithelium, and has been described as an organ involved in selective absorption of nutrients in the form of volatile fatty acids (VFAs) from rumen bacterial fermentation<sup>38</sup>. Therefore, active and secondary active transport systems mediate nutrient absorption<sup>38,39</sup>. Previous studies have demonstrated Claudin family genes to play a role in the formation of the permeability barrier and to help with tight junction formation in the rumen<sup>38</sup>. In our GWAS, we identified a significant association with claudin-18 (CLDN18) in chromosome 1 in the 132–133 Mb region. This region was associated with the specialized propionate producer *Succiniclasticum ruminis* that utilizes conversion of succinate to propionate as its sole energy producing reaction<sup>31</sup>. As such, it is tempting to speculate that this association between claudin-18 and *Succiniclasticum ruminis* is involved in selective VFA absorption to the host. In addition to the described associations above, associations with GDP-Mannose 4,6-Dehydratase (GMDS) in chromosome 23, Alpha-1,4-N-Acetylglucosaminyltransferase (A4GNT) in chromosome 1, and ANXA5 involved in endocytotic and exocytotic pathways in chromosome 6 all suggests implication of the host in modulating the rumen microbes to increase energy absorption. A majority of the SNPs identified by Li et al. were present in the non-coding region and thus they were unable to identify how the associated taxa may influence host metabolism or health. The few genes identified in that study suggested, the associated regions to increase nutrient absorption to the host. Results presented in this study is consistent with this notion and the associations observed between the host and the microbial species further suggests that the associations between the microbiota and host genotype is directed toward selective absorption of volatile fatty acids from the rumen to increase energy availability to the host animal.

In the ruminant animal, 50–70% of the animal's protein needs and up to 70% of the energy needs<sup>7</sup> are met through the metabolism of rumen microbes. The rumen microbiota can be viewed as an environmental factor that impacts animal health, nutrition and performance. Studies have demonstrated the rumen bacterial species composition to influence feed efficiency, Average Daily Gain (ADG) and intake<sup>40</sup>. As such demonstrating that heritable trait in the host genome can impact rumen bacterial species composition provides new opportunities to using genome selection to improve animal health and productivity. Future work evaluating the SNP identified herein and their relative effects in other populations as well as the recruitment of associated bacterial taxa in subsequent generations when parents are selected based on genotypes at these loci would be interesting. Such investigations may lead to the possibility of selection for microbiome manipulation. Additionally, similar associations may exist between the host genotype and the fungal, protozoal and viral populations within the rumen. Future studies focused on other members of the rumen microbiome may shed light into the role of host genetics shaping these microbial populations.

## Materials and methods

**Experimental design and animals.** All animal procedures implemented in this study were approved by the University of Nebraska—Lincoln and U.S. Meat Animal Research Center (USMARC) Animal Care and Use Committee. All experiments were performed using relevant guidelines and regulations described by the Animal Care and Use Committee. The data presented in this study were collected between 2009 and 2015 from different cohorts of animals which included a cohort of heifers ( $n=127$ ) and a cohort of steers ( $n=131$ ) from USMARC ( $n=258$ ) and a cohort of steers ( $n=328$ ) from University of Nebraska, Lincoln (UNL) research feedlot. Animals within USMARC and UNL were fed in 3 (USMARC1-3) and 5 (UNL1-5) cohort groups, respectively. The number of animals within each cohort ranged from 51 to 127 for animals fed at USMARC and 51 to 111 for animals fed at UNL. The animals from USMARC were part of the USMARC Germplasm Evaluation project (GPE)<sup>41</sup>. The animals from UNL were purchased from Nebraska and surrounding states and were a cross bred population of unknown breed makeup.

