RESEARCH



Predicting foodborne pathogens and probiotics taxa within poultry-related microbiomes using a machine learning approach

Moses B. Ayoola^{1,2}, Nisha Pillai³, Bindu Nanduri², Michael J. Rothrock Jr⁴ and Mahalingam Ramkumar^{3*}

Abstract

Background Microbiomes that can serve as an indicator of gut, intestinal, and general health of humans and animals are largely influenced by food consumed and contaminant bioagents. Microbiome studies usually focus on estimating the alpha (within sample) and beta (similarity/dissimilarity among samples) diversities. This study took a combinatorial approach and applied machine learning to microbiome data to predict the presence of disease-causing pathogens and their association with known/potential probiotic taxa. Probiotics are beneficial living microorganisms capable of improving the host organism's digestive system, immune function and ultimately overall health. Here, 16 S rRNA gene high-throughput Illumina sequencing of temporal pre-harvest (feces, soil) samples of 42 pastured poultry flocks (poultry in this entire work solely refers to chickens) from southeastern U.S. farms was used to generate the relative abundance of operational taxonomic units (OTUs) as machine learning input. Unique genera from the OTUs were used as predictors of the prevalence of foodborne pathogens (*Salmonella, Campylobacter* and *Listeria*) at different stages of poultry growth (START (2–4 weeks old), MID (5–7 weeks old), END (8–11 weeks old)), association with farm management practices and physicochemical properties.

Result While we did not see any significant associations between known probiotics and *Salmonella* or *Listeria*, we observed significant negative correlations between known probiotics (*Bacillus* and *Clostridium*) and *Campylobacter* at the mid-time point of sample collection. Our data indicates a negative correlation between potential probiotics and *Campylobacter* at both early and end-time points of sample collection. Furthermore, our model prediction shows that changes in farm operations such as how often the houses are moved on the pasture, age at which chickens are introduced to the pasture, diet composition and presence of other animals on the farm could favorably increase the abundance and activity of probiotics that could reduce *Campylobacter* prevalence.

Conclusion Integration of microbiome data with farm management practices using machine learning provided insights on how to reduce *Campylobacter* prevalence and transmission along the farm-to-fork continuum. Altering management practices to support proliferation of beneficial probiotics to reduce pathogen prevalence identified here could constitute a complementary method to the existing but ineffective interventions such as vaccination and

*Correspondence: Mahalingam Ramkumar ramkumar@cse.msstate.edu

Full list of author information is available at the end of the article



© The Author(s) 2023. **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/.

Page 2 of 14

bacteriophage cocktails usage. Study findings also corroborate the presence of bacterial genera such as *Caloramator*, *DA101*, *Parabacteroides* and *Faecalibacterium* as potential probiotics.

Keywords Microbiome, Machine learning, Pastured poultry, Foodborne pathogens, Probiotics

Background

Demand for poultry meat is growing exponentially as an affordable source of protein and it is currently the most consumed meat worldwide [1]. Chickens are the most efficient feed converter among the traditional livestock (such as pig, cattle and turkey) and this contributes to the increased demand for poultry products [2]. However, colonization of the chickens while on the farm and contamination of the final poultry product with zoonotic pathogens are of public health concern, oftentimes leading to shortage in supply and economic loss. Microbial colonization of newly hatched chicks is reported to start at the hatching stage, but there is a possibility that it could be earlier as evidenced by the passage of pathogens through the egg pores [3]. Strict hygiene measures are usually implemented in commercial poultry to reduce colonization of eggs and newly hatched chickens by microbes from the environment, but further evidence suggests that inheritance of the colonizers from the parent is possible at the embryonic stage [4].

The chicken gut helps maintain intestinal homeostasis by competitively excluding pathogens or preventing colonization. This competitive exclusion is expected to lower the energy required to maintain the immune system and ultimately improve chicken performance [5]. Prophylactic use of antibiotics as part of management practice or abuse of antibiotics to treat infections can lead to imbalance in the microbiome, drug resistance and exacerbation of infection. Due to the growing concern regarding the transfer of antibiotic resistance from animals to human, the United States banned the use of antibiotics as growth promoters in livestock farming, including poultry [6], which mandates the development of alternate strategies to limit the prevalence of pathogens in the production environment.

Dysbiosis in the chicken gut microbiome could affect intestinal microbiota, which in turn could impact immunity, digestion and intestinal integrity, which in turn affect energy available to the chicken host and have been established as a function of microbiome composition and diversity [7, 8]. Currently, beneficial microbes such as probiotics are being directly applied to the chicken or inside the egg for creating a healthy microbiome as a suitable alternative to antibiotics in both poultry and alternative poultry systems for animal welfare [9, 10]. Increased poultry growth, performance and immune response, as well as improved meat and egg quality, have been associated with healthy microbiomes nurtured by probiotics, but their usage is still largely at the evaluation stage as a suitable alternative to antibiotics usage [11]. Identifying relationships between management practices, impact on gut microbiome (including probiotics) and pathogen prevalence in pastured poultry management systems, where antibiotics are not used historically, could provide useful information on appropriate probiotic intervention mechanisms.

The application of machine learning to find biologically relevant patterns from large datasets is becoming common in many domains, including microbiome data analysis. The random forest algorithm has been successfully used to find relationships between human gut microbiomes and disease states such as atherosclerosis, diabetes and arthritis [12]. Associations between gut microbial single nucleotide variants and colorectal cancer have also been predicted and established using the random forest algorithm [13]. The application of machine learning tools enables analyses to move beyond the routine alpha- and beta-diversity comparisons often associated with microbiome data. This study generated microbiome datasets from pre-harvest (feces, soil) poultry-related samples from 42 pastured poultry flocks in the southeastern U.S. and (1) focused on how changes in the microbiome of temporal pastured poultry pre-harvest samples could help predict prevalence of common poultry pathogen species such as Salmonella, Campylobacter and Listeria in the poultry environment, (2) examined their associations with known and potential probiotics in the microbiome and (3) determined how these changes are related to different farm management practices and physicochemical properties of the farms. We initially employed three different machine algorithms, random forest (RF), support vector machine (SVM) and logistic regression (LogReg), for training and evaluation purposes to access their suitability for our microbiome and pastured poultry data. Cross-validation analysis of the results indicated that RF performed optimally in learning from training data and making the best prediction about the evaluation data, and this algorithm was employed for the downstream analysis of predicting presence of pathogens, associated changes in farm management practices and sample physicochemical levels.

Results

Microbiome and poultry data

A total of 1,393 feces and soil samples were collected at different times throughout the chicken growth cycle (Feces_START (2–4 weeks old)=200, Feces_MID (5–7 weeks old)=313, Feces_END (8–11 weeks old)=185,

Table 1 Comparison of random forest (RF), support vectormachine (SVM) and logistic regression (LogReg). Modelperformance was considered good when 5-fold cross-validationvalue ≥ 0.7

