Differences in the microbiome of the small intestine of Leghorn lines divergently selected for antibody titer to sheep erythrocytes suggest roles for commensals in host humoral response

Shelly J. Nolin[1]*, Paul B. Siegel[2] and Christopher M. Ashwell[3]

[1]Prestage Department of Poultry Science, North Carolina State University, Raleigh, NC, United States, [2]School of Animal Science, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, [3]Davis College of Agriculture, Natural Resources, and Design, West Virginia University, Morgantown, WV, United States

For forty generations, two lines of White Leghorn chickens have been selected for high (HAS) or low (LAS) antibody response to a low dose injection of sheep red blood cells (SRBCs). Their gut is home to billons of microorganisms and the largest number of immune cells in the body; therefore, the objective of this experiment was to gain understanding of the ways the microbiome may influence the differential antibody response observed in these lines. We achieved this by characterizing the small intestinal microbiome of HAS and LAS chickens, determining their functional microbiome profiles, and by using machine learning to identify microbes which best differentiate HAS from LAS and associating the abundance of those microbes with host gene expression. Microbiome sequencing revealed greater diversity in LAS but statistically higher abundance of several strains, particularly those of *Lactobacillus,* in HAS. Enrichment of microbial metabolites implicated in immune response such as lactic acid, short chain fatty acids, amino acids, and vitamins were different between HAS and LAS. The abundance of several microbial strains corresponds to enriched host gene expression pathways related to immune response. These data provide a compelling argument that the microbiome is both likely affected by host divergent genetic selection and that it exerts influence on host antibody response by various mechanisms.

KEYWORDS

microbiome, antibody response, VA tech HAS and LAS, machine learning, immunogenetics

Introduction

Gut microbes are integral to intestinal physiology. Necessary for maintaining homeostatic balance between roles of nutrient absorption and pathogen response; resident microbes train the intestine to tolerate commensals while recognizing and responding to pathogens (Shulzhenko et al., 2011; Oakley et al., 2014). The accessibility

of next-generation tools, such as 16S ribosomal RNA sequencing, has facilitated better characterization of chicken microbiomes and correlated associations between commensal microbes and bird health and physiology (Benson et al., 2010; Oakley et al., 2014; Al-Marzooqi et al., 2020). Research is emerging which further connects host humoral response to vaccines and the commensal microbiota (Zimmermann and Curtis, 2018; Gonçalves et al., 2021; Jordan, Carding, and Hall, 2022).

Lines of White Leghorns have been divergently selected for high or low antibody response 5 days post injection with intravenous sheep red blood cells (Siegel and Gross, 1980). These lines, HAS and LAS, have been under constant divergent selection for over 40 generations and are valuable models for avian immunology and genetics, however the majority of the phenotypic differences remain unexplained (Gehad et al., 1999; Dorshorst, Siegel, and Ashwell, 2011; Lillie et al., 2017). Fecal 16S sequencing of HAS and LAS, along with associated lines where selection was relaxed, HAR and LAR, show selection associated differences in microbial abundance (L. Yang et al., 2017). Understanding how selection affects changes in microbiome composition and abundance is important in exploring the host microbe relationship. Equally important however, is unravelling ways in which the resident microbes may influence host phenotype.

In addition to simply characterizing the resident microbes, there is much interest in functional analysis to better explain host-microbe interactions. For amplicon-based sequencing, functional analysis is inferentially determined using software tools such as Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) (Douglas, Beiko, and Langille, 2018). Functional profile inference provides more information to better understand what molecules may be involved in host-microbe interactions, however this approach has obvious limitations both in terms of incomplete information due to dependence on previously sequenced microbial genomes as well as understanding how microbial gene products may affect host physiology. More integrative approaches are necessary to make accurate associations between microbes and host molecular responses.

Machine learning is revolutionizing omics data analysis (Lin and Lane, 2017; Leite et al., 2018; Oh et al., 2021). Predictive algorithms make it possible to identify patterns in large data sets and make relevant associations. This is useful in microbiome data analysis because the abundance of specific microbes can identify microbial "signatures" specific to, and able to differentiate between, two or more groups. This analysis can complement conventional methods to define which microbes best characterize a phenotypic group. Additionally, once microbial signatures are identified, that information can be used as the "classifier" for subsequent machine learning with other data. We have RNA sequencing data for the jejuna segments from which the microbes for 16S sequencing were obtained (Nolin Shelly et al., 2023). Machine learning utilizing microbial signatures with gene expression data, enables the identification of patterns of gene expression which best predict microbe abundance. By determining what molecular pathways are altered by genes differentially expressed in chickens with specific microbial signatures we are then able to make associations of host gene expression with commensal microbiomes. The integration of microbial data, with genetic line and gene expression data becomes a novel means to further explain

how microbes may influence underlying host physiology and in turn host phenotype.

Materials and methods

Animal work

Eggs from the 40th generations of lines HAS and LAS were obtained from Virginia Polytechnic Institute and State University and co-incubated until hatch. At hatch, all chicks were tagged for line identification and randomly transferred to battery cages such that chickens from both lines were raised together for the duration of the experiment. At 46 days of age, six chickens from each line were randomly selected, euthanized, and intestinal content samples collected from the duodenum, jejunum, and ileum for microbial DNA isolation. Corresponding intestinal tissue samples were collected for RNA isolation. This time point was selected because it is the traditional age of selection and the age corresponding to that of a cohort of chickens which had been injected with SRBC 5 days prior. All animal research was done in accordance with the North Carolina State University Institutional Animal Care and Use Committee.

Nucleic acid isolation, quantitation, and sequencing

Microbial DNA was isolated from intestinal contents using the QiaAmp DNA Stool Mini Kit and RNA was isolated from each jejunum sample using the RNEasy mini kit (Qiagen, Hilden, Germany) using the manufacturer's protocols. RNA and microbial DNA quantity were measured, and purity was assessed using the NanoDrop ND2000 spectrophotometer (Thermo Fisher Scientific, Waltham MA). Five hundred nanograms of each sample for 16S rRNA sequencing targeting the V3-V4 hypervariable region were sent to the University of North Carolina Microbiome center for library preparation, barcoding, and pyrosequencing on the Roche 454 sequencer. Two micrograms of RNA from each sample were taken to the North Carolina State University Genomics Sciences Laboratory for library preparation and sequencing on the Illumina HiSeq 2,500 sequencer.

Sequencing data analysis

CLC Genomics Workbench (Qiagen, Hilden, Germany) was used for all sequencing and statistical analyses. Statistical comparisons use a generalized linear model and calculate a p-value as well as false discovery rate p-value (FDRp) and Bonferroni correction to correct for multiple testing.

16S amplicon sequencing data was analyzed using the CLC Microbial Genomics Module. Sequence reads passing quality control were grouped into operational taxonomic units (OTUs) based on 97% sequence similarity to reference database SILVA SSU 99 version 138.1 (Pruesse et al., 2007; Quast et al., 2012; Yilmaz et al., 2014; Glöckner et al., 2017) and representative sequences were selected from each OTU for taxonomic assignment. Metrics were calculated

for alpha (within line) diversity using total number of OTUs and beta (between line) diversity using weighted unifrac. Differences in OTU abundance between HAS and LAS were considered significant at $p \leq 0.05$. Additionally, weighted unifrac permutational multivariate analysis of variance (PERMANOVA) was run to investigate line by tissue differences and were also considered significant at $p \leq 0.05$.

The microbial genomics module also contains a tool for inferring functional microbial profiles utilizing PICRUSt2 (Douglas et al., 2020). By importing a table containing Kmer frequency profiles with term multipliers and associated 16S copy numbers, the software can produce approximate functional profiles for mapped OTUs. This tool was used to create functional profiles as represented by KEGG (Kyoto encyclopedia of genes and genomes) (M. Kanehisa and Goto, 2000; Kanehisa, 2019; Kanehisa et al., 2021) identifiers present for HAS and LAS. Differential abundance of microbial molecules was calculated using the same GLM tool for determining OTU abundance. Due to the large number of functional molecules as well as the inferential nature of the data, the most conservative cut-off, Bonferroni correction ≤0.05 was considered significant. Differentially abundant KEGG molecules were entered into the KEGG orthology database and analyzed for module enrichment between lines.