The Heifers at USMARC were fed a growing diet at time of sample collection that included 70% corn silage and 30% alfalfa hay on dry matter basis<sup>40</sup>. The USMARC steers were fed a finishing diet composed of 57.6% dry-rolled corn, 30% wet distillers grains with solubles, 8% alfalfa hay, and 4.4% vitamin and mineral supplement on a dry matter basis<sup>40</sup>. All UNL animals were fed a common basal diet (UNL-Diet) containing 50:50 blend of Alfalfa and Sweet Bran<sup>®</sup>. After adaptation to each diet for at least 21 days, rumen samples were collected via esophageal tubing for bacterial community analysis. The diets were formulated to meet or exceed NRC recommendations for growth and vitamin and mineral supplementation of growing and finishing beef cattle.

**Sample collection for rumen microbiota composition analysis.** Rumen samples were collected after adaptation to each diet via esophageal tubing as described previously<sup>42</sup>. Briefly, the animal was restrained in the chute, and a stomach tube was inserted through a speculum and passed through the esophagus until it reached the rumen. A vacuum pump was attached to the free end of the tube and the sample was withdrawn from the rumen. The samples collected contained both rumen fluid and feed particles and were a representative sample of the rumen. Samples collected via esophageal tubing have been shown to represent the rumen to contain a similar microbial community composition to a sample collected via a rumen fistula<sup>42</sup>. The samples collected were snap frozen in liquid nitrogen and were stored in  $-80^{\circ}\text{C}$  until used for DNA extraction and microbial community analysis.

**Microbial DNA extraction.** DNA was extracted from the rumen samples using the PowerMag Soil DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA) according to the manufacturer's protocol with a few modifications as described by Paz et al.<sup>40</sup>. The modified protocol also included adding RNase A to the lysis solution to ensure removal of RNA during DNA extraction. The isolated DNA was stored at  $-20^{\circ}\text{C}$  until used for bacterial community analysis.

The V4 region of the 16S rRNA gene was amplified and sequenced on the MiSeq platform as described previously<sup>40,43</sup>. Briefly, barcoded universal primers specific for the V4 region of eubacteria were amplified using 25  $\mu\text{L}$  PCR amplification reactions<sup>43</sup>. Each 25  $\mu\text{L}$  PCR reaction contained 0.75 Units Terra PCR Direct Polymerase Mix, 1X Terra PCR Direct Buffer, 10  $\mu\text{M}$  indexed fusion primers, and 5–20 ng of DNA. The cycling conditions contained 98  $^{\circ}\text{C}$  for 2 min, followed by 25 cycles of 98  $^{\circ}\text{C}$  for 10 s, 55  $^{\circ}\text{C}$  for 30 s, and 68  $^{\circ}\text{C}$  for 30 s; and a final elongation step of 68  $^{\circ}\text{C}$  for 4 min<sup>40</sup>. Following amplification, PCR products were normalized using the Sequel-Prep Normalization Plate Kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. The normalized libraries were pooled and were further purified using the MinElute PCR Purification Kit (Qiagen, Valencia, CA, USA) as described by the manufacturer. The resulting pooled sample was subjected to size selection and purification using the Pippin Prep (Sage Science, Inc., Beverly, MA, USA) automated size selection instrument. The resulting sequence ready libraries were further analyzed using the Agilent BioAnalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) and were subjected to 250 bp paired end sequencing using the Illumina Miseq System (Illumina, San Diego, CA, USA) using the V2 500 cycle sequencing kit as described by the manufacturer. Bridge amplification and reversible dye-terminator -based sequencing on the MiSeq was performed as described by the manufacturer. Raw sequences have been deposited at the NCBI Sequence Read Archive (SRA) under the accession no. SRP100776 and PRJNA55259.