Pathogens 1 Salmonella 0.85 0.85 0.78 2 Campylobacter 0.80 0.65 0.74 3 Listeria 0.86 0.86 0.79 Common Farm Practice Variables 4 BroodBedding 0.91 0.91 0.88 5 BrGMOFree 0.78 0.69 0.72 6 BrSoyFree 0.84 0.83 0.88 7 BrMedicated 0.97 0.94 0.84 8 AvgAgeToPasture 0.73 0.61 0.71 9 PastureHousing 0.72 0.55 0.62 10 FreqHousingMove 0.98 0.98 0.96 11 AlwaysNewPasture 0.9 0.87 0.86 12 PaGMOFree 0.77 0.67 0.71 13 PaSoyFree 0.76 0.64 0.71 14 PaMedicated 0.98 0.95 0.93 15 LayersOnFarm 0.74 </th <th>Number</th> <th>Target Variables</th> <th>RF</th> <th>SVM</th> <th>LogReg</th>	Number	Target Variables	RF	SVM	LogReg		
1 Salmonella 0.85 0.85 0.78 2 Campylobacter 0.80 0.65 0.74 3 Listeria 0.80 0.65 0.74 4 BroodBedding 0.91 0.91 0.88 5 BrGMOFree 0.78 0.69 0.72 6 BrSoyFree 0.84 0.83 0.8 7 BrMedicated 0.97 0.97 0.94 8 AvgAgeToPasture 0.73 0.61 0.71 9 PastureHousingMove 0.98 0.98 0.96 11 AlwaysNewPasture 0.9 0.87 0.86 12 PaGMOFree 0.77 0.67 0.73 13 PaSoyFree 0.76 0.64 0.71 14 PaMedicated 0.98 0.98 0.95 15 LayersOnFarm 0.77 0.6 0.7 17 SwineOnFarm 0.74 0.63 0.7 18 <td< td=""><td></td><td>Pathogens</td><td></td><td></td><td></td></td<>		Pathogens					
2Campylobacter0.800.650.743Listeria0.860.860.794BroodBedding0.910.910.885BrGMOFree0.840.830.815BrSoyFree0.840.830.816BrSoyFree0.840.830.817BrMedicated0.970.940.918AvgAgeToPasture0.720.550.6210FreqHousingMove0.880.980.9611AlwaySNewPasture0.970.8412PaGMOFree0.760.440.7113PaSoyFree0.760.440.7114PaMedicated0.980.950.9315LayersOnFarm0.950.930.7116CattleOnFarm0.740.630.7117SwineOnFarm0.740.630.7118SobepOnFarm0.740.820.8220WaterSource0.70.520.8221PreqBirdHandling0.940.810.8122AnimalSource0.940.810.8123Pi0.550.820.8124Moisture0.940.630.8125EC0.940.840.8126Kation0.940.840.8127TotalN0.940.840.8128CutalN0.940.850.8229CutalN<	1	Salmonella	0.85	0.85	0.78		
3Listeria0.860.860.794BroodBedding0.910.910.885BrGMOFree0.780.690.726BrSoyFree0.780.970.447BrMedicated0.970.970.448AvageToPasture0.730.610.719PastureHousing0.720.550.6210FreqHousingMove0.980.980.9611AlwaysNewPasture0.90.870.8612PaGMOFree0.770.670.7313PaSoyFree0.770.670.7314PaMedicated0.980.950.9315LayersOnFarm0.740.630.7116CattleOnFarm0.740.630.7117SwineOnFarm0.740.630.7118GoatsOnFarm0.740.630.7119SheepOnFarm0.740.630.8120WaterSource0.700.850.8121FreqBirdHandling0.90.870.8622AnimalSource0.940.850.8123CatlaC0.940.630.8124PH0.940.630.8125CatlaC0.940.630.8126Moisture0.940.630.8127CatlaC0.940.550.8228CatlaC0.940.550.8229 </td <td>2</td> <td>Campylobacter</td> <td>0.80</td> <td>0.65</td> <td>0.74</td>	2	Campylobacter	0.80	0.65	0.74		
Common Farent Parent	3	Listeria	0.86	0.86	0.79		
4BroodBedding0.910.910.885BrGMOFree0.780.690.726BrSoyFree0.840.830.87BrMedicated0.970.970.948AvgAgeToPasture0.730.610.719PastureHousing0.720.550.6210FreqHousingMove0.980.980.9811AlwaysNewPasture0.90.870.8612PaGMOFree0.770.670.7313PaSoyFree0.760.640.7114PaMedicated0.980.980.9315LayersOnFarm0.950.930.5116CattleOnFarm0.740.630.7117SwineOnFarm0.740.630.7118GoatSOnFarm0.740.630.7119SheepOnFarm0.780.550.7220WaterSource0.700.520.6221PreqBirdHandling0.900.850.8122AnyABXUse0.980.950.8123EC0.960.850.8124PH0.920.660.8125EC0.940.630.8326Moisture0.940.630.8227TotalN0.920.550.7232Ca0.950.630.8128Ci0.910.550.9229Cital <t< td=""><td></td><td colspan="6">Common Farm Practice Variables</td></t<>		Common Farm Practice Variables					
5BrGMOFree0.780.690.726BrSoyFree0.840.830.87BrMedicated0.970.970.948AvgAgeToPasture0.730.610.719PastureHousing0.720.550.6210FreqHousingMove0.980.980.9611AlwaysNewPasture0.970.670.7312PaGMOFree0.770.670.7313PaSoyFree0.760.640.7114PaMedicated0.980.980.9315LayerSOnFarm0.950.930.9516CattleOnFarm0.770.660.717SwineOnFarm0.780.550.718GoatsOnFarm0.740.630.719SheepOnFarm0.780.550.6220WaterSource0.900.870.8621FreqBirdHandling0.90.870.8622AnyABXUse0.980.950.8123EC0.960.650.8124PH0.930.550.8225EC0.940.630.8326CNAtio0.940.630.8327CAa0.950.630.8428CotalN0.940.550.9230Al0.940.550.9231B0.940.550.9232Ca0.950.	4	BroodBedding	0.91	0.91	0.88		
6BrSoyFree0.840.830.87BrMedicated0.970.970.948AvgAgeToPasture0.730.610.719PastureHousing0.720.550.6210FreqHousingMove0.980.980.9811AlwaySNewPasture0.90.770.6712PaGMOFree0.760.640.7113PaSoyFree0.760.640.7114PaMedicated0.980.980.9515LayersOnFarm0.750.620.7316CattleOnFarm0.770.60.717SwineOnFarm0.830.810.7118GoatsOnFarm0.780.550.720WaterSource0.700.520.6221FreqBirdHandling0.900.870.8622AnyABXUse0.980.980.8123EC0.960.650.8124PH0.930.560.8125EC0.940.630.8226Ala0.940.650.8227TotalC0.940.630.8328Cla0.940.550.8229CNRatio0.940.650.9231B0.940.550.9232Ca0.950.630.8433Cd0.940.650.9234Cd0.940.650.	5	BrGMOFree	0.78	0.69	0.72		
7Brideciated0.970.970.948AvgAgeToPasture0.730.610.719PastureHousing0.720.550.6210FreqHousingMove0.980.980.9611AlwaysNewPasture0.90.870.8612PaGMOFree0.770.670.7313PaSoyFree0.760.640.7114PaMedicated0.980.980.9515LayeronFarm0.950.930.5116CatteOnFarm0.740.630.717SwineOnFarm0.740.630.718GoatsOnFarm0.740.630.719SheepOnFarm0.780.550.720WaterSource0.70.520.6221FreqBirdHandling0.90.850.8122AnyABXUse0.980.950.8123EC0.930.650.8124PH0.930.550.8225EC0.940.630.8126Moisture0.940.630.8127TotalC0.940.650.8228TotalN0.940.650.8229CNRatio0.940.650.8231B0.940.650.8232Ca0.940.950.8234Cf0.950.650.8235Ca0.940.55 <td>6</td> <td>BrSoyFree</td> <td>0.84</td> <td>0.83</td> <td>0.8</td>	6	BrSoyFree	0.84	0.83	0.8		
8AvgAgeToPasture0.730.610.719PastureHousing0.720.550.6210FreqHousingMove0.980.980.9811AlwaysNewPasture0.90.870.8612PaGMOFree0.770.670.7313PaSoyFree0.760.540.7114PaMedicated0.980.980.9515LayersOnFarm0.950.950.9316CattleOnFarm0.770.60.717SwineOnFarm0.780.550.719SheepOnFarm0.780.550.720WaterSource0.70.520.6221FreqBirdHandling0.90.870.8622AnyABXUse0.980.980.9523AnimalSource0.900.850.8125EC0.960.650.8125EC0.960.650.8126Moisture0.940.630.8327TotalN0.940.630.8128Cd0.940.630.8230Al0.940.630.8231B0.940.630.8432Cu0.940.630.8433Cd0.940.630.8434Cr0.950.610.8735Cu0.950.610.8736Ma0.950.610.87	7	BrMedicated	0.97	0.97	0.94		
9 PastureHousing 0.72 0.55 0.62 10 FreqHousingMove 0.98 0.98 0.96 11 AlwaysNewPasture 0.9 0.87 0.86 12 PaGMOFree 0.77 0.67 0.73 13 PaSoyFree 0.76 0.64 0.71 14 PaMedicated 0.98 0.98 0.95 15 LayersOnFarm 0.95 0.95 0.93 16 CattleOnFarm 0.77 0.6 0.7 17 SwineOnFarm 0.78 0.55 0.7 19 SheepOnFarm 0.78 0.55 0.7 20 WaterSource 0.90 0.87 0.86 21 FreqBirdHandling 0.9 0.85 0.87 22 AnyABXUse 0.90 0.85 0.87 23 AnimalSource 0.90 0.85 0.81 25 EC 0.96 0.65 0.82 26	8	AvgAgeToPasture	0.73	0.61	0.71		
Ind FreqHousingNove 0.98 0.98 0.96 11 AlwaysNewPasture 0.9 0.87 0.86 12 PaGMOFree 0.77 0.67 0.73 13 PaSoyFree 0.76 0.64 0.71 14 PaMedicated 0.98 0.98 0.95 15 LayersOnFarm 0.95 0.95 0.93 16 CattleOnFarm 0.77 0.60 0.7 17 SwineOnFarm 0.78 0.55 0.7 19 SheepOnFarm 0.78 0.55 0.62 20 WaterSource 0.70 0.52 0.62 21 FreqBirdHandling 0.91 0.87 0.86 22 AnyABXUse 0.98 0.95 0.81 23 EC 0.96 0.65 0.81 24 PH 0.93 0.55 0.82 25 EC 0.96 0.65 0.81 26 Natiore	9	PastureHousing	0.72	0.55	0.62		
11AlwaysNewPasture0.90.870.8612PaGMOFree0.770.670.7313PaSoyFree0.760.640.7114PaMedicated0.980.980.9515LayersOnFarm0.950.950.9316CattleOnFarm0.770.60.717SwineOnFarm0.830.80.8118GoatsOnFarm0.740.630.719SheepOnFarm0.780.550.720WaterSource0.70.520.6221FreqBirdHandling0.90.870.8622AnyABXUse0.980.980.9523AnimalSource0.900.850.8125EC0.960.650.8126Moisture0.970.660.9127TotalC0.940.630.8328TotalN0.940.580.7731B0.940.580.7732Ca0.950.630.8433Cd0.940.630.8334Cr0.980.950.9235Cu0.940.650.9136Fe0.950.610.8737K0.970.550.9138Mg0.950.630.8139Mn0.910.630.8139Mn0.910.630.8140N	10	FreqHousingMove	0.98	0.98	0.96		
12 PaGMOFree 0.77 0.67 0.73 13 PaSoyFree 0.76 0.64 0.71 14 PaMedicated 0.98 0.98 0.93 15 LayersOnFarm 0.95 0.95 0.93 16 CattleOnFarm 0.77 0.6 0.7 17 SwineOnFarm 0.74 0.63 0.7 18 GoatsOnFarm 0.74 0.63 0.7 19 SheepOnFarm 0.78 0.55 0.7 20 WaterSource 0.7 0.52 0.62 21 FreqBirdHandling 0.9 0.87 0.88 22 AnyABXUse 0.98 0.98 0.95 23 AnimalSource 0.90 0.85 0.81 25 EC 0.96 0.65 0.89 26 Moisture 0.94 0.63 0.83 27 TotalC 0.94 0.65 0.92 28 Otal	11	AlwaysNewPasture	0.9	0.87	0.86		
13PaSoyFree0.760.640.7114PaMedicated0.980.980.9515LayersOnFarm0.950.950.9316CattleOnFarm0.770.60.717SwineOnFarm0.730.630.718GoatsOnFarm0.740.630.719SheepOnFarm0.780.550.720WaterSource0.70.520.6221FreqBirdHandling0.90.870.8622AnyABXUse0.900.850.8723AnimalSource0.900.850.8125EC0.960.650.8926Moisture0.970.660.9127TotalC0.940.630.8328TotalN0.940.630.8329CNRatio0.980.550.8230Al0.920.550.7331B0.940.630.8433Cd0.940.640.8434Cr0.960.650.9235Cu0.970.550.936Ma0.970.550.937K0.970.550.938Mg0.970.550.939Mn0.910.630.8140Na0.960.650.8939Mn0.960.650.8941Na0.96	12	PaGMOFree	0.77	0.67	0.73		
14PaMedicated0.980.980.980.9515LayersOnFarm0.950.950.9316CattleOnFarm0.770.60.717SwineOnFarm0.830.80.8118GoatsOnFarm0.740.630.719SheepOnFarm0.780.550.720WaterSource0.70.520.6221FreqBirdHandling0.90.870.8622AnyABXUse0.980.980.9523AnimalSource0.900.850.8724pH0.930.560.8125EC0.960.650.8926Moisture0.970.660.9127TotalC0.940.630.8328TotalN0.940.630.8329CNRatio0.980.550.8230Al0.920.550.7331B0.940.630.8433Cd0.940.650.9234Cr0.940.650.9235Cu0.940.650.9236Mg0.950.610.8737K0.970.550.9138Mg0.950.630.8839Mn0.910.630.8841Na0.960.650.8939Mn0.910.650.8941Na <td< td=""><td>13</td><td>PaSoyFree</td><td>0.76</td><td>0.64</td><td>0.71</td></td<>	13	PaSoyFree	0.76	0.64	0.71		
15LayersOnFarm0.950.950.9316CattleOnFarm0.770.60.717SwineOnFarm0.830.80.8118GoatsOnFarm0.740.630.719SheepOnFarm0.780.550.720WaterSource0.70.520.6221FreqBirdHandling0.90.870.8622AnyABXUse0.980.980.9523AnimalSource0.900.850.8724pH0.930.560.8125EC0.960.650.8126Moisture0.970.660.9127TotalC0.940.630.8328TotalN0.940.630.8329CNRatio0.940.550.7331B0.940.550.7331B0.940.650.9234Cr0.940.650.9235Cu0.940.650.9236Fe0.950.610.8437K0.970.550.938Mg0.950.630.8441Na0.910.630.8142Ni0.940.650.8939Mn0.910.650.8943P0.970.670.8944Pb0.910.550.7945Si0.940.57 <t< td=""><td>14</td><td>PaMedicated</td><td>0.98</td><td>0.98</td><td>0.95</td></t<>	14	PaMedicated	0.98	0.98	0.95		
16CattleOnFarm0.770.60.717SwineOnFarm0.830.80.8118GoatsOnFarm0.740.630.719SheepOnFarm0.780.550.720WaterSource0.70.520.6221FreqBirdHandling0.90.870.8622AnyABXUse0.980.980.9523AnimaSource0.900.870.8124pH0.930.560.8125EC0.960.650.8926Moisture0.970.660.9127TotalC0.940.630.8328TotalN0.940.630.8329CNRatio0.980.550.8230Al0.920.550.7331B0.940.630.8433Cd0.940.630.8434Cr0.950.610.8735Cu0.970.550.936Fe0.950.610.8737K0.970.550.938Mg0.960.680.8841Na0.960.650.8939Mn0.910.630.8141Na0.960.650.8939Mn0.910.630.8141Na0.960.650.8939Mn0.910.630.81 <td>15</td> <td>LayersOnFarm</td> <td>0.95</td> <td>0.95</td> <td>0.93</td>	15	LayersOnFarm	0.95	0.95	0.93		