CLC Genomics Workbench version 11 (Qiagen, Hilden, Germany) was used for RNA sequencing analysis as described previously (Nolin Shelly et al., 2023). High quality RNA sequencing reads were mapped to the Galgal6 reference genome (GCA_000002315.5) and differentially expressed genes were determined between lines, with FDRp ≤0.05 considered statistically significant.

Machine learning

Waikato Environment for Knowledge Analysis suite (WEKA) (Hall et al., 2009) was used for Machine learning. Three algorithms were used: support vector machines, artificial neural networks, and a decision tree. Algorithms were validated using two cross validation methods, a %-split and a K-fold stratified hold-out. A % split cross validation randomly splits the data into training or test. In our experiment we used a 66%–34% split, wherein 66% of the data was assigned to be the training set and the remaining 34% used to test. Sample groups are equally represented in the test and training sets. The K-fold stratified hold-out works by randomly dividing the data into K datasets, where K-1 datasets are used for training with the remaining 1 "hold-out" dataset as the test set. For our experiment, 6 is the number of biological replicates for each line/tissue, so K = 6. This algorithm runs sequentially 6 times, such that each dataset is the test set only once and included in the training set the remainder; performance is given as the average for the 6 runs. Performance for all machine learning is calculated as the average for the three algorithms and two validation methods.

To utilize machine learning to associate microbes with genetic line and later host gene expression, first we ran the algorithms to determine which jejunal microbes were most predictive of line. We started by running the algorithms with all OTUs to determine the predictive performance using the entire dataset. Next the OTUs were ranked by combining the entropy based InfoGainAttribute ranker function in WEKA (Li et al., 2004) with the p-value which was calculated for differential abundance for each OTU. A portion of OTUs were then removed (those with the highest p-value and lowest entropy) and the algorithms were re-run with the smaller set of OTUs. This process was repeated until the optimal microbial signatures were determined for highest prediction performance. This is known as reduction of data dimensionality and allows for identifying the optimal microbial "signature" for line prediction. Once these microbes were identified, they were each used as the sole attribute for running the algorithms to determine their individual predictive performance. Microbes with <75% correct prediction performance were eliminated from further analysis, as were those exclusively present in only one line.

These remaining OTUs were divided into groups by their relative abundance as high, low, or absent, or if the microbe was present in all samples as high, medium or low, as shown in Table 1. Abundance could then be used as the classifier and each sample, previously identified as HAS# or LAS#, was now identified by relative OTU abundance. OTU counts greater than the mean (for all present samples) were classified as high and those less than the mean as low, samples which did not contain any OTU counts were obviously grouped as absent. For microbes which were present in all samples, high was calculated as the mean OTU counts +50%, low as mean −50%, and medium were counts in between. Dividing the abundances for the microbes into three groups was necessary to avoid samples being classified solely by line. RNA sequence data for each sample was then used as the attributes with microbial abundance as the classifiers for machine learning in order to determine the gene expression patterns which predict microbe abundance. As statistical analysis of differential gene expression could not be performed based on microbial abundance, the InfoGainAttribute ranker function in WEKA (Li et al., 2004) was used to rank the genes for purposes of reducing data dimensionality as before. The optimal gene lists for each microbe were run in Ingenuity Pathway Analysis (IPA, Qiagen, Hilden, Germany) to determine enriched host pathways.

Negative controls for machine learning were also run for line prediction and microbial abundance. Datasets were randomized, wherein each sample identifier was randomly assigned to the data for a different sample, this allows for disassociation between classifier and attributes. Ten random datasets were used for each comparison and the average performance was calculated for each of the ten runs. Given two groups, HAS *versus* LAS 50% correct prediction would indicate random chance, whereas for microbial abundance with three classifiers, it would be ~33%. If the negative controls performance is close to that of random probability, we can be confident that true machine learning occurred in our experimental data sets.

Results

Microbial diversity and abundance

A total of 2,543 OTUs were identified, however those with less than 10 total OTU counts were filtered out, leaving 450 OTUs remaining for further analysis. LAS had the slightly more diverse microbiota of the two lines, with 350 OTUs present *versus* 304 OTUs

TABLE 1 Explanation of how samples were identified by microbial abundance as, absent (A), low (L), medium (M), or high (H).

OTU	HAS1	HAS2	HAS3	HAS4	HAS5	HAS6	LAS1	LAS2	LAS3	LAS4	LAS5	LAS6
1	A	L	A	A	A	A	H	H	H	L	L	H
2	M	H	H	H	M	M	L	L	L	L	L	L
3	M	H	H	H	M	M	M	L	L	L	L	L
4	M	H	H	H	M	M	M	M	L	L	M	M
5	L	A	A	A	L	L	H	H	L	A	L	L
6	L	H	H	H	H	M	M	L	L	L	M	L
7	L	H	L	H	H	L	L	A	L	A	L	L
8	H	L	H	H	H	L	A	L	L	A	A	L
9	L	L	H	H	H	A	A	A	L	L	L	A

present in HAS. There was overlap of 204 OTUs common to both lines, leaving a substantial percentage of the microbiota unique to one line or the other, see Figures 1A–F. As previously reported (Yang et al., 2017), *firmicutes* were the most abundant phyla in both lines, as well as for all intestinal segments, accounting for >80% of all microbes in our samples, followed by *actinobacteria*, *cyanobacteria*, and *proteobacteria* as shown in Figure 2. Within the *firmicutes*, lactic acid bacteria of the closely related genera *Lactobacillus*, *Ligilactobacillus*, and *Limosilactobacillus* are by far the most dominant for HAS and LAS, both in terms of relative abundance (>95%) as well as number of individual OTUs. Each intestinal segment also had a microbial signature unique from the others. At the order level, the duodenum had the most diversity, followed by the ileum, and very little diversity was found in the jejunum. These data are shown in Figure 3.

Alpha diversity box plots in Figure 4 further illustrate the microbial richness of LAS vs. HAS, as the total number of OTUs observed for LAS samples are greater than those of HAS for each intestinal segment. Additionally, while rarefaction curves are not shown, the diversity captured in HAS seems thorough, whereas additional sequencing in LAS may have identified additional OTUs.

Beta diversity, variation between samples, as shown by principal coordinate analysis in Figure 5 illustrates how the samples cluster by tissue and line. In general, it seems that within line microbial differences became less prominent while between line differences became more prominent from anterior to distal segments, given that samples cluster somewhat more closely in the ileum and jejunum than in the duodenum, particularly so for the HAS line.

Statistical differences in abundance

We found 206 OTUs which differed in abundance between lines, irrespective of tissue, at FDRp ≤0.05. Of those, 65 OTUs were observed in both lines. In contrast to most of the data where LAS had exhibited more diversity only 18 of the common OTUs were more abundant in LAS while 47 were more abundant in HAS as shown in Figure 6. Additionally, 100 of the 206 OTUs only present in one line were found exclusively in HAS while 41 were only found in LAS.