**Data processing pipelines for microbial community analysis.** Raw reads generated from Illumina MiSeq sequencing were processed using the quality filter and analysis pipeline described by Paz et al.<sup>40</sup>. Following preliminary quality filtering and read processing, the resulting reads were used for microbial community analysis as described below. Complete information of the bioinformatics pipeline describing data analysis is available at <https://github.com/FernandoLab>. Briefly, forward and reverse reads were assembled to generate contigs of the V4 region and further quality filtering was performed on subsequent contigs to remove sequences with ambiguous bases, incorrect contig length, or incorrect assembly using MOTHUR v.1.38.1<sup>23</sup>. Following secondary quality filtering, subsequent reads were clustered into operational taxonomic units (OTUs) using the UPARSE pipeline (USEARCH v7.0.1090)<sup>44</sup> after dereplication, discarding singletons, clustering sequences into OTUs at 97% similarity, and removing chimeric sequences using UCHIME<sup>45</sup> as described previously<sup>40</sup>. Representative OTU sequences from each OTU were aligned against the SILVA reference alignment v123 to ensure the OTU reads came from the V4 region of the 16S rDNA gene. The resulting representative OTU sequences that fail to map to the V4 region were discarded. Taxonomy was assigned to each OTU using QIIME v.1.9.1<sup>46</sup> as described previously<sup>23</sup>. The Greengenes database (gg\_13\_8\_otus)<sup>24</sup> were used as the reference database for taxonomic assignment. OTUs classified as Cyanobacteria were filtered from the data set as cyanobacterial sequences may arise from 16S remnants present in the plant chloroplasts<sup>47</sup>. However, recent studies have reported Cyano-

bacterial orders such as YS2, SM1D11, and mle1-12 to be a new class of Cyanobacteria to be present in gut, soil and plants<sup>48–50</sup>. However, these orders were not identified within the cyanobacterial sequences that were removed. A rarefied (7,000 reads) OTU table was used to generate a Bray Curtis dissimilarity matrix that was used for Principal Coordinate Analysis (PCoA) within QIIME v.1.9.1 pipeline. The R package ggplot2<sup>51</sup> was used to generate the PCoA plot in R by using first two components of the Bray Curtis coordinates. The Bray Curtis dissimilarity matrix was used to perform the multivariate analysis of variance (PERMANOVA)<sup>52</sup> within R package vegan<sup>53</sup>. The complete information of the bioinformatics pipeline describing data analysis is available at <https://github.com/FernandoLab>.

**Genotyping the resource population.** All animals (n = 586) were genotyped with either a 770 K, 150 K, or 80 K SNP assay. Blood samples were collected onto blood collection cards and the collection cards were sent to Geneseek (Lincoln, NE) for subsequent DNA extraction and genotyping. The SNP in common across all panels were utilized in the analysis and included 61,974 SNP. Genotypes that were missing were replaced with the mean genotype at that locus across all genotyped individuals.

**Identifying “positional candidate genes”.** The 1 Mb windows that account for the greatest proportion of the genetic variance were used to identify candidate genes using the “positional candidate gene approach” using the *Bos taurus* build UMD\_3.1 assembly<sup>54</sup> and UCSC genome (ars-ucd1.2/*bos taurus*) browser. Due to the limited understanding of the *Bos taurus* genome compared to *Homo sapiens*, human orthologs of beef cattle “positional candidate genes” was identified using the BioMart data mining tool (68) and the Ensembl Genes (release 69). Human gene orthologs to bovine “positional candidate genes” was utilized to identify gene ontology terms, and pathways using the UCSC genome (ars-ucd1.2/*bos taurus*) browser and NCBI gene functions.

**Statistical analysis.** An OTU table generated for the 586 animals was used for subsequent analysis. Rarefaction analysis was performed using QIIME v.1.9.1<sup>46</sup> as described previously<sup>40</sup>. Prior to the analysis, operational taxonomic unit (OTU), family and phylum read counts were transformed into relative abundance.