17SwineOnFarm0.830.80.8118GoatsOnFarm0.740.630.719SheepOnFarm0.780.550.720WaterSource0.70.520.6221FreqBirdHandling0.90.870.8622AnyABXUse0.980.980.9523AnimalSource0.900.850.8124pH0.930.560.8125EC0.960.650.8926Moisture0.970.660.9127TotalC0.940.630.8328TotalN0.940.630.8329CNRatio0.980.550.8230Al0.940.580.7732Ca0.940.580.7733Cd0.940.630.8434Cr0.980.550.9235Cu0.940.610.8436Fe0.950.610.8737K0.970.550.9138Mg0.910.630.8140Mo0.980.70.9241Na0.960.650.8942Ni0.960.650.8943P0.910.550.7945S0.910.550.7945Si0.940.650.8946Si0.940.570.78<	16	CattleOnFarm	0.77	0.6	0.7		
18GoatsOnFarm0.740.630.719SheepOnFarm0.780.550.720WaterSource0.70.520.6221FreqBirdHandling0.90.870.8622AnyABXUse0.980.980.9523AnimalSource0.900.850.8724PH0.930.560.8125EC0.960.650.8926Moisture0.970.660.9127TotalC0.940.630.8328TotalN0.940.630.8329CNRatio0.940.550.8230Al0.920.550.7331B0.940.630.8433Cd0.940.610.8734Cr0.980.920.9235Cu0.940.610.8736Mg0.950.610.8737K0.970.550.938Mg0.910.630.8140Mo0.960.680.8839Mn0.910.630.8141Na0.960.650.8939Mn0.960.650.8941Na0.960.650.8942Ni0.970.550.7943P0.910.550.7944Pb0.910.540.734	17	SwineOnFarm	0.83	0.8	0.81		
19SheepOnFarm0.780.550.720WaterSource0.70.520.6221FreqBirdHandling0.90.870.8622AnyABXUse0.980.980.9523AnimalSource0.900.850.87 Physicochemical Propertur 24pH0.930.560.8125EC0.960.650.8926Moisture0.940.630.8328TotalN0.940.630.8329CNRatio0.980.550.8230Al0.920.550.7331B0.940.630.8433Cd0.980.650.9234Cr0.980.650.9235Cu0.940.630.8436Fe0.950.610.8737K0.970.550.9138Mg0.970.550.9139Mn0.910.630.8140Mo0.940.640.8441Na0.960.650.8943P0.970.670.8944Pb0.910.550.7945Si0.930.540.7347X0.930.540.73	18	GoatsOnFarm	0.74	0.63	0.7		
20WaterSource0.70.520.6221FreqBirdHandling0.90.870.8622AnyABXUse0.980.980.9523AnimalSource0.900.850.87 Physicochemical Properties 24pH0.930.560.8125EC0.960.650.8926Moisture0.970.660.9127TotalC0.940.630.8328TotalN0.940.630.8329CNRatio0.980.550.8230Al0.920.550.7331B0.940.630.8433Cd0.980.650.9234Cr0.980.650.9235Cu0.940.610.8737K0.970.550.9138Mg0.970.550.9139Mn0.910.630.8140Mo0.980.770.9241Na0.960.680.8842Ni0.960.650.8943P0.970.670.8944Pb0.910.550.7945Si0.930.540.7347ZNi0.930.540.73	19	SheepOnFarm	0.78	0.55	0.7		
21FreqBirdHandling0.90.870.8622AnyABXUse0.980.950.9523AnimalSource0.900.850.87 Physicochemical Properter 24pH0.930.560.8125EC0.960.650.8926Moisture0.970.660.9127TotalC0.940.630.8328TotalN0.940.630.8329CNRatio0.980.550.8230Al0.920.550.7331B0.940.630.8433Cd0.980.650.9234Cr0.980.650.9235Cu0.940.610.8737K0.970.550.938Mg0.970.550.939Mn0.910.630.8141Na0.960.680.8842Ni0.960.650.8943P0.970.670.8944Pb0.910.550.7945Si0.930.540.7347Xi0.930.540.73	20	WaterSource	0.7	0.52	0.62		
22 AnyABXUse 0.98 0.98 0.98 0.95 23 AnimalSource 0.90 0.85 0.87 Physicochemical Properties 24 pH 0.93 0.56 0.81 25 EC 0.96 0.65 0.89 26 Moisture 0.97 0.66 0.91 27 TotalC 0.94 0.63 0.83 28 TotalN 0.94 0.63 0.83 29 CNRatio 0.98 0.55 0.82 30 Al 0.92 0.55 0.73 31 B 0.94 0.58 0.77 32 Ca 0.95 0.63 0.84 33 Cd 0.98 0.55 0.92 34 Cr 0.98 0.90 0.92 35 Cu 0.94 0.66 0.84 36 Fe 0.95 0.61 0.87 37 K 0.97 0.55 0.9 38 Mg 0.96	21	FregBirdHandling	0.9	0.87	0.86		
23 AnimalSource 0.90 0.85 0.87 Physicochemical Properties 24 pH 0.93 0.56 0.81 25 EC 0.96 0.65 0.89 26 Moisture 0.97 0.66 0.91 27 TotalC 0.94 0.6 0.85 28 TotalN 0.94 0.63 0.83 29 CNRatio 0.98 0.55 0.82 30 Al 0.92 0.55 0.73 31 B 0.94 0.58 0.77 32 Ca 0.95 0.63 0.84 33 Cd 0.98 0.90 0.92 34 Cr 0.98 0.90 0.92 35 Cu 0.94 0.66 0.84 36 Fe 0.95 0.61 0.87 37 K 0.97 0.55 0.9 38 Mg 0.95 0.69 0.85 39 Mn 0.91 0.63 0	22	AnyABXUse	0.98	0.98	0.95		
Physicochemical Properties24pH0.930.560.8125EC0.960.650.8926Moisture0.970.660.9127TotalC0.940.630.8328TotalN0.940.630.8329CNRatio0.980.550.8230Al0.920.550.7331B0.940.580.7732Ca0.940.630.8433Cd0.940.630.8434Cf0.980.650.9235Cu0.940.610.8736Fe0.950.610.8737K0.970.550.938Mg0.970.550.939Mn0.910.630.8140Mo0.960.680.8842Ni0.960.650.8943P0.970.670.8944Pb0.910.550.7945S0.940.570.7846Si0.940.570.73	23	AnimalSource	0.90	0.85	0.87		
24pH0.930.560.8125EC0.960.650.8926Moisture0.970.660.9127TotalC0.940.60.8528TotalN0.940.630.8329CNRatio0.980.550.8230Al0.920.550.7331B0.940.580.7732Ca0.950.630.8433Cd0.980.920.9234Cr0.980.90.9235Cu0.940.60.8436Fe0.950.610.8737K0.970.550.938Mg0.950.690.8539Mn0.910.630.8140Mo0.980.70.9241Na0.960.650.8943P0.970.670.8944Pb0.910.550.7945S0.940.570.7846Si0.930.540.73		Physicochemical Prope	erties				
25 EC 0.96 0.65 0.89 26 Moisture 0.97 0.66 0.91 27 TotalC 0.94 0.6 0.85 28 TotalN 0.94 0.63 0.83 29 CNRatio 0.98 0.55 0.82 30 Al 0.92 0.55 0.73 31 B 0.94 0.58 0.77 32 Ca 0.95 0.63 0.84 33 Cd 0.98 0.55 0.92 34 Cr 0.98 0.65 0.92 35 Cu 0.94 0.6 0.84 36 Fe 0.95 0.61 0.87 37 K 0.97 0.55 0.9 38 Mg 0.95 0.69 0.85 39 Mn 0.91 0.63 0.81 40 Mo 0.96 0.68 0.88 42	24	рН	0.93	0.56	0.81		
26 Moisture 0.97 0.66 0.91 27 TotalC 0.94 0.6 0.85 28 TotalN 0.94 0.63 0.83 29 CNRatio 0.98 0.55 0.82 30 Al 0.92 0.55 0.73 31 B 0.94 0.58 0.77 32 Ca 0.95 0.63 0.84 33 Cd 0.98 0.55 0.92 34 Cr 0.98 0.65 0.92 35 Cu 0.98 0.65 0.92 36 Fe 0.95 0.61 0.84 36 Fe 0.95 0.61 0.87 37 K 0.97 0.55 0.9 38 Mg 0.95 0.69 0.85 39 Mn 0.91 0.63 0.81 40 Mo 0.96 0.65 0.89 41 Na 0.96 0.65 0.89 43 P 0.91	25	EC	0.96	0.65	0.89		
27 TotalC 0.94 0.6 0.85 28 TotalN 0.94 0.63 0.83 29 CNRatio 0.98 0.55 0.82 30 Al 0.92 0.55 0.73 31 B 0.94 0.58 0.77 32 Ca 0.95 0.63 0.84 33 Cd 0.98 0.65 0.92 34 Cr 0.98 0.65 0.92 35 Cu 0.98 0.90 0.92 36 Fe 0.98 0.90 0.92 35 Cu 0.94 0.63 0.84 36 Fe 0.95 0.61 0.87 37 K 0.97 0.55 0.9 38 Mg 0.95 0.69 0.85 39 Mn 0.91 0.63 0.81 40 Na 0.96 0.65 0.89 41 Na 0.96 0.65 0.89 42 Ni 0.91	26	Moisture	0.97	0.66	0.91		
28 TotalN 0.94 0.63 0.83 29 CNRatio 0.98 0.55 0.82 30 Al 0.92 0.55 0.73 31 B 0.94 0.58 0.77 32 Ca 0.95 0.63 0.84 33 Cd 0.98 0.65 0.92 34 Cr 0.98 0.90 0.92 35 Cu 0.94 0.65 0.92 34 Cr 0.98 0.90 0.92 35 Cu 0.94 0.65 0.92 36 Fe 0.95 0.61 0.87 37 K 0.97 0.55 0.9 38 Mg 0.95 0.69 0.85 39 Mn 0.91 0.63 0.81 40 Na 0.96 0.68 0.88 42 Ni 0.96 0.65 0.89 43 <	27	TotalC	0.94	0.6	0.85		
29 CNRatio 0.98 0.55 0.82 30 Al 0.92 0.55 0.73 31 B 0.94 0.58 0.77 32 Ca 0.95 0.63 0.84 33 Cd 0.98 0.65 0.92 34 Cr 0.98 0.65 0.92 35 Cu 0.94 0.6 0.84 36 Fe 0.98 0.9 0.92 35 Cu 0.94 0.6 0.84 36 Fe 0.95 0.61 0.87 37 K 0.97 0.55 0.9 38 Mg 0.95 0.69 0.85 39 Mn 0.91 0.63 0.81 40 Mo 0.96 0.68 0.88 42 Ni 0.96 0.65 0.89 43 P 0.97 0.67 0.89 44 Pb 0.91 0.55 0.79 45 S 0.94 0.57 <td>28</td> <td>TotalN</td> <td>0.94</td> <td>0.63</td> <td>0.83</td>	28	TotalN	0.94	0.63	0.83		
30 Al 0.92 0.55 0.73 31 B 0.94 0.58 0.77 32 Ca 0.95 0.63 0.84 33 Cd 0.98 0.65 0.92 34 Cr 0.98 0.90 0.92 35 Cu 0.94 0.6 0.84 36 Fe 0.95 0.61 0.87 37 K 0.97 0.55 0.9 38 Mg 0.95 0.69 0.85 39 Mn 0.91 0.63 0.81 40 Mo 0.98 0.7 0.92 41 Na 0.96 0.68 0.88 42 Ni 0.96 0.65 0.89 43 P 0.97 0.67 0.89 44 Pb 0.91 0.55 0.79 45 S 0.94 0.57 0.78 46 Si 0.93 0.54 0.73 47 Zn N.92 0.64	29	CNRatio	0.98	0.55	0.82		
31 B 0.94 0.58 0.77 32 Ca 0.95 0.63 0.84 33 Cd 0.98 0.65 0.92 34 Cr 0.98 0.9 0.92 35 Cu 0.94 0.6 0.84 36 Fe 0.95 0.61 0.87 37 K 0.97 0.55 0.9 38 Mg 0.95 0.69 0.85 39 Mn 0.91 0.63 0.81 40 Mo 0.98 0.7 0.92 41 Na 0.96 0.68 0.88 42 Ni 0.96 0.65 0.89 43 P 0.97 0.67 0.89 44 Pb 0.91 0.55 0.79 45 S 0.94 0.57 0.78 46 Si 0.93 0.54 0.73 47 Zn Nn 0.92 0.64 0.85	30	Al	0.92	0.55	0.73		
32 Ca 0.95 0.63 0.84 33 Cd 0.98 0.65 0.92 34 Cr 0.98 0.9 0.92 35 Cu 0.94 0.6 0.84 36 Fe 0.95 0.61 0.87 37 K 0.97 0.55 0.9 38 Mg 0.95 0.69 0.85 39 Mn 0.91 0.63 0.81 40 Mo 0.98 0.7 0.92 41 Na 0.96 0.68 0.88 42 Ni 0.96 0.65 0.89 43 P 0.97 0.67 0.89 44 Pb 0.91 0.55 0.79 45 S 0.94 0.57 0.78 46 Si 0.93 0.54 0.73 47 Zn N.92 0.64 0.85	31	В	0.94	0.58	0.77		
33 Cd 0.98 0.65 0.92 34 Cr 0.98 0.9 0.92 35 Cu 0.94 0.6 0.84 36 Fe 0.95 0.61 0.87 37 K 0.97 0.55 0.9 38 Mg 0.95 0.69 0.85 39 Mn 0.91 0.63 0.81 40 Mo 0.98 0.7 0.92 41 Na 0.96 0.68 0.88 42 Ni 0.96 0.65 0.89 43 P 0.91 0.55 0.79 44 Pb 0.91 0.55 0.79 45 S 0.94 0.57 0.78 46 Si 0.93 0.54 0.73 47 Zn 0.92 0.64 0.85	32	Са	0.95	0.63	0.84		
34 Cr 0.98 0.9 0.92 35 Cu 0.94 0.6 0.84 36 Fe 0.95 0.61 0.87 37 K 0.97 0.55 0.9 38 Mg 0.95 0.69 0.85 39 Mn 0.91 0.63 0.81 40 Mo 0.98 0.7 0.92 41 Na 0.96 0.65 0.89 42 Ni 0.96 0.65 0.89 43 P 0.91 0.55 0.79 44 Pb 0.91 0.55 0.79 45 S 0.94 0.57 0.78 46 Si 0.93 0.54 0.73 47 Zn 0.92 0.64 0.85	33	Cd	0.98	0.65	0.92		
35 Cu 0.94 0.6 0.84 36 Fe 0.95 0.61 0.87 37 K 0.97 0.55 0.9 38 Mg 0.95 0.69 0.85 39 Mn 0.91 0.63 0.81 40 Mo 0.98 0.7 0.92 41 Na 0.96 0.65 0.89 42 Ni 0.96 0.65 0.89 43 P 0.97 0.67 0.89 44 Pb 0.91 0.55 0.79 45 S 0.94 0.57 0.78 46 Si 0.93 0.54 0.73 47 Zn 0.92 0.64 0.85	34	Cr	0.98	0.9	0.92		
36 Fe 0.95 0.61 0.87 37 K 0.97 0.55 0.9 38 Mg 0.95 0.69 0.85 39 Mn 0.91 0.63 0.81 40 Mo 0.98 0.7 0.92 41 Na 0.96 0.68 0.88 42 Ni 0.96 0.65 0.89 43 P 0.97 0.67 0.89 44 Pb 0.91 0.55 0.79 45 S 0.94 0.57 0.78 46 Si 0.93 0.54 0.73 47 Zn 0.92 0.64 0.85	35	Cu	0.94	0.6	0.84		
37K0.970.550.938Mg0.950.690.8539Mn0.910.630.8140Mo0.980.70.9241Na0.960.680.8842Ni0.960.650.8943P0.970.670.8944Pb0.910.550.7945S0.940.570.7846Si0.930.540.7347Zn0.920.640.85	36	Fe	0.95	0.61	0.87		
38Mg0.950.690.8539Mn0.910.630.8140Mo0.980.70.9241Na0.960.680.8842Ni0.960.650.8943P0.970.670.8944Pb0.910.550.7945S0.940.570.7846Si0.930.540.7347Zn0.920.640.85	37	К	0.97	0.55	0.9		
39Mn0.910.630.8140Mo0.980.70.9241Na0.960.680.8842Ni0.960.650.8943P0.970.670.8944Pb0.910.550.7945S0.940.570.7846Si0.930.540.7347Zn0.920.640.85	38	Mg	0.95	0.69	0.85		
40Mo0.980.70.9241Na0.960.680.8842Ni0.960.650.8943P0.970.670.8944Pb0.910.550.7945S0.940.570.7846Si0.930.540.7347Zn0.920.640.85	39	Mn	0.91	0.63	0.81		
41Na0.960.680.8842Ni0.960.650.8943P0.970.670.8944Pb0.910.550.7945S0.940.570.7846Si0.930.540.7347Zn0.920.640.85	40	Мо	0.98	0.7	0.92		
42Ni0.960.650.8943P0.970.670.8944Pb0.910.550.7945S0.940.570.7846Si0.930.540.7347Zn0.920.640.85	41	Na	0.96	0.68	0.88		
43P0.970.670.8944Pb0.910.550.7945S0.940.570.7846Si0.930.540.7347Zn0.920.640.85	42	Ni	0.96	0.65	0.89		
44Pb0.910.550.7945S0.940.570.7846Si0.930.540.7347Zn0.920.640.85	43	Р	0.97	0.67	0.89		
45S0.940.570.7846Si0.930.540.7347Zn0.920.640.85	44	Pb	0.91	0.55	0.79		
46Si0.930.540.7347Zn0.920.640.85	45	S	0.94	0.57	0.78		
47 Zn 0.92 0.64 0.85	46	Si	0.93	0.54	0.73		
	47	Zn	0.92	0.64	0.85		