PERMANOVA revealed statistically significant line by tissue differences. The jejunum and ileum were significantly different between lines, though there was not a significant difference between the duodenum of HAS *versus* LAS. Within line, significant differences were observed between HAS duodenum and jejunum, and between LAS jejunum and ileum.
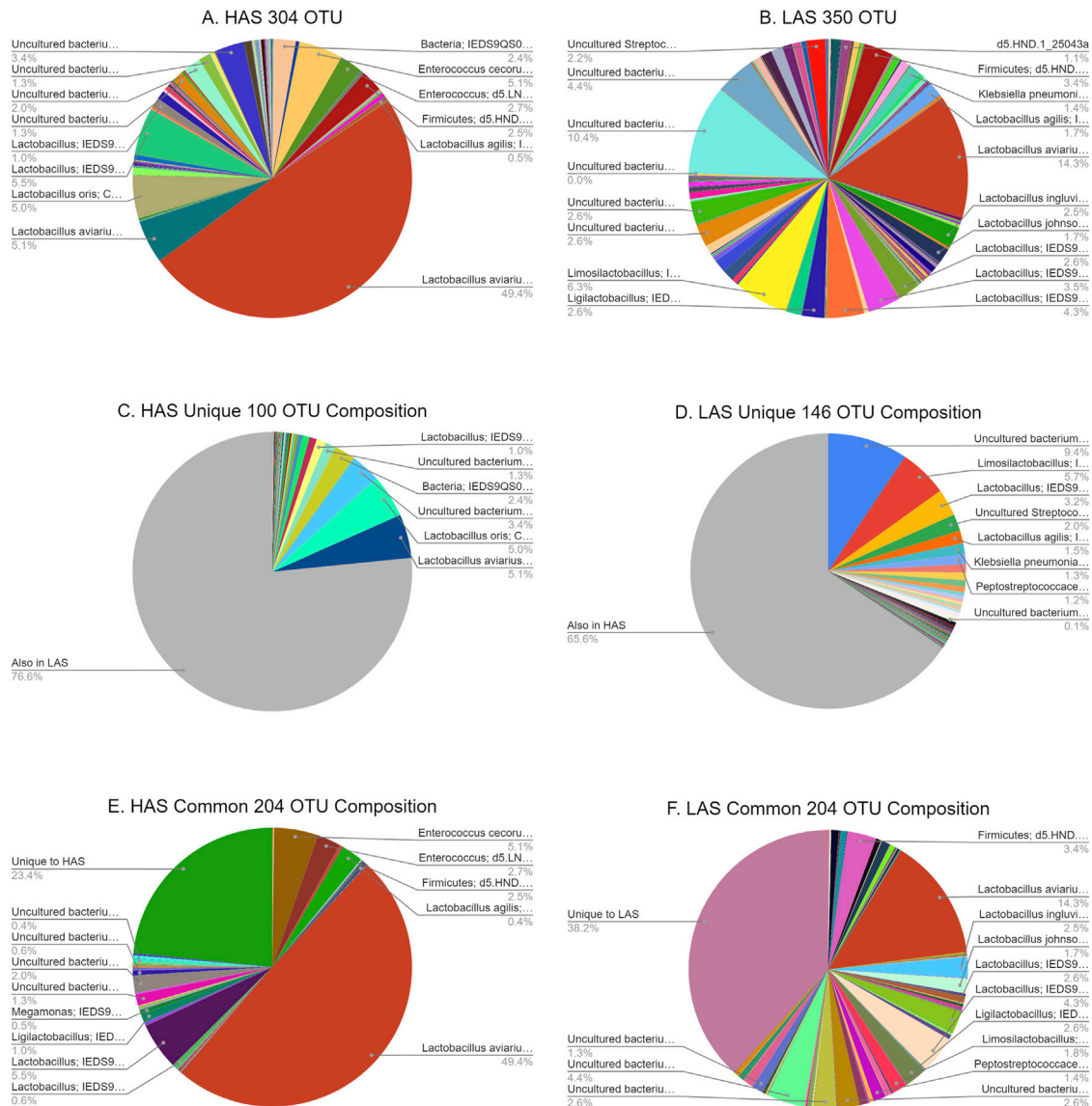
Functional analysis

Functional analysis identified 1385 KEGG molecules which were differentially abundant between lines; 704 more abundant in LAS *versus* 681 more abundant in HAS. Using the KEGG mapper tool for each list of molecules revealed enrichment of modules for different types of metabolic pathways between lines. Of the top 10 modules in HAS five were associated with biosynthesis of molecules fatty acid, lysine, pyrimidine, riboflavin, and coenzyme A. Four were involved in the 3-Hydroxypropionate bi-cycle, glycolysis, antimicrobial resistance, and ATPase activity. Conversely, LAS had three modules involved in degradation: AMP, GMP, and phenylacetate. Three modules were enriched for reductive carbon conversion, two were involved in menaquinone biosynthesis, and one in NAD biosynthesis. LAS and HAS both had enriched modules for C5 isoprenoid, though HAS was enriched for the mevalonate pathway whereas the non-mevalonate pathway was enriched in LAS.

Machine learning and host pathway analysis

Figure 7 summarizes the machine learning workflow. In the jejunum, the combination of the top twenty OTUs identified via machine learning resulted in an average predictive performance of 98.61%, that is to say the algorithms were able to identify patterns in the abundance data for these microbes which could be used to accurately classify the samples as being from the HAS line or LAS line on average 98.61% of the time. The randomized datasets performed at an average of 45%, which is very close to the 50% expected by random probability and assured us that "true" machine learning occurred. Of those top twenty OTUs, eleven had average individual predictive performance >75%, however two were only present in the LAS line. The nine remaining OTUs which were
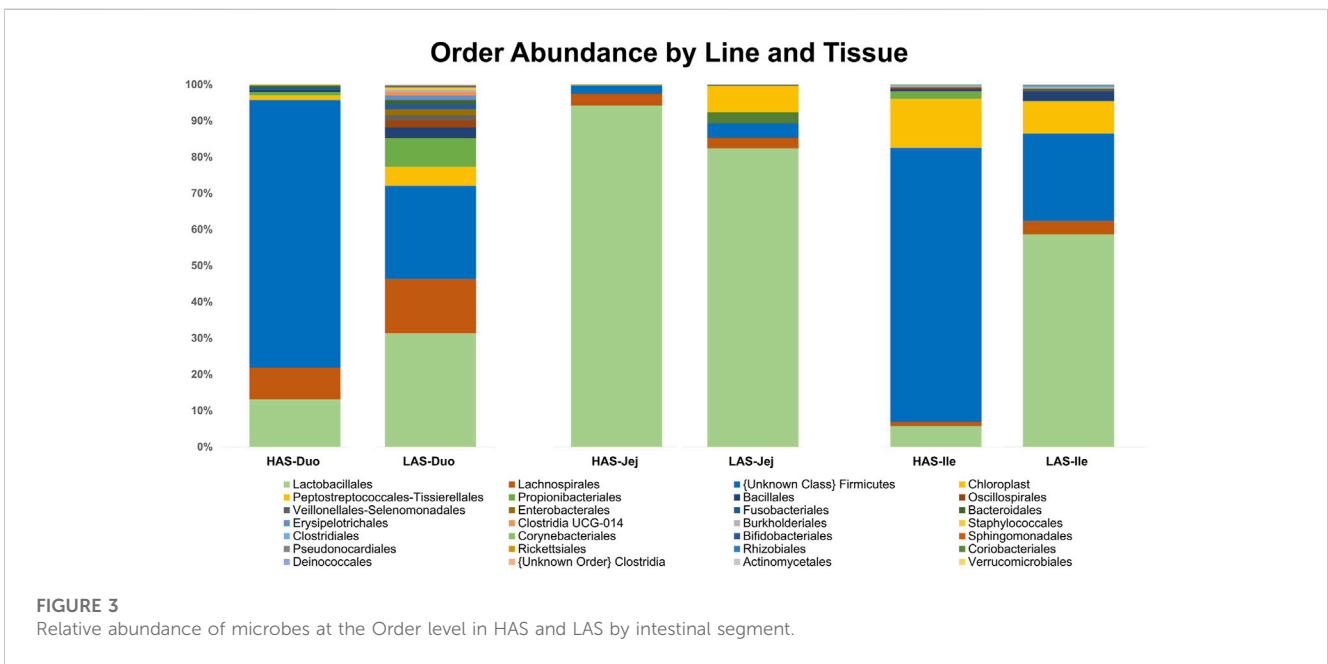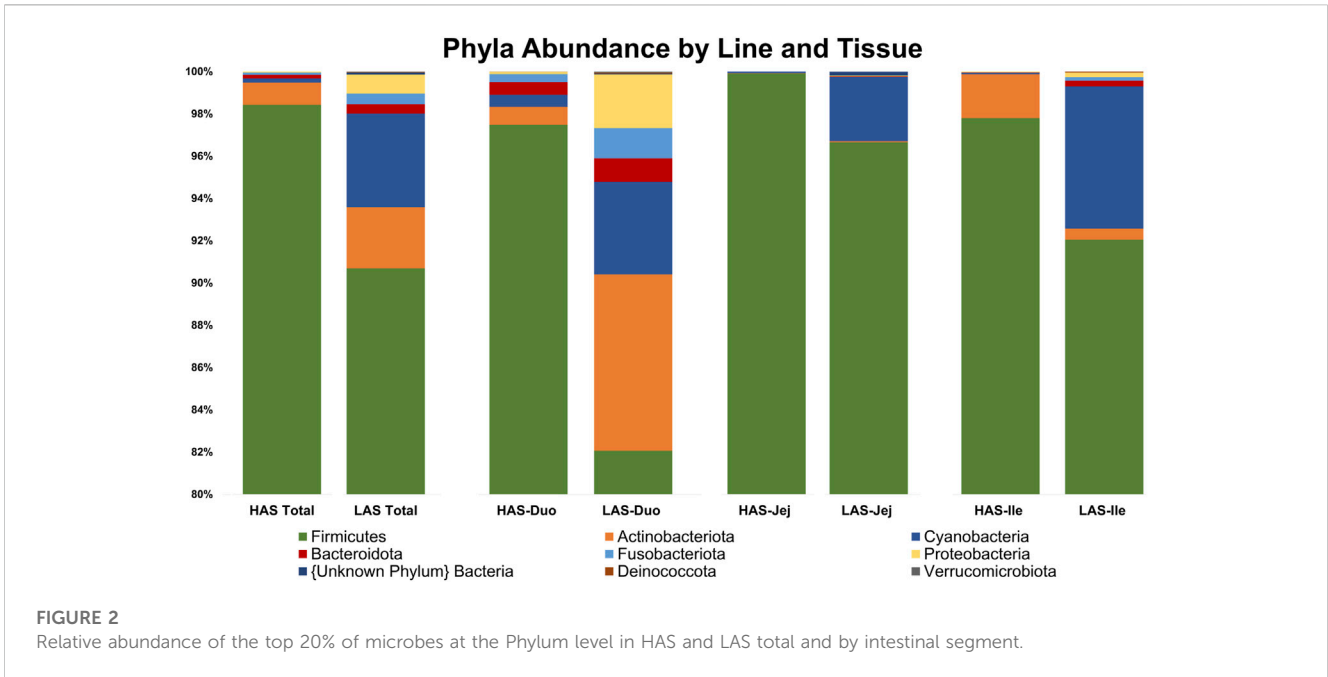
FIGURE 1
(A–F) Relative OTU composition of HAS and LAS small intestine.