The heritability of an OTU, family, or phylum class was estimated using a Bayesian genomic best linear unbiased prediction (GBLUP) model utilizing the BGLR package in R<sup>55</sup>. Within each OTU, family, and phylum the following model was fitted:

$$y = Xb + Zu + e,$$

where  $y$  is the proportional abundance,  $b$  is a vector of fixed effects,  $u$  is a vector of random additive genetic effects,  $e$  is a vector of random residuals and  $X$  and  $Z$  are incidence matrices relating observations to the fixed and random additive genetic effects, respectively. To determine the extent of and to correct for population structure, a principle component analysis (PCA) on a genomic relationship matrix ( $G$ ) was utilized. The  $G$  matrix was constructed as:

$$G = \frac{MM'}{2 \sum p_j(1 - p_j)},$$

where  $M$  is a genotype incidence matrix that has been centered based on allele frequencies (VanRaden, 2008) and  $p$  is the allele frequency of the second allele at the  $j$ th SNP across all loci.

The fixed effects included the intercept, the first two PC, and contemporary group (concatenation of management type and year).

The random additive genetic effect was assumed  $\sim N(0, G\sigma_u^2)$ , where  $G$  is a genomic relationship as outlined previously. A chain length of 102,000 iterations was run with the first 2000 discarded as burn-in and a thinning rate of 10 was utilized. The default priors were utilized within BGLR, which included a flat and bounded prior for the fixed effects and a scaled inverse Chi-squared distribution for the additive genetic and residual variances. Across both random effects the degrees of freedom were set at their default value of 5 and the scale factors were set based on the rules described by de los Campos and Pérez-Rodríguez<sup>56</sup> with the assumption that the model explains 20% of the phenotypic variance. The posterior mean  $\pm$ PSD heritability estimates for a given OTU, family, or phylum was calculated as the mean of  $\frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$  across all samples that remained after thinning. The number of effective samples across models was estimated using the CODA package in R<sup>57</sup>.

**Genome-wide association study.** Across OTU, family, or phylum class the estimated additive genetic effects across animals were decomposed into marker effects for a subset of groups within each class<sup>58</sup>. The classes investigated are outlined in Table S1. The marker effects ( $a$ ) were estimated by backsolving using  $G$  and the incidence matrix (i.e.,  $Z$ ) as outlined below:

$$\hat{a} = \frac{M'G^{-1}\hat{u}}{2 \sum p_j(1 - p_j)},$$

where  $M$  is the genotype incidence matrix as outlined previously,  $G^{-1}$  is the inverse of  $G$  and  $\hat{u}$  is the estimated additive genetic value of an individual which was derived from the Bayesian GBLUP model. After estimating marker effect solutions, the variance of 1 Mb non-overlapping window genomic estimated breeding values (WGEBV) was computed for each window. Within a window the WGEBV for each individual was estimated by multiplying the estimated SNP effects with their respective genotypes and summing across all SNP within the

window. The WGEV variance of a 1-Mb window was calculated as the variance in WGEV across individuals. All OTUs ( $n = 17$ ) with at least 40,000 reads and present in at least 95% of the animals, all families ( $n = 12$ ) that represented greater than 1.8% total reads and were the top families (12 families were selected as families ranked 11th and 12th had very similar read percentages to the family ranked 10th) and top 10 Phyla ( $n = 10$ ) with the largest variance in WGEV were investigated further. The relative abundances of the identified OTUs, families and phyla were transformed with natural log ( $\log(1 + x)$ ) and then pairwise Pearson correlation analyses were performed. The pairwise correlation among OTUs, families and Phyla associated with chromosome 9 and 27 were plotted using R package GGally<sup>59</sup> while correlation heatmap among all the identified OTUs, families and phyla was generated by using R package ggplot2<sup>51</sup>.

### Data availability

Raw sequences generated in this study have been deposited at the NCBI Sequence Read Archive (SRA) under the accession no. SRP100776 and PRJNA55259. The complete information of the bioinformatics pipeline describing data analysis is available at <https://github.com/FernandoLab>.