Soil_START (2-4 weeks old)=199, Soil_MID (5-7 weeks old)=313, Soil_END (8-11 weeks old)=183) from eleven pastured farms in the southeastern United States. DNA was extracted from all the samples and the V4 domain of the bacterial 16 S rRNA gene was amplified. After sequencing and preliminary analysis using the QIIME pipeline, 1,823 microbiome operational taxonomic units (OTUs) were identified. During the data preprocessing stage, the dimension of the OTUs were reduced to 877 unique genera. These microbiome genera were used as input to predict the prevalence of 3 important foodborne pathogens (Salmonella, Campylobacter and Listeria) that were culturally isolated, levels of physicochemical properties (acidity/basicity (pH), electrical conductivity (EC), moisture, total carbon (TotalC), total nitrogen (TotalN), carbon to nitrogen ratio (CNRatio), aluminum (Al), boron (B), calcium (Ca), cadmium (Cd), chromium (Cr), copper (Cu), iron (Fe), potassium (K), magnesium (Mg), manganese (Mn), molybdenum (Mo), sodium (Na), nickel (Ni), phosphorus (P), lead (Pb), sulphur (S), silicon (Si), zinc (Zn)) and changes in farm management practices associated with the samples as machine learning targets. The 32 farm management practices (AvgNumBirds, AvgNumFlocks, YearsFarming, EggSource, BroodBedding, BroodFeed, BrGMOFree, BrSoyFree, BrMedicated, BroodCleanFrequency, AvgAgeToPasture, PastureHousing, FreqHousingMove, AlwaysNewPasture, PastureFeed, PaGMOFree, PaSoyFree, PaMedicated, LayersOnFarm, CattleOnFarm, SwineOnFarm, GoatsOnFarm, SheepOn-Farm, WaterSource, FreqBirdHandling, AnyABXUse, LengthFeedRestrixProcess, Seasons, FlockAgeDays, Breed, FlockSize, AnimalSource) have been previously described [14, 15].