both predictive of line and present in at least one sample from each line included eight uncultured strains of the genus *Lactobacillus* identified as EU776306.1.1423, AY958858.1.1540, EU460910.1.1421, EU774852.1.1423, EU774852.1.1423, IEDS9QS05FX4Q0, sp. HM218868.1.1554, and HM218952.1.1524, one species of *Lactobacillus crispatus*: EU559595.1.1562, and one of the genus *Enterococcus* EU459393.1.1420 (also uncultured). These OTUs were subsequently used to identify host gene expression associated with the relative abundance of each microbe, which were used as input for pathway enrichment analysis.

Each OTU dataset was run independently using the three machine learning algorithms and two cross validation methods. The relative abundance of microbes was used as the classifier (absent, low, medium, or high) and the predictive dataset attributes were composed of the gene expression data in transcripts per million. RNA was sequenced from the jejunum tissue samples, the contents from which the microbial DNA was isolated. The list of genes whose expression patterns most accurately predicted relative OTU abundance was determined for each of the nine OTUs and used for pathway enrichment analysis. Predictive performance for each optimized gene list was between 67% and 90%. Genes lists varied from 35–500 and IPA was able to map >60% of those in each list. The randomized negative control dataset performances were between 27% and 38%, close to the expected 33.3% indicative of random chance. The results of machine learning and IPA are summarized in Table 2, and all metrics for the machine learning algorithm runs with optimized feature sets can be found in Supplementary Table S1.

FIGURE 2
Relative abundance of the top 20% of microbes at the Phylum level in HAS and LAS total and by intestinal segment.

FIGURE 3
Relative abundance of microbes at the Order level in HAS and LAS by intestinal segment.
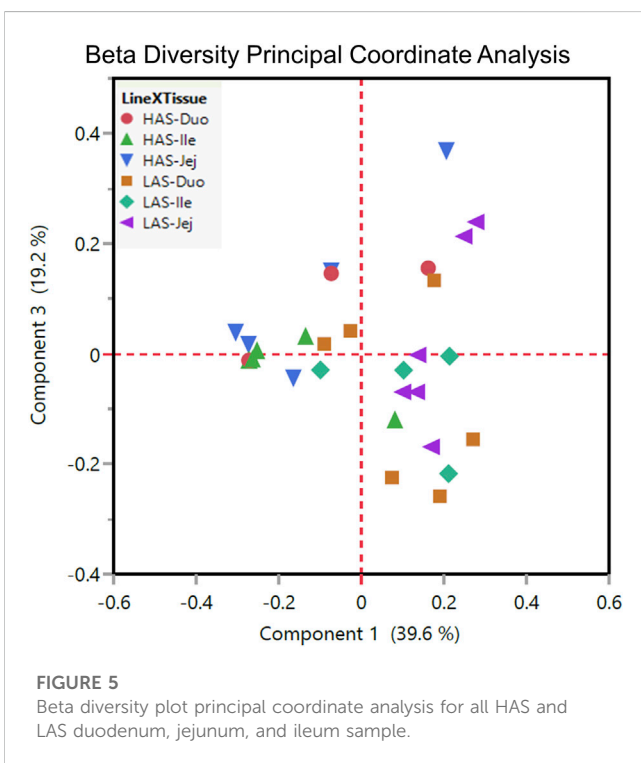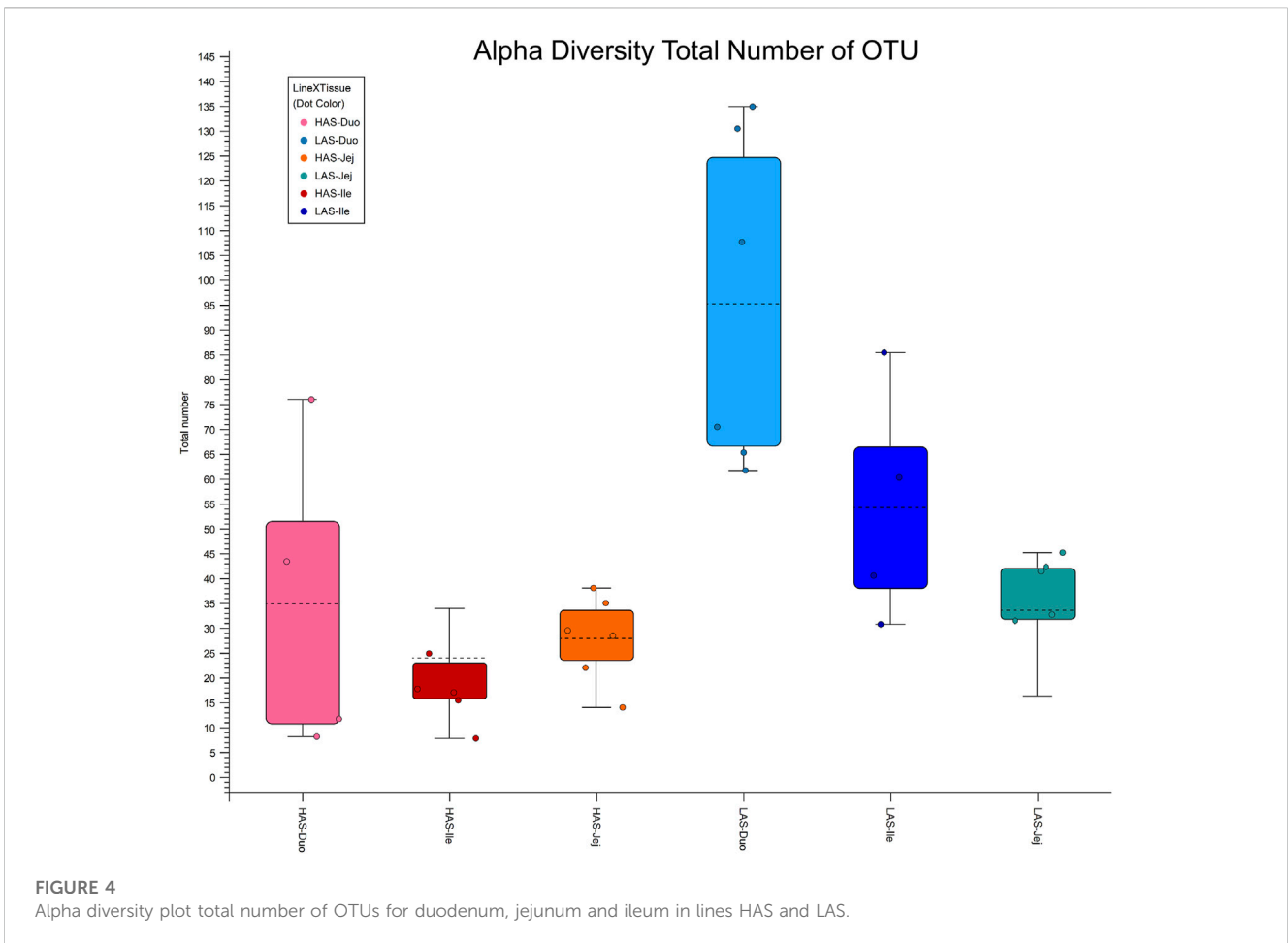
Discussion