Received: 21 January 2020; Accepted: 19 August 2020

Published online: 15 September 2020

### References

- Anderson, C. L., Schneider, C. J., Erickson, G. E., MacDonald, J. C. & Fernando, S. C. Rumen bacterial communities can be acclimated faster to high concentrate diets than currently implemented feedlot programs. *J. Appl. Microbiol.* **120**, 588–599. <https://doi.org/10.1111/jam.13039> (2016).
- Church, D. C. *The Ruminant Animal Digestive Physiology and Nutrition* (Waveland Press Inc, Long Grove, 1993).
- Hobson, P. N. *The Rumen Microbial Ecosystem* (Elsevier Applied Science, London, 1988).
- Hungate, R. E. *The Rumen and Its Microbes* (Academic Press Inc, London, 1966).
- Li, F. *et al.* Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. *Microbiome* **7**, 92. <https://doi.org/10.1186/s40168-019-0699-1> (2019).
- Sasson, G. *et al.* Heritable bovine rumen bacteria are phylogenetically Related and Correlated with the Cow's Capacity To Harvest Energy from Its Feed. *mBio* <https://doi.org/10.1128/mBio.00703-17> (2017).
- Flint, H. J. Polysaccharide breakdown by anaerobic microorganisms inhabiting the mammalian gut. *Advances in Applied Microbiology, Vol 56* **56**, 89–+ (2004).
- Jami, E., Israel, A., Kotser, A. & Mizrahi, I. Exploring the bovine rumen bacterial community from birth to adulthood. *ISME J.* **7**, 1069–1079. <https://doi.org/10.1038/ismej.2013.2> (2013).
- Furman, O. *et al.* Stochasticity constrained by deterministic effects of diet and age drive rumen microbiome assembly dynamics. *Nat. Commun.* **11**, 13 (2020).
- Rey, M. *et al.* Establishment of ruminal bacterial community in dairy calves from birth to weaning is sequential. *J. Appl. Microbiol.* **116**, 245–257. <https://doi.org/10.1111/jam.12405> (2014).
- Wang, L. *et al.* Dynamics and stabilization of the rumen microbiome in yearling Tibetan sheep. *Sci. Rep.* <https://doi.org/10.1038/s41598-019-56206-3> (2019).
- Benson, A. K. *et al.* Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc. Natl. Acad. Sci. USA.* **107**, 18933–18938. <https://doi.org/10.1073/pnas.1007028107> (2010).
- Spor, A., Koren, O. & Ley, R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat. Rev. Microbiol.* **9**, 279–290. <https://doi.org/10.1038/nrmicro2540> (2011).
- Fernando, S. C. *et al.* Rumen microbial population dynamics during adaptation to a high-grain diet. *Appl. Environ. Microbiol.* **76**, 7482–7490. <https://doi.org/10.1128/Aem.00388-10> (2010).
- Fernando, S. C. *et al.* Meta-functional genomics of the rumen biome. *J. Anim. Sci.* **85**, 569–569 (2007).
- Henderson, G. *et al.* Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci. Rep.* <https://doi.org/10.1038/srep14567> (2015).
- Huws, S. A. *et al.* Addressing global ruminant agricultural challenges through understanding the rumen microbiome: Past, present, and future. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2018.02161> (2018).
- Mizrahi, I. & Jami, E. Review: The compositional variation of the rumen microbiome and its effect on host performance and methane emission. *Anim. Int. J. Anim. Biosci.* **12**, s220–s232. <https://doi.org/10.1017/S1751731118001957> (2018).
- Jiao, J. Z. *et al.* Shifts in host mucosal innate immune function are associated with ruminal microbial succession in supplemental feeding and grazing goats at different ages. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2017.01655> (2017).
- Glass, E. J., Baxter, R., Leach, R. J. & Jann, O. C. Genes controlling vaccine responses and disease resistance to respiratory viral pathogens in cattle. *Vet. Immunol. Immunopathol.* **148**, 90–99. <https://doi.org/10.1016/j.vetimm.2011.05.009> (2012).
- Wallace, R. J. *et al.* A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. *Sci. Adv.* <https://doi.org/10.1126/sciadv.aav8391> (2019).
- Hernandez-Sanabria, E. *et al.* Influence of sire breed on the interplay among rumen microbial populations inhabiting the rumen liquid of the progeny in beef cattle. *PLoS ONE* **8**, e58461. <https://doi.org/10.1371/journal.pone.0058461> (2013).
- Schloss, P. D. *et al.* Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541. <https://doi.org/10.1128/Aem.01541-09> (2009).
- McDonald, D. *et al.* An improved greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* **6**, 610–618. <https://doi.org/10.1038/ismej.2011.139> (2012).
- Brulc, J. M. *et al.* Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl. Acad. Sci. USA.* **106**, 1948–1953. <https://doi.org/10.1073/Pnas.0806191105> (2009).
- Kim, M., McSweeney, C. S., Morrison, M. & Yu, Z. An in silico generated census of the rumen bacterial microbiome. *Microb. Ecol.* **57**, 574–574 (2009).
- Anderson, C. L., Sullivan, M. B. & Fernando, S. C. Dietary energy drives the dynamic response of bovine rumen viral communities. *Microbiome* <https://doi.org/10.1186/s40168-017-0374-3> (2017).
- Wallace, R. J. *et al.* The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* **16**, 839. <https://doi.org/10.1186/s12864-015-2032-0> (2015).
- Difford, G. F. *et al.* Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. *PLoS Genet.* **14**, e1007580. <https://doi.org/10.1371/journal.pgen.1007580> (2018).
- Geier, R. R., Kwon, I. H., Cann, I. K. & Mackie, R. I. Interspecies hydrogen transfer and its effects on global transcript abundance in *Ruminococcus albus*, a predominant fiber-degrading species in the rumen. *FASEB J.* **30**, 1102 (2016).