Model evaluation

Cross-validation is a generally accepted resampling technique suitable for machine learning model evaluation as it trains on different parts of the data and evaluates prediction accuracy on the rest of the data in many iterations. Out of the three algorithms tested with microbiome relative abundance as the input and different pathogens, farm practices and physicochemical properties as targets, random forest performed optimally the best with the average 5-fold cross-validation of 0.83, followed by logistic regression with 0.76 and support vector machine with 0.65 average performance (data not shown). With 59 target variables, 47 models performed relatively well with RF, above 70% confidence threshold with 5-fold cross-validation (Table 1) while 12 models performed poorly with less than 70% with 5-fold crossvalidation (Supplementary Table 1). Definitions of these variables have been previously published [14, 15].

Feature importance

Using the random forest feature importance package in scikit-learn library (v0.24.2), which is based on impurity decrease within each decision tree, we identified the top 10 important microbiome genera associated with each target variable. A representative plot showing the top 10 important genera out of a total of 877 genera that are associated with *Campylobacter* positive samples (as determined using traditional cultural protocols) is shown in Fig. 1. The corresponding area under the receiver operating characteristic curve (AUROC) is indicated on the plot and any prediction model with AUROC<0.70 was discarded.

Correlation of pathogens and probiotics

After successfully classifying the top 10 genera that predict the presence of Salmonella spp., Campylobacter spp. and Listeria spp. in different samples, we focused on genera among the top 10 that are identified in literature as probiotics such as Bacillus, Bifidobacterium, Clostridium, Enterococcus, Lactobacillus, Pediococcus, Propionibacterium and Streptococcus for further analysis. We used the Pearson statistical module in SciPy package (v1.7.0) to identify a significant relationship (r) between the identified probiotics and pathogens (p < 0.05). This correlation analysis is expected to yield useful information on potentially suitable probiotic taxa within pastured poultry management systems. No significant negative associations were found between Salmonella- and Listeria-positive samples and any of the known probiotics and no relationships were found between probiotics and the pathogens of interest in the soil sample type. Therefore, all subsequent analyses focused on Campylobacter-positive feces samples. Campylobacter-positive feces samples were negatively correlated with both probiotics *Bacillus* (r = -0.77, p-value= 3×10^{-21}) and *Clostridium* (r = -0.82, p-value= 7.0×10^{-26}) and positively correlated with *Lac*-tobacillus (r=0.83, p-value= 5.0×10^{-27}).

Following the identification of relationships between *Campylobacter*-positive samples and probiotic taxa *Bacillus, Clostridium* and *Lactobacillus* in the fecal microbiomes of those samples, we used RF to build new models based on the time of sample collection to identify the time point at which these correlations actually occur as this information will help identify at what flock age potential probiotic application would be most impactful. Using seaborn package (v 0.10.1), our observation is presented as the heatmap in Fig. 2.

Correlation of Campylobacter and other genera

The correlations between Campylobacter-positive fecal samples and other "non-probiotics" taxa identified among the top 10 in feces at different time points of collection using RF, its feature importance package and Pearson correlation analysis (Fig. 3) are discussed in this section. While taxa that are negatively correlated with *Campylobacter* are potentially novel probiotics, a positive correlation of a particular genus could serve as an indicator species that forewarns the presence of Campylobacter. We identified genera such as Dorea, Faecalibacterium, Parabacteroides and Solibacillus that are negatively correlated with Campylobacter at the START phase and Caloramator, DA101, Proteus, Rumellibacillus and Veillonella that are also negatively correlated at the END production phase. The only correlation identified at the MID phase belongs to taxa with uncertain genera. We did not identify any positive correlation between



Fig. 1 Representative feature importance plots showing the top 10 genera (including probiotic genera) that are important for the prediction of *Campylobacter* in the feces samples when the flock is first introduced onto the pasture (Feces_START). The plot ranks the genus with the highest importance score as 100



Fig. 2 Heatmap showing the correlation of probiotic taxa with *Campylobacter*-positive feces at different stages of pastured poultry production (START, MID, END). Strong positive correlations ($r \ge 0.7$, p-value < 0.05) are depicted in different shades of green while strong negative correlations ($r \le -0.7$, p-value < 0.05) are shown in different shades of red. A strong negative correlation is observed with both *Bacillus* and *Clostridium* taxa in *Campylobacter*-positive feces at the second time point (MID) of feces collection. Probiotic taxa *Lactobacillus* is positively correlated to *Campylobacter*-positive feces at both START and MID time points of sample collection



Fig. 3 Heatmap showing the correlation between "non-probiotics" microbiome taxa and *Campylobacter*-positive feces samples. All genera observed in the analysis have a negative correlation with *Campylobacter*-positive feces samples at the start and end-time point of sample collection. The only negative correlation identified at the mid-time point belongs to the group of OTUs that could not be identified at the genus level (g_). Strong negative correlations ($r \le -0.7$, p-value < 0.05) are depicted in red

the "non-probiotics" taxa and *Campylobacter* that could serve as indicators.