Forty generations of divergent selection have resulted in a greater than six-fold difference in 5 day post injection antibody titer to sheep red blood cells between the HAS and LAS lines. Previous research has identified differences which help explain the phenotype, including their major histocompatibility B-haplotypes, as well as other genetic loci identified via pooled resequencing and quantitative trait loci mapping of an advanced intercross line (Gehad et al., 1999; Dorshorst, Siegel, and Ashwell, 2011; Lillie et al., 2017). However, much of the variation in antibody response has yet to be elucidated. The microbiome has become a key area of

immunology research, and given that commensal microbes are thought to coevolve with their host (Zhao et al., 2013; Meng et al., 2014; Yang et al., 2017; Koskella and Joy, 2020), and influence host humoral response to vaccines (Zimmermann and Curtis, 2018; Gonçalves et al., 2021) the microbiomes of HAS and LAS likely play a role in their differential antibody response phenotypes.

Sequencing the small intestinal microbiome of HAS and LAS has demonstrated an association between genetic selection and microbial composition and diversity. The microbiome of the small intestine were distinctly different between lines HAS and LAS, wherein LAS exhibited greater microbial diversity than

FIGURE 4
Alpha diversity plot total number of OTUs for duodenum, jejunum and ileum in lines HAS and LAS.

FIGURE 5
Beta diversity plot principal coordinate analysis for all HAS and LAS duodenum, jejunum, and ileum sample.

HAS, in agreement with previous fecal microbiome research (Zhao et al., 2013). Genetic selection for high or low antibody response to SRBCs appears to have resulted in an inverse relationship with commensal microbe diversity. While LAS exhibited greater numerical diversity, it is also important to note that from a statistical standpoint HAS displayed an increased abundance of more microbial strains. The discordant observances in statistical *versus* numerical abundance may be attributed to greater within line diversity in LAS than HAS. That is to say, HAS may have more microbes which differ statistically from LAS, but that is because LAS samples differ more from one another than HAS samples do. This is further supported by the results of the alpha diversity analysis, where the diversity captured in the HAS samples appears more comprehensive.