31. van Gylswyk, N. O. *Succiniclasticum ruminis* gen. nov., sp. Nov., a ruminal bacterium converting succinate to propionate as the sole energy-yielding mechanism. *Int. J. Syst. Bacteriol.* **45**, 297–300. <https://doi.org/10.1099/00207713-45-2-297> (1995).
32. Wu, C. W. *et al.* Generation and characterization of acid tolerant *Fibrobacter succinogenes* S85. *Sci. Rep.* **7**, 2277. <https://doi.org/10.1038/s41598-017-02628-w> (2017).
33. Suen, G. *et al.* The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. *PLoS ONE* **6**, e18814. <https://doi.org/10.1371/journal.pone.0018814> (2011).
34. Jun, H. S., Qi, M., Ha, J. K. & Forsberg, C. W. *Fibrobacter succinogenes*, a dominant fibrolytic ruminal bacterium: Transition to the post genomic era. *Asian Austral. J. Anim.* **20**, 802–810. <https://doi.org/10.5713/ajas.2007.802> (2007).
35. Avgustin, G., Ramsak, A. & Peterka, M. Systematics and evolution of ruminal species of the genus *Prevotella*. *Folia Microbiol.* **46**, 40–44 (2001).
36. Stelzer, G. *et al.* The GeneCards Suite: From gene data mining to disease genome sequence analyses. *Curr. Protocols Bioinform.* **54**, 1–30. <https://doi.org/10.1002/cpbi.5> (2016).
37. Leavy, O. Thymocyte development: The identification of THEMIS. *Nat. Rev. Immunol.* **9**, 604–604. <https://doi.org/10.1038/nri2624> (2009).
38. Graham, C. & Simmons, N. L. Functional organization of the bovine rumen epithelium. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **288**, R173–181. <https://doi.org/10.1152/ajpregu.00425.2004> (2005).
39. Sehested, J., Diernaes, L., Moller, P. D. & Skadhauge, E. Transport of sodium across the isolated bovine rumen epithelium: Interaction with short-chain fatty acids, chloride and bicarbonate. *Exp. Physiol.* **81**, 79–94 (1996).
40. Paz, H. A. *et al.* Rumen bacterial community structure impacts feed efficiency in beef cattle. *J. Anim. Sci.* **96**, 1045–1058. <https://doi.org/10.1093/jas/skx081> (2018).
41. Schiermister, L. N., Thallman, R. M., Kuehn, L. A., Kachman, S. D. & Spangler, M. L. Estimation of breed-specific heterosis effects for birth, weaning, and yearling weight in cattle. *J. Anim. Sci.* **93**, 46–52. <https://doi.org/10.2527/jas.2014-8493> (2015).
42. Paz, H. A., Anderson, C. L., Muller, M. J., Kononoff, P. J. & Fernando, S. C. 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. <https://doi.org/10.3389/fmicb.2016.01206> (2016).
43. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120. <https://doi.org/10.1128/AEM.01043-13> (2013).
44. Edgar, R. C. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998. <https://doi.org/10.1038/nmeth.2604> (2013).
45. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200. <https://doi.org/10.1093/bioinformatics/btr381> (2011).
46. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336. <https://doi.org/10.1038/Nmeth.F303> (2010).
47. Giovannoni, S. J. *et al.* Evolutionary relationships among cyanobacteria and green chloroplasts. *J. Bacteriol.* **170**, 3584–3592. <https://doi.org/10.1128/jb.170.8.3584-3592.1988> (1988).