Correlation of farm management practices, feces physicochemical properties and probiotics

Once we established that there is a correlation between the microbiome (probiotics and other non-probiotics genera) with Campylobacter, we next identified farm management practices and feces physicochemical properties that are associated with these microbiome signatures that correlate with Campylobacter by building RF models. We used the relative abundance for all taxa at the genus level of the microbiome genera from 698 feces samples as input to RF models with different farm management practices and feces physicochemical properties as the individual targets. Using random forest feature importance, Pearson correlation, and scikit-learn seaborn package, we identified associations between known probiotics and potential probiotic taxa (identified in the previous section) and farm management or physicochemical properties (Figs. 4 and 5, respectively). Predicting associations between probiotic taxa and Campylobacter prevalence, based on cultural identification (Figs. 2 and 3) as well as probiotic taxa and farm management or fecal physicochemical properties (Figs. 4 and 5, respectively) from the same input microbiome could identify management practices or physicochemical properties that could be modified to optimize a healthy microbiome and selectively target Campylobacter and reduce its prevalence. We observed that GMO free feed and presence of sheep on the farm have a negative correlation with the level of Bacillus while GMO free feed and having cattle on the farm have a negative correlation with Enterococcus (Fig. 4). Keeping brood and pasture feeds free of soy products, especially at the START and MID phases of production could potentially increase the abundance of Clostridium, Lactobacillus, Pediococcus and Streptococcus. Always moving the chicken to new pasture without rotating back to previously used farm lots, putting chicks on pasture later at 4 weeks instead of 3 weeks, having goats on the farm, and feeds that are GMO free could all negatively impact the abundance of Lactobacillus. The presence of swine and layers on the farm and frequently moving pasture housing everyday instead of every 2 days may negatively impact the abundance of Streptococcus. Some farm management practices such as AvgAgeTo-Pasture and SheepOnFarm have both negative and positive correlations, depending on the probiotic genus. For example, AvgAgeToPasture has a negative correlation with Lactobacillus but a positive correlation with Bacillus. GMO free diets, moving the pasture house daily and having other animals such as swine on the farms appears to not only negatively impact common probiotics but the



Fig. 4 Heatmap showing the correlation of common/potential probiotic taxa associations with diverse farm management practices. Strong positive correlations ($r \ge 0.7$, p-value < 0.05) are shown in different shades of green while strong negative correlations ($r \le -0.7$, p-value < 0.05) are shown in different shades of green while strong negative correlations ($r \le -0.7$, p-value < 0.05) are shown in different shades of red. There is a negative correlation between most known probiotics and farm practices such as AlwaysNewPasture, FreqHousingMove, PaG-MOFree, CattleOnFarm and SwineOnFarm and a positive correlation with BrSoyFree, PaSoyFree and GoatOnFarm. Similar to our observation with known probiotics, FreqHousingMove, PaGMOFree and SwineOnFarm have a negative correlation with potential probiotics

potential probiotics (*DA101, Dorea, Parabacteroides* and *Solibacillus*) as well (Fig. 4).

Looking at the relationship between physicochemical properties and probiotics, four elements identified (Cd, Mo, P, S) and pH have a negative correlation with known probiotics, mostly *Lactobacillus*, while three elements (Cd, Cu, Mn) have a negative correlation with potential probiotics, especially *g_Faecalibacterium*, which has a negative association with all the three elemental metals (Fig. 5). All association between physicochemical elements of feces are negatively correlated ($r \leq -0.7$, *p*-value < 0.05) to both the common probiotics and potential probiotics. Cadmium (Cd) in particular negatively impacts *Lactobacillus*, *Streptococcus*, and *g_Faecalibacterium*, while copper (Cu) also has a negative correlation with *g_Faecalibacterium* as well as *g_Parabacteroides*.

Discussion

With the growing capability of artificial intelligence to find patterns in large datasets, machine learning has found applications in the biomedical research in diverse areas such as cancer detection, drug development, treatment recommendation and prediction of disease outcome [16, 17]. Machine learning is currently being explored for a quick turnaround in finding association between disease states, causative agents and the grey areas in between. Traditional biological research to elucidate such associations usually takes years, requires exorbitant funding and is oftentimes laborious. Interactions between bacterial pathogen (*Yersinia pestis, Bacillus anthracis* and *Francisella tularensis*) proteins and human proteins have been successfully predicted by machine learning [18]. Relationships between microbiome dysbiosis and diseases such as ulcerative colitis, obesity and amyotrophic lateral sclerosis in humans has been established [19–21]. Traditional machine learning algorithms, such as support vector machine, random forest and logistic regression, have been used to identify available microbiome features like relative abundance that is linked to obesity and diabetes [22]. Random forest, a popular machine learning algorithm, has been used to identify structural features of beta-glucans, making it suitable for developing prebiotics [23].

Probiotics which are microorganisms from a healthy host are being considered in the food safety industry as they are carefully chosen for their non-pathogenic nature and ability to confer health benefits to the new host they are administered to. Although the exact mechanism of action is not well understood and often is strain dependent, improving host nutrient absorption, secretion of toxins such as bacteriocin, altering pH of the intestinal lumen, production of short-chain fatty acids, modulation of the immune response and competing for nutrients are some of the ways in which probiotics have been proposed to competitively exclude the pathogens and benefit both humans and animals [24, 25]. While the weight gain benefit of probiotics in livestock has been reported [26], their ability to reduce contamination of pathogens



Fig. 5 Heatmap showing the correlation between common/potential probiotic taxa associations with physicochemical properties in feces. All association between physicochemical elements of feces are negatively correlated ($r \le -0.7$, p-value < 0.05) to the common probiotics and potential probiotics

such as *Salmonella* spp. in chickens, *E. coli* in pigs and *Aeromonas salmonicida* in fish and improved food safety has also been recorded [27]. The benefits of probiotics in conventional poultry have been documented and a similar positive effect has been proposed in the alternative poultry production systems [10]. However, baseline level and dynamics of these useful microorganisms which are part of the natural microflora have not been established in pastured poultry.

In this study, microbiome relative abundance values were used as input to the RF algorithm to predict the presence of enteric poultry foodborne pathogen species (Salmonella, Campylobacter and Listeria), find possible relationships with known probiotic genera (Bacillus, Bifidobacterium, Clostridium, Enterococcus, Lactobacillus, Pediococcus, Propionibacterium, Streptococcus [28, 29]) and identify novel genera that represent potential probiotics [30] in a time dependent manner based on different stages of microbiome sample collection across poultry production phases. Samples from which the relative abundance was generated include feces as a representation of the gut microbiome in pastured poultry and soil as a representation of communities of microbes in the poultry environment. Existing literature on microbiome analysis usually focus on alpha diversity, such as measures of microbiome richness and evenness within samples, and beta diversity, dissimilarities among microbiome samples to identify prevalent OTUs or attribute them with certain disease conditions [31, 32]. This approach is inadequate to answer the question about the correlation between enteric pathogens and those found in the poultry environment with respect to the level of the probiotics within the microbiome as an innate defense system. Here, using the machine learning approach, we wanted to establish if microbiome data could be used successfully to predict pathogen prevalence, and if so, is there a correlation between the pathogens and the probiotics contained in the microbiome. A similar approach using random forest and relative abundance feature has been used to find association between the microbiome and disease conditions such as cirrhosis, colorectal cancer, diabetes and obesity [22]. Furthermore, we explored the relationship between the probiotics and different farm management practices and physicochemical properties. We hypothesized that certain practices and/or properties could be altered to enhance the abundance of probiotics and potentially reduce pathogen prevalence.