Lactobacillus, being a predominant genus in both lines, warrants additional exploration. The species/strains of lactobacilli present in one line are often significantly reduced or entirely absent in the other and it is important to consider the additive effects of small differences in microbial abundance and composition that may contribute to host physiology. Studies of feeding different probiotic *Lactobacillus* strains result in various degrees and types of immunostimulant (Perdigón, Fuller, and Raya, 2001). One of the main ways microbes exert an effect on host immune response in the gut is via fermentation products such as lactic acid, and the short
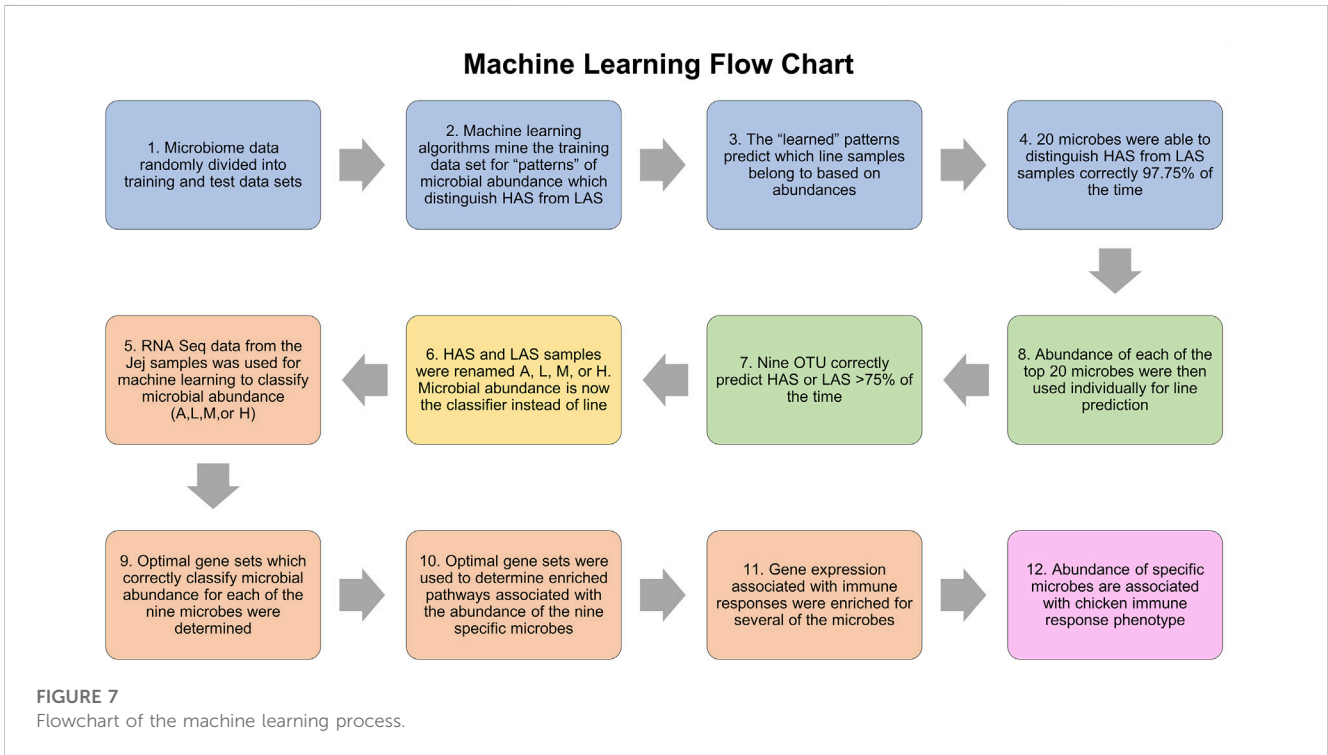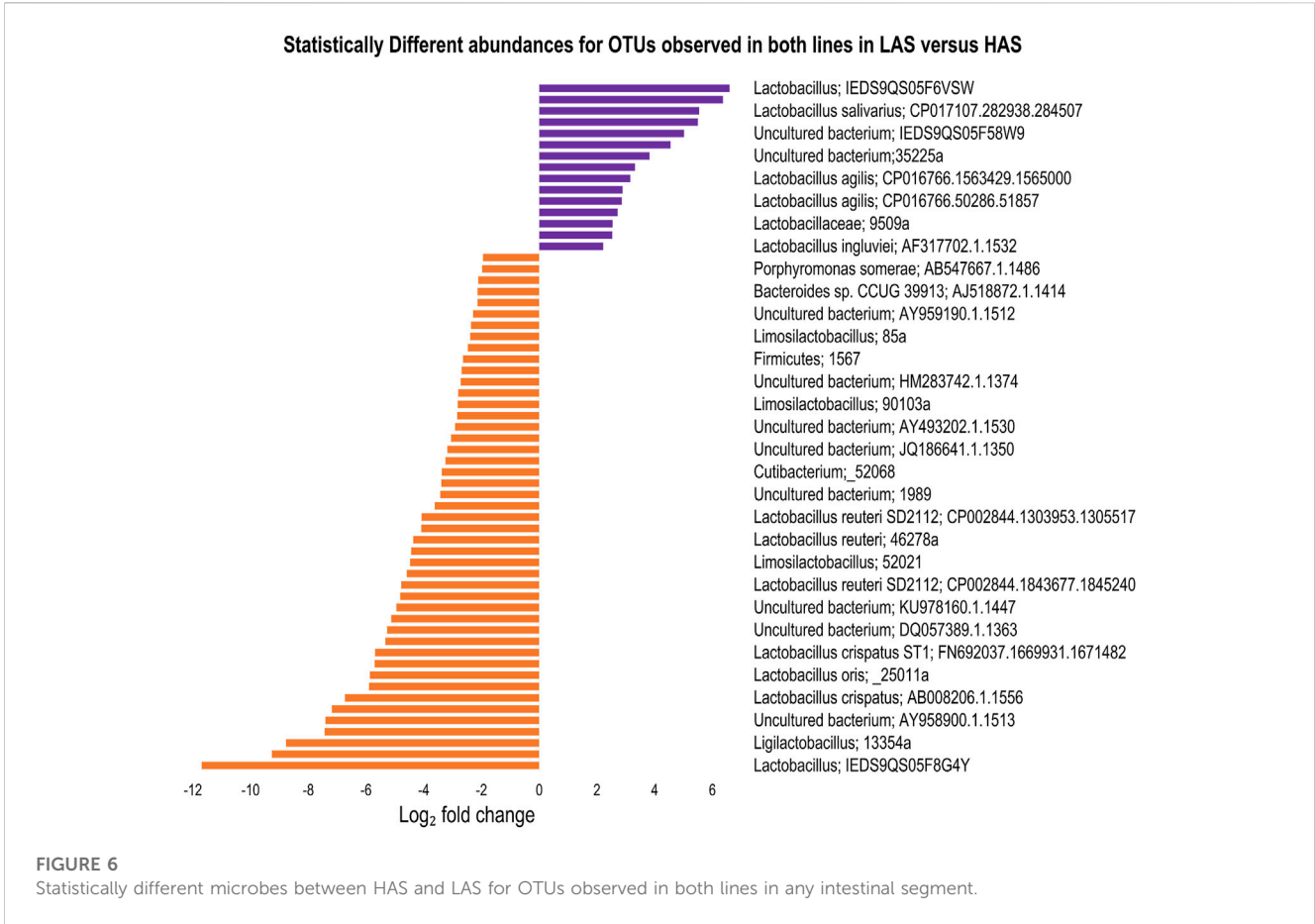
FIGURE 6
Statistically different microbes between HAS and LAS for OTUs observed in both lines in any intestinal segment.

FIGURE 7
Flowchart of the machine learning process.

TABLE 2 Functional pathway enrichment for genes associated with most predictive microbes.

Taxa	% Correct line prediction	Top ave % performance for RNASeq	# Genes for maximum performance	#IPA mapped genes	Top enriched pathways
Ent_UnCul_EU459393.1.1420	90	82	339	244	Neuroinflammation Signaling Pathway, Antigen Presentation Pathway, Virus Entry via Endocytic Pathways
LB_crispatus_EU559595.1.1562	83	85	35	22	Flavin Biosynthesis IV, Tetrapyrrole Biosynthesis II, Selenocysteine Biosynthesis II
LB_UnCul_EU776306.1.1423	78	78	325	253	Glycolysis I, Colanic Acid Building Blocks Biosynthesis, Heme Degradation
LB_UnCul_AY958858.1.1540	88	74	349	264	Gnaq Signaling, fMLP Signaling in Neutrophils, IL-1 Signaling
LB_UnCul_EU460910.1.1421	76	88	83	58	Synaptogenesis Signaling Pathway, Neuregulin Signaling, UDP-N-acetyl-D-galactosamine Biosynthesis II
LB_UnCul_EU774852.1.1423	79	85	500	444	Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis, ERK/MAPK Signaling, Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis
LB_UnCul_IEDS9QS05FX4Q0	76	90	228	186	Phagosome Maturation, Antigen Presentation Pathway, Colanic Acid Building Blocks Biosynthesis
LB_UnCul_sp.HM218868.1.1554	79	75	312	219	WNT/B-catenin Signaling, Neuroinflammation Signaling Pathway, Stearate Biosynthesis I, (SLC25A51)
LB_UnCul_HM218952.1.1524	78	67	45	40	NADH Repair, Salvage Pathways of Pyrimidine Deoxyribonucleotides, Insulin Secretion Signaling Pathway

chain fatty acids (SCFA) (Kim, 2018). As its name suggests, *Lactobacillus* is responsible for producing lactic acid and while some strains can produce short chain fatty acids or manipulate the microbial environment leading to SCFA production by other microbes (H. Li et al., 2020) there is a negative correlation between *Lactobacillus* and SCFA content, particularly butyrate, in the small intestine of poultry (H. Yang et al., 2018; Du et al., 2020; Ranjitkar et al., 2016). Butyrate, and to a lesser degree propionate, have been reported to inhibit dendritic cells, T helper cells, and B cells (Millard et al., 2002; Singh et al., 2010; Arpaia et al., 2013; Trompette et al., 2014; Sanchez et al., 2020) all of which would contribute to diminished immunoglobulin production. If the intestinal microbiome composition of HAS favors a reduction in SCFA producing bacteria compared to LAS, it could help explain the enhanced antibody response to sheep red blood cells observed in that line. Conversely, while intestinal lactic acid concentration has not been studied specifically in regard to antibody response, there are data to indicate that it may contribute positively. Cell culture experiments with supplemented lactic acid resulted in an increase in antibody production (Kromenaker and Frirdrich, 1994). Children

suffering from rotavirus induce diarrhea given oral probiotic *Lactobacillus gg* exhibited an increase in antibody secreting cells (Kaila et al., 1992) and broilers fed probiotic *Lactobacillus bulgaris* had increased antibody titer to Newcastle's Disease Virus vaccination (Apata, 2008). Additionally lactic acid bacteria are currently being investigated as a useful vector for vaccines (Szatraj, Szczepankowska, and Chmielewska-Jeznach, 2017). Whether it is a result of the higher levels of lactic acid, or some other metabolite produced by *Lactobacillus*, the antibody response to SRBCs observed in HAS is likely influenced by the increased abundance of commensal *Lactobacillus*.