48. Soo, R. M. *et al.* An expanded genomic representation of the phylum Cyanobacteria. *Genome Biol. Evol.* **6**, 1031–1045. <https://doi.org/10.1093/gbe/evu073> (2014).
49. Di Rienzi, S. C. *et al.* The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *Elife* <https://doi.org/10.7554/eLife.01102> (2013).
50. McGorum, B. C. *et al.* Grazing livestock are exposed to terrestrial cyanobacteria. *Vet. Res.* <https://doi.org/10.1186/s13567-015-0143-x> (2015).
51. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York, 2009).
52. Anderson, M. J. & Walsh, D. C. I. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing?. *Ecol. Monogr.* **83**, 557–574. <https://doi.org/10.1890/12-2010.1> (2013).
53. Oksanen, J. *et al.* vegan: Community Ecology Package v. R package version 2.5–3. (2018).
54. Zimin, A. V. *et al.* A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* **10**, R42. <https://doi.org/10.1186/gb-2009-10-4-r42> (2009).
55. Pérez, P. & de los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **198**, 483–495 (2014).
56. de los Campos, G. & Perez-Rodriguez, P. Bayesian Generalized Linear Regression v. R package version 1.0.4. (2014).
57. Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: Convergence diagnosis and output analysis for MCMC. *R News* **6**, 7–11 (2006).
58. Wang, H., Misztal, I., Aguilar, I., Legarra, A. & Muir, W. M. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res.* **94**, 73–83. <https://doi.org/10.1017/S0016672312000274> (2012).
59. Schloerke, B.C.J. *et al.* GGally: Extension to 'ggplot2' v. R package version 1.4.0. (2018).

## Acknowledgements

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. USDA is an equal opportunity provider and employer.

## Author contributions

J.W., L.K., M.S. and S.F. designed the research project. K.H., H.P. and G.E. helped collect data for the study. H.P. performed laboratory analysis. H.P., W.A. and J.H. performed bioinformatic and GWAS analysis. W.A., M.S., J.H. and S.F. prepared the manuscript. M.S. and S.F. are co-corresponding authors of this manuscript and J.W. and W.A. are co-first authors of this manuscript.

## Funding

This work is supported by Animal Nutrition, Growth and Lactation grant no. 2018-67015-27496, Effective Mitigation Strategies for Antimicrobial Resistance grant no. 2018-68003-27545, and Multi-state research project accession no. 1000579 from the USDA National Institute of Food and Agriculture awarded to SCF.

## Competing interests

Samodha C. Fernando, author of this publication has disclosed a significant financial interest in NuGUT LLC. In accordance with its Conflict of Interest policy, the University of Nebraska-Lincoln's Conflict of Interest in Research Committee has determined that this must be disclosed. The rest of the authors have nothing to disclose.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-72011-9>.

**Correspondence** and requests for materials should be addressed to M.L.S. or S.C.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020