Of the three poultry foodborne pathogens (*Salmonella*, *Campylobacter* and *Listeria*) and two sample types (feces and soil) specifically targeted in this study, the machine learning models were only able to identify significant associations for *Campylobacter*-positive feces samples. *Campylobacter*, the zoonotic causative agent of campylobacteriosis and gastroenteritis, was the predominant

foodborne pathogen identified in all the eleven pastured poultry farms sampled in this study with negative correlations to both known and potential probiotics (Figs. 2 and 3). Campylobacter is a thermophilic bacterium that colonizes chicken due to their higher body temperature. It is often found in farmhouses, poultry house water and even capable of surviving on slaughtering equipment despite sanitization. Campylobacter has a high infection rate, with about 95% of birds becoming infected within 4 to 7 days after colonization of the first broiler [33]. Due to controlled use of antibiotics to curb the growing trend of antibiotic resistance and failure of several physical biosecurity and hygiene measures to prevent Campylobacter contamination of poultry product [34], vaccination, bacteriocins, bacteriophage cocktails and probiotics administrations are being proposed as alternatives to control this pathogen [35]. Supplementation with probiotics, which are a natural part of the chickens' microbiome, is being suggested as the most viable alternative as they could not only competitively exclude the pathogen but also enhance immune response and improve digestibility and nutrient absorption [36, 37]. We observed that the prevalence of *Campylobacter* detection by the cultural isolation method is higher in both feces and soil samples when the poultry chickens were young and old than at mid-age of their lifetime (Supplementary Fig. 1), suggesting that age may play a role in colonization by Campylobacter spp. Colonization of chicken with Campylobacter has been reported to start in the first three to four weeks of life [38]. And similar to our observation, Babacan et al. showed higher Campylobacter colonization towards the old-age (around 7 to 8 weeks) than at mid-age (5 to 6 weeks) [39]. The interpretation of this finding is that probiotics containing these taxa could be administered when the chicks are newly introduced to the pasture when their gut microbiomes are still maturing and re-applied closer to slaughter time to potentially control Campylobacter before entering the processing/post-harvest stage. Identifying the best time of probiotic application will also invariably reduce production costs in terms of the amount of probiotics required to be administered. Interestingly, our analysis identifies a strong negative correlation between probiotics Bacillus and Clostridium and Campylobacter at mid-age of the chickens (Fig. 2). Based on these findings, we recommend administration of existing probiotics, especially Bacillus and Clostridium, preferably at the early age of the chicks and late age as a suitable strategy to reduce Campylobacter contamination in poultry.

Lactobacillus, Bacillus and Bifidobacterium species as probiotics have been previously used to inhibit the growth and reduce the virulence of Campylobacter [40]. Our analysis predicts negative associations between Campylobacter and common probiotics Clostridium and Bacillus as expected, however it identified a positive association with Lactobacillus. Although most Lactobacillus species act as probiotics, certain species such as L. rhamnosus have been positively associated with several infections [41]. Since the microbiome analysis in this study could only identify the taxa down to the genus level, this could possibly explain the positive association between Campylobacter and an unknown strain of Lactobacillus that could be pathogenic such as L. rhamnosus, an indicator species for Campylobacter. In addition, efforts to identify other microorganisms, other than the conventional ones, that could be beneficial as probiotics are ongoing in agriculture and food industry [42]. In aquaculture, Lysinibacillus macrolides was recently shown to be a potential probiotic as it significantly helped to improve the growth rate and weight gain of Cyprinus carpio fish [30]. In this study, we identified 9 different genera such as Caloramator, DA101, Dorea, Faecalibacterium, Parabacteroides, Proteus, Rumellibacillus, Solibacillus and Veillonella in a feces model (Fig. 3) that could potentially be considered as probiotics against Campylobacter. Existing literature shows that *Caloramator* is closely related to probiotic *Clostridium* as they both belong to the same family, Clostridiaceae [43]. Also, positive roles of Caloramator in regulating inflammation and the immune system and promotion of gut health have been described in broiler chickens [44]. Supplementation with probiotic Bacillus increased the proportion of DA101 and significantly improved the performance of the broiler chickens [45], suggesting that genus *DA101* could also potentially be a probiotic. Faecalibacterium was recently shown as both a biomarker of obesity and a probiotic in treatment of this condition [46] while Parabacteroides has been proposed as a next-generation probiotic [47]. The potential of Solibacillus as a probiotic against pathogens such as Aeromonas and Pseudomonas has been adequately demonstrated [48]. Additionally, the ability of Veillonella to effectively inhibit the growth of Salmonella enteritidis by producing acetate and propionate intermediates has also been reported [49]. In summary our model has identified not only known probiotics to target as potential interventional strategies against Campylobacter, but also discovered other taxa with known probiotic potential for consideration in future studies to validate their beneficial role in commercial poultry.

In evaluation of different farm practices that could impact the abundance of probiotic taxa in the pastured poultry gut microbiome (as represented by feces samples), farm management variables had a greater impact on the known probiotic taxa than the identified potential probiotics (Fig. 4). It should be stated that the variability of management practices on pastured poultry farms is much greater than what is observed on conventional poultry farms, with the broiler flocks having much greater interaction with the farm environment and the other animals being raised on those farms. Out of the eight probiotics commonly used in poultry, only Bacillus, Enterococcus, Lactobacillus and Streptococcus are mostly affected by the farm activities in this study. We observed that the presence of sheep and a GMO-free diet have a negative correlation with the relative abundance of Bacillus within the broiler fecal microbiome, while having layers on the farm was associated with higher Bacillus abundance (Fig. 4). Our model suggests that soy-free diets have significantly positive correlation with the abundance of Clostridium, Lactobacillus, Pediococcus and Streptococcus in the fecal microbiome. While a soy-free diet may improve the abundance of probiotics, it has also been shown to reduce the microbiome relative abundance of pathogens such as Campylobacter and Acinetobacter [50]. GMO-free feed and the presence of cattle on the farm negatively impact Enterococcus abundance while having goats on the farm and delaying the age chickens are put on pasture from 3 weeks to 4 weeks could increase its abundance. Similar to previous observation, goats on the farm appear to improve the abundance of Streptococcus while layers and swine on the farm reduce the abundance of Streptococcus within the poultry fecal microbiome. While the mechanisms are not clearly understood at this time, the presence of goats with the poultry, a soy-free diet and increasing the age to put chickens on pasture from 3 weeks to 4 weeks should help improve the abundance of these probiotics. Having swine and cattle on the farm, a GMO-free diet and daily moving of pasture houses may be discouraged as they do not just negatively impact the abundance of the known probiotics alone but the potential novel probiotics as well.

Out of the 24 physicochemical properties modeled as machine learning targets, we observed that the presence of six elements (Cd, Cu, Mn, Mo, P, and Si), especially Cd, Cu and Si, have a varying degree of negative impacts on the relative abundance of probiotic taxa in the fecal microbiome data (Fig. 5). Literature pertaining to the exact mechanism of action of these elements in modulating the probiotics activities is very limited. The ability of Lactobacillus and Streptococcus as probiotics to effectively bind Cd as a heavy metal and source of oxidative stress and reduce its toxicity has been demonstrated in both mice and rats [51-53]. Cd has been shown to significantly reduce Lactobacillus and microbiome abundance in general [54]. Similar to our finding here, Zhang and colleagues show that an increased level of Cu decreases the abundance of genus Parabacteroides in the rat microbiome [55] while the ability of Faecalibacterium in the microbiome to detoxify arsenic, often found in complexes with Cu, has been demonstrated [56]. In contrast, Si, with its antioxidant properties, is reported to have positive correlation with beneficial Lactobacillus reuteri and *Lactobacillus murinus* [57]. Si has a negative correlation with *Clostridium* and *Lactobacillus* in our dataset, although specific *Lactobacillus* species that have negative correlation could not be ascertained.

Conclusion

In this study we describe a combinatorial approach using machine learning and microbiome relative abundance for predicting pathogens prevalence, identifying probiotics (Bacillus and Clostridium) suitable for control of Campylobacter, validating potential probiotics candidates (Caloramator, DA101, Dorea, Faecalibacterium, Parabacteroides, Proteus, Rumellibacillus, Solibacillus and Veillonella) and identifying farm management practices and farm physiochemical properties that may affect the probiotic strains that could limit the spread of Campylobacter along the pastured poultry farm-to-fork continuum. Although we initially considered checking for associations between microbiome taxa and different individual farms, the total number of samples collected is not evenly distributed among individual farms, making it difficult to apply machine learning algorithms such as RF to effectively train and learn from individual farms. Therefore, in this study, we combined all samples into a single comprehensive dataset. However, we acknowledge that Farm "A" and "I," which account for 18 out of 42 flocks (i.e., 43% of the total flocks) studied (Table 2), could have correlated observations that were not considered by RF due to the assumption of independence. Despite this limitation, it is important to note that while the total number of samples in this animal study (1393 samples) may be fewer than those reported in similar human studies (4347 samples [12] and 2424 samples [22]), our results offer valuable insights and generate hypothesis regarding

Table 2 Comparison of the 11 antibiotic-free pastured broiler farms in this study

Farm	Breed ¹	No. of flocks	Flock size	Multi- use farm?	Animal type
A	FR	10	>500	Yes	Layers, swine, cattle, sheep
В	FR, CC	5	< 50	Yes	Layers, swine, goats
С	FR	1	< 50	No	NA
D	FR	1	< 50	No	NA
E	FR, CC	5	50-100	Yes	Layers, swine, cattle, sheep
Н	FR	2	>500	Yes	Layers
I.	FR, CC	8	100-500	Yes	Layers, swine, goats
J	FR, CC	2	50	Yes	Layers
К	FR	4	100-500	Yes	Layers, cattle, goats
L	FR	2	>500	Yes	Layers, swine, cattle, sheep
Μ	CC	2	50-100	Yes	Layers, swine

¹ FR=Freedom Ranger, CC=Cornish Cross

microbiome signatures that could reduce pathogen prevalence and the impact of farm practices on these potential beneficial microorganisms for food safety. These results demonstrate the utility of this combinatorial approach with microbiome datasets for predicting microbiome signatures that could be targeted for developing "all natural" probiotic-based interventions for reduction of pathogens within pastured poultry management systems.