Other microbial metabolites can also stimulate host physiology. Modules for microbial riboflavin and lysine biosynthesis were more enriched in HAS compared to LAS, which have been shown to impact immune cell function and antibody response. Antibodies to *Salmonella pullorum* were observed to be impaired in animals with diets deficient in riboflavin (Panda and Combs, 1963) and influenza vaccines containing riboflavin adjuvants resulted in increased antibody titers (Quintilio et al., 2016). Additionally, microbes have been shown to present riboflavin metabolites to the MHCI

related protein (MR1) of mucosal associated invariant T-cells (Treiner et al., 2003; Toubal et al., 2019). Lysine deficiency has been associated with reduced antibody response to Newcastle disease in broilers, and chickens on diets with supplemented lysine were found to have an increase in antibody titer (Chen, Sander, and Dale, 2003; Faluyi et al., 2015). Future studies may include full metagenomic sequencing to better characterize the spectrum of gut microbial metabolites which may impact chicken antibody response.

Machine learning allows for the addition of host pathway analysis related to microbe abundance. These data further support the involvement of commensal microbes in antibody response as the abundance of the microbial strains most predictive of host line are associated with enriched expression of host genes involved in antigen presentation, inflammation, and immune cell responses. Future studies may look to better determine how these specific microbes may be impacting the molecular pathways identified using this integrated machine learning approach.

By taking an integrated approach which includes 16S microbiome sequencing, functional microbiome analysis, and machine learning to characterizing the small intestinal microbiome of genetically selected lines, HAS and LAS we have been able to demonstrate differences which may influence antibody response to SRBCs. In agreement with the results of fecal microbiome sequencing (Yang et al., 2017) we saw an increase in microbial diversity in LAS birds and an increase in *Lactobacillus* abundance in HAS birds. In addition to characterizing the microbiota of the separate small intestinal segments *versus* the combined microbiota of feces, another novel aspect of our experiment is that the chickens in our study had not been injected with sheep red blood cells, thus the differences we observe are native and not influenced by antigen exposure. Based on what has been published with regard to *Lactobacillus* and microbial metabolites, our results support a role for the involvement of the associated gut microbiome with the high and low antibody response to SRBCs in these divergently selected lines. Including functional microbiome data has allowed for the inferential identification of enriched microbial modules, the products of which may influence host phenotype. Finally, machine learning allows for identifying specific microbe strains whose abundance accurately predicts line (HAS or LAS) and enables associating those microbial abundances with host gene expression data to further explore the host-microbe relationship to host antibody response. This is to our knowledge the first time these different data types have been integrated together in this way via the use of machine learning algorithms and offers a novel approach to better understanding host physiology with regard to microbiota composition.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, GSE206804 https://www.ncbi.nlm.nih.gov/, GSE240908.

Ethics statement

The animal study was approved by North Carolina State University Institutional Animal Care and Use Committee. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

SN: Conceptualization, Formal Analysis, Investigation, Methodology, Software, Writing–original draft, Writing–review and editing. CA: Conceptualization, Project administration, Resources, Supervision, Writing–review and editing. PS: Resources, Writing–review and editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2023.1304051/full#supplementary-material

References

Al-Marzooqi, W., Al-Maskari, S., Al-Kharousi, K., Johnson, E. H., and El Tahir, Y. (2020). Diversity of intestinal bacterial microbiota of indigenous and commercial strains of chickens using 16S rDNA-based analysis. *Animals* 10 (3), 391. doi:10.3390/ani10030391

Apata, D. (2008). Growth performance, nutrient digestibility and immune response of broiler chicks fed diets supplemented with a culture of Lactobacillus bulgaricus. *J. Sci. Food Agric.* 88 (7), 1253–1258. doi:10.1002/jsfa.3214

Arpaia, N., Campbell, C., Fan, X., Dikiy, S., van der Veeken, J., Paul, D. R., et al. (2013). Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* 504 (7480), 451–455. doi:10.1038/nature12726

Benson, A. K., Kelly, S. A., Ryan, L., and Ma, F. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc. Natl. Acad. Sci.* 107, 201007028. doi:10.1073/pnas.1007028107

Chen, C., Sander, J. E., and Dale, N. M. (2003). The effect of dietary lysine deficiency on the immune response to Newcastle disease vaccination in chickens. *Avian Dis.* 47 (4), 1346–1351. doi:10.1637/7008

Dorshorst, B. J., Siegel, P. B., and Ashwell, C. M. (2011). Genomic regions associated with antibody response to sheep red blood cells in the chicken. *Anim. Genet.* 42 (3), 300–308. doi:10.1111/j.1365-2052.2010.02146.x

Douglas, G. M., Beiko, R. G., and Langille, M. G. I. (2018). Predicting the functional potential of the microbiome from marker genes using PICRUSt. *Microbiome Analysis Methods Protoc.* 1849, 169–177. doi:10.1007/978-1-4939-8728-3_11

Douglas, G. M., Curtis, H., Langille, M. G. I., Yurgel, S. N., Brown, J. R., Taylor, C. M., et al. (2020). PICRUSt2 for prediction of metagenome functions. *Nat. Biotechnol.* 38 (6), 685–688. doi:10.1038/s41587-020-0548-6

Du, X., Xiang, Y., Lou, F., Tu, P., Zhang, X., Hu, X., et al. (2020). Microbial community and short-chain fatty acid mapping in the intestinal tract of quail. *Animals* 10 (6), 1006. doi:10.3390/ani10061006

Faluyi, O. B., Agbede, J. O., and Adebayo, I. A. (2015). Growth performance and immunological response to Newcastle disease vaccinations of broiler chickens fed lysine supplemented diets. *J. Veterinary Med. Animal Health* 7 (3), 77–84. doi:10.5897/jvmah2014.0328

Gehad, A. E., Mashaly, M. M., S Siegel, H., Dunnington, E. A., and Siegel, P. (1999). Effect of genetic selection and MHC haplotypes on lymphocyte proliferation and interleukin-2 like activity in chicken lines selected for high and low antibody production against sheep red blood cells. *Veterinary Immunol. Immunopathol.* 68 (1), 13–24. doi:10.1016/S0165-2427(99)00008-2

Glöckner, F. O., Yilmaz, P., Quast, C., Jan, G., Beccati, A., Ciuprina, A., et al. (2017). 25 Years of serving the community with ribosomal RNA gene reference databases and tools. *J. Biotechnol.* 261, 169–176. doi:10.1016/j.jbiotec.2017.06.1198

Gonçalves, E., Guillén, Y., Lama, J. R., Sanchez, J., Brander, C., Paredes, R., et al. (2021). Host transcriptome and microbiota signatures prior to immunization profile vaccine humoral responsiveness. *Front. Immunol.* 12, 657162. doi:10.3389/fimmu.2021.657162

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Peter, R., and Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explor. Newsl.* 11 (1), 10–18. doi:10.1145/1656274.1656278

Jordan, A., Carding, S. R., and Hall, L. J. (2022). The early-life gut microbiome and vaccine efficacy. *Lancet Microbe* 3 (10), e787–e794. doi:10.1016/S2666-5247(22)00185-9

Kaila, M., Isolauri, E., Soppi, E., Virtanen, E., Laine, S., and Arvilommi, H. (1992). Enhancement of the circulating antibody secreting cell response in human diarrhea by a human Lactobacillus strain. *Pediatr. Res.* 32 (2), 141–144. doi:10.1203/00006450-199208000-00002

Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein Sci. A Publ. Protein Soc.* 28 (11), 1947–1951. doi:10.1002/pro.3715

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 49 (D1), D545–D551. doi:10.1093/nar/gkaa970

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi:10.1093/nar/28.1.27

Kim, C. H. (2018). Immune regulation by microbiome metabolites. *Immunology* 154 (2), 220–229. doi:10.1111/imm.12930

Koskella, B., and Joy, B. (2020). The study of host–microbiome (Co)evolution across levels of selection. *Philosophical Trans. R. Soc. B Biol. Sci.* 375 (1808), 20190604. doi:10.1098/rstb.2019.0604

Kromenaker, S. J., and Friedrich, S. (1994). Effect of lactic acid on the kinetics of growth and antibody production in a murine hybridoma: secretion patterns during the cell cycle. *J. Biotechnol.* 34 (1), 13–34. doi:10.1016/0168-1656(94)90162-7

Leite, D. M. C., Brochet, X., Resch, G., Que, Y. A., Neves, A., and Peña-Reyes, C. (2018). Computational prediction of inter-species relationships through omics data analysis and machine learning. *BMC Bioinforma.* 19 (14), 420. doi:10.1186/s12859-018-2388-7

Li, G., Pan, Y., Zhang, L., Xie, B., and Shao, W. (2004). "Attribute ranking: an entropy-based approach to accelerating browsing-based component retrieval," in *Software reuse: methods, techniques and tools: 8th international conference* (Berlin Heidelberg: Springer).

Li, H., Liu, F., Lu, J., Shi, J., Guan, J., Yan, F., et al. (2020). Probiotic mixture of Lactobacillus plantarum strains improves lipid metabolism and gut microbiota structure in high fat diet-fed mice. *Front. Microbiol.* 11, 512. doi:10.3389/fmicb.2020.00512

Lillie, M., Sheng, Z., Honaker, C. F., Dorshorst, B. J., Ashwell, C. M., Siegel, P. B., et al. (2017). Genome-wide standing variation facilitates long-term response to bidirectional selection for antibody response in chickens. *BMC Genomics* 18, 99. doi:10.1186/s12864-016-3414-7

Lin, E., and Lane, H.-Y. (2017). Machine learning and systems genomics approaches for multi-omics data. *Biomark. Res.* 5 (1), 2. doi:10.1186/s40364-017-0082-y

Meng, H., Zhang, Y., Zhao, L., Zhao, W., He, C., Honaker, C. F., et al. (2014). Body weight selection affects quantitative genetic correlated responses in gut microbiota. *PLoS ONE* 9 (3), e89862. doi:10.1371/journal.pone.0089862

Millard, A. L., Mertes, P. M., Ittelet, D., Villard, F., Jeannesson, P., and Bernard, J. (2002). Butyrate affects differentiation, maturation and function of human monocyte-derived dendritic cells and macrophages. *Clin. Exp. Immunol.* 130 (2), 245–255. doi:10.1046/j.0009-9104.2002.01977.x

Nolin, S. J., Taylor, R. L., Edens, F. W., Siegel, P. B., Ashwell, C. M., Siegel, P. B., et al. (2023). Combining supervised machine learning with statistics reveals differential gene expression patterns related to energy metabolism in the jejuna of chickens divergently selected for antibody response to sheep red blood cells. *Poult. Sci.* 102, 102751. doi:10.1016/j.psj.2023.102751

Oakley, B. B., Hyun, S., Pedroso, A., Lee, M. D., Stephen, R., Collett, T. J. J., et al. (2014). The chicken gastrointestinal microbiome. *FEMS Microbiol. Lett.* 360 (2), 100–112. doi:10.1111/1574-6968.12608

Oh, M., Park, S., Sun, K., and Chae, H. (2021). Machine learning-based analysis of multi-omics data on the cloud for investigating gene regulations. *Briefings Bioinforma.* 22 (1), 66–76. doi:10.1093/bib/bbaa032

Panda, B., and Combs, G. F. (1963). Impaired antibody production in chicks fed diets low in vitamin A, pantothenic acid or riboflavin. *Proc. Soc. Exp. Biol. Med.* 113 (3), 530–534. doi:10.3181/00379727-113-28418

Perdigón, G., Fuller, R., and Raya, R. (2001). Lactic acid bacteria and their effect on the immune system. *Curr. Issues Intestinal Microbiol.* 2 (1), 27–42.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with arb. *Nucleic Acids Res.* 35 (21), 7188–7196. doi:10.1093/nar/gkm864

Quast, C., Pruesse, E., Yilmaz, P., Jan, G., Schweer, T., Yarza, P., et al. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41 (D1), D590–D596. doi:10.1093/nar/gks1219

Quintilio, W., Alessandro de Freitas, F., Rodriguez, D., Saldanha Kubrusly, F., Yourtov, D., Miyaki, C., et al. (2016). Vitamins as influenza vaccine adjuvant components. *Archives Virology* 161 (10), 2787–2795. doi:10.1007/s00705-016-2994-5

Ranjitkar, S., Lawley, B., Tannock, G., and Ricarda, M. E. (2016). Bacterial succession in the broiler gastrointestinal tract. *Appl. Environ. Microbiol.* 82 (8), 2399–2410. doi:10.1128/AEM.02549-15

Sanchez, H. N., Moroney, J. B., Gan, H., Shen, T., Im, J. L., Li, T., et al. (2020). B cell-intrinsic epigenetic modulation of antibody responses by dietary fiber-derived short-chain fatty acids. *Nat. Commun.* 11 (1), 60. doi:10.1038/s41467-019-13603-6

Shulzhenko, N., Morgun, A., Hsiao, W., Battle, M., Yao, M., Gavrilova, O., et al. (2011). Crosstalk between B Lymphocytes, microbiota and the intestinal epithelium governs immunity versus metabolism in the gut. *Nat. Med.* 17 (12), 1585–1593. doi:10.1038/nm.2505

Siegel, P. B., and Gross, W. B. (1980). Production and persistence of antibodies in chickens to sheep erythrocytes. 1. Directional selection. *Poult. Sci.* 59 (1), 1–5. doi:10.3382/ps.0590001

Singh, N., Thangaraju, M., Prasad, P. D., Martin, P. M., Lambert, N. A., Boettger, T., et al. (2010). Blockade of dendritic cell development by bacterial fermentation products butyrate and propionate through a transporter (Slc5a8)-Dependent inhibition of histone deacetylases. *J. Biol. Chem.* 285 (36), 27601–27608. doi:10.1074/jbc.M110.102947

Szatraj, K., Szczepankowska, A. K., and Chmielewska-Jeznach, M. (2017). Lactic acid bacteria — promising vaccine vectors: possibilities, limitations, doubts. *J. Appl. Microbiol.* 123 (2), 325–339. doi:10.1111/jam.13446

Toubal, A., Lotersztajn, S., Lehuen, A., and Lehuen, A. (2019). Mucosal-associated invariant T cells and disease. *Nat. Rev. Immunol.* 19 (10), 643–657. doi:10.1038/s41577-019-0191-y

Treiner, E., Duban, L., Bahram, S., Radosavljevic, M., Wanner, V., Tilloy, F., et al. (2003). Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. *Nature* 422 (6928), 164–169. doi:10.1038/nature01433

Trompette, A., Gollwitzer, E. S., Yadava, K., Sichelstiel, A. K., Sprenger, N., Ngom-Bru, C., et al. (2014). Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nat. Med.* 20 (2), 159–166. doi:10.1038/nm.3444

Yang, H., Xiao, Y., Gui, G., Li, J., Wang, J., and Li, D. (2018). Microbial community and short-chain fatty acid profile in gastrointestinal tract of goose. *Poult. Sci.* 97 (4), 1420–1428. doi:10.3382/ps/pex438

Yang, L., Liu, S., Ding, J., Dai, R., He, C., Xu, K., et al. (2017). Gut microbiota Co-microevolution with selection for host humoral immunity. *Front. Microbiol.* 8, 1243. doi:10.3389/fmicb.2017.01243

Yilmaz, P., Wegener Parfrey, L., Yarza, P., Jan, G., Pruesse, E., Quast, C., et al. (2014). The SILVA and 'all-species living tree project (LTP)' taxonomic frameworks. *Nucleic Acids Res.* 42 (D1), D643–D648. doi:10.1093/nar/gkt1209

Zhao, L., Wang, G., Siegel, P., He, C., Wang, H., Zhao, W., et al. (2013). Quantitative genetic background of the host influences gut microbiomes in chickens. *Sci. Rep.* 3, 1163. doi:10.1038/srep01163

Zimmermann, P., and Curtis, N. (2018). The influence of the intestinal microbiome on vaccine responses. *Vaccine* 36 (30), 4433–4439. doi:10.1016/j.vaccine.2018.04.066