Materials and methods

Flock management, sample collection and preparation

This study was conducted in eleven pasture-raised broiler farms located in the southeastern United States. Flock sizes range between 25 and over 1500 with Freedom Ranger and Cornish Cross breeds being placed in different farms (Table 2). Soil and feces samples were collected from the pasture where the flock was currently residing at the time of sampling. Samplings occurred within a few days of being placed in the pasture (START=2-4 weeks old), halfway through their time on pasture (MID=5-7 weeks old), and on the day the flock was processed (END=8–11 weeks). At each sampling time, the pasture area was divided into five separate sections and five subsamples in each section were pooled into a single sample for each section (a total of five soil samples and five feces samples were collected on each sampling day). Soil samples were collected from the surface (0-7 cm) with sterile scoops, and feces samples were collected from fresh droppings on the soil surface. Gloves and scoops were changed between sample types and between sampling areas. To prepare the environmental samples for homogenization, 3 g (feces, soil) were combined within filtered stomacher bags (Seward Laboratory Systems, Davie, FL, USA), and diluted 1:3 using 10 mM phosphate-buffered saline (PBS). All fecal and soil samples were transported back to the laboratory on ice and processed within 2 h of collection. All samples were homogenized for 60 s and these homogenates were used for all downstream cultural isolations and DNA extraction.

Cultural isolation methods

To determine the foodborne pathogen status for each fecal and soil sample in the study, the following traditional cultural isolation protocols were used.

Salmonella spp

As a pre-enrichment step, the stomached homogenates remained in the filtered stomacher bags and were incubated overnight at 35 °C. Two different enrichment broths were used to isolate *Salmonella spp.* from these environmental samples: Tetrathionate (TT; Becton Dickinson, Sparks, MD, USA) broth and Rappaport-Vassiliadis (RV; Becton Dickinson, Sparks, MD, USA) media. After overnight incubation at 42 °C in both of these

enrichment broths, one loopful from each enrichment broth was spread on two different differential media: Brilliant Green Sulfa with novobiocin (BGS; Becton Dickinson, Sparks, MD, USA) agar and xylose lysine tergitol-4 (XLT-4; Becton Dickinson, Sparks, MD, USA) agar. These plates were incubated overnight at 35 °C, and on each plate, 3 Salmonella-like colonies per subsample were picked and confirmed using triple sugar iron agar (TSI; Becton Dickinson, Sparks, MD, USA) and lysine iron agar fermentation (LIA; Becton Dickinson, Sparks, MD, USA) using an incubation period of 18-24 h at 35 °C. Final confirmation of suspect TSI/LIA isolates was performed using Salmonella polyvalent O antiserum agglutination (Becton Dickinson, Sparks, MD, USA), using the manufacturer's specifications. Positive Salmonellae were serogrouped using individual Salmonella poly O antisera for O groups A through I, following the Kauffman-White scheme [58].

Campylobacter spp

Recovery of *Campylobacter* spp. from homogenized samples was performed as previously described [59]. Initially, 100 mL of homogenized suspension was removed, plated onto Campy–Cefex agar and incubated at 42 ± 1 °C for 36 to 48 h in a microaerobic atmosphere (5% O₂, 10% CO₂, 85% N₂). Putative *Campylobacter spp.* colonies were enumerated, and up to five colonies per sample were subcultured on Brucella agar supplemented with 10% lysed horse blood (BAB plates) for isolation and incubated as previously described.

Listeria spp

As a pre-enrichment step, the stomached homogenates remained in the filtered stomacher bags and were incubated overnight at 35 °C. This pre-enrichment was followed by two subsequent enrichments in UVM Modified Listeria Enrichment Broth (UVM, Becton Dickinson, Sparks, MD, USA) and Fraser Broth (Becton Dickinson, Sparks, MD, USA), both requiring an overnight incubation period at 30 °C. One loopful of the Fraser enrichment was streaked for isolation on *Listeria*-selective agar (Becton Dickinson, Sparks, MD, USA). These plates were incubated overnight at 30 °C, and on each plate, 3 *Listeria*-like colonies per positive subsample were picked and confirmed as *Listeria* using the appropriate BAX PCR assay (DuPont, Wilmington, DE, USA).

Fecal and soil physiochemical analysis

The moisture content of the fecal and soil samples was determined by drying overnight at 65 °C and calculating the difference between the wet and dried weights of the litter. Fecal and soil pH and electrical conductivity (EC) were determined using an Orion Versa Star Advanced Electrochemistry Meter (Thermo Fisher Scientific, Waltham, MA, USA) and using a 1:5 dilution in distilled water. Fecal and soil samples were submitted to the University of Georgia Soils Testing Laboratory for Total C, Total N, and elemental (Al, B, Ca, Cd, Cr, Cu, Fe, K, Mg, Mn, Mo, Na, Ni, P, Pb, S, Si, Zn) composition.

DNA extraction and quantification

DNA extractions were performed on 0.33 g of feces and 0.33 g of soil. DNA was extracted from samples according to a semi-automated hybrid DNA extraction protocol previously described [60]. This method was a combination of a mechanical method using the FastDNA Spin Kit for Feces (MP Biomedicals, Solon, OH, USA) and an enzymatic method based on the QIAamp DNA Stool Mini Kit (QIAGEN, Valencia, CA, USA). DNA purification was performed using the DNA Stool—Human Stool—Pathogen Detection Protocol of the QIAcube Robotic Workstation. After purification, the DNA concentration in each sample was determined spectrophotometrically using the Take3 plate in conjunction with the Synergy H4 multimode plate reader (BioTek, Winooski, VT, USA).

Illumina MiSeq library construction and analyses

Library construction and sequencing were performed by the Earth Microbiome Project Laboratory at the U.S. Department of Energy, Argonne National Laboratory (Argonne, IL, USA). In short, the hypervariable V4 domain of the bacterial 16 S rRNA gene was amplified using the F515 (5'-CACGGTCGKCGGCGCCATT-3') and R806 (5'-GGACTACHVGGGTWTCT AAT-3') primer set with each primer containing Illumina adapter regions (Illumina, Inc., San Diego, CA, USA) and the reverse primer containing the Golay barcodes to facilitate multiplexing [61]. Raw reads were obtained by using the Illumina MiSeq platform. A total of 3,297,242 raw sequence reads were generated and processed by the QIIME v1.9.1 (Quantitative Insights Into Microbial Ecology) pipeline [62]. Quality filtering and library splitting according to the Golay barcode sequences were performed on the R1 read (split_library_fastq.py script, default parameters). Sequences were chimera checked against the gold.fa database (http://drive5.com/uchime/ gold.fa) and clustered into Operational Taxonomic Units (OTUs) according to their sequence similarity (97%) using the usearch option [63] with pick_otus.py script (-m usearch, all other parameters were default). A representative sequence for each OTU was selected with pick_rep_set.py script (using the most abundant method for picking, all other parameters were default) and used for taxonomic assignment using UCLUST and the Greengenes 13_8 database [64] with assign_taxonomy. py (default parameters). Sequences were aligned (align_ seqs.py script, default parameters) using PyNAST [65]

and filtered (filter_alignment.py, default parameters). A phylogenetic tree was subsequently produced with the make_phylogeny.py script (with default settings and Fast-Tree program). A total of 1,823 OTUs were identified among all the 1,393 samples collected (Feces_Start=200, Feces_Mid=313, Feces_End=185, Soil_Start=199, Soil_Mid=313, Soil_End=183). During the data preprocess-ing stage, the dimensions of the OTUs were reduced to 877 with unique genera.

Model evaluation and prediction

In this study, we evaluated the performance of three machine learning algorithms (random forest (RF), support vector machine (SVM) and logistic regression (LogReg)) on our microbiome dataset as input with foodborne zoonotic poultry pathogens (Salmonella, Campylobacter and Listeria), farm management practices and physicochemical properties as targets. We dropped only microbiome genera columns belonging to Salmonella, Campylobacter and Listeria and selected all other microbiome genera as input features for our model building. Multiple binary classifiers were built separately for model prediction and evaluation. We used 5-fold cross-validation for all the models to train, evaluate and predict the data sampled from the dataset. In 5-fold cross-validation, the dataset is divided into five equal subsets (train on four subsets and evaluate on the remaining subset) and then does so repeatedly five times, changing the subset used for evaluation. This is done to reduce the impact of data variability and get more accurate representation on how well the model will perform on unseen data. The average prediction accuracy is computed by scoring metric of individual estimator algorithms (RF, LogReg, and SVM) in scikit-learn (v0.24.2) using cross-val-score function. AUCROC as a measure of trade-off between the "true positive rate" (sensitivity) and the "false positive rate" (1 - specificity) in class prediction was computed using roc auc score function in scikit-learn. AUROC with a value closer to 1.0 indicates a higher level of performance for an overall classification model. Data analysis, statistical tests and visualization were performed in Jupyter Notebook environment (v 6.0.3). Our analyses were performed using Python (v3.8.3), scikit-learn (v0.24.2), pandas (v1.2.5) and SciPy (v1.7.0) and results were presented by seaborn (v 0.10.1) heatmaps. Target variables include 3 pathogens (Salmonella, Campylobacter and Listeria), 24 physicochemical properties (pH, EC, Moisture, TotalC, TotalN, CNRatio, Al, B, Ca, Cd, Cr, Cu, Fe, K, Mg, Mn, Mo, Na, Ni, P, Pb, S, Si, Zn) and 32 farm management practices (AvgNumBirds, AvgNumFlocks, YearsFarming, EggSource, BroodBedding, BroodFeed, BrGMOFree, BrSoyFree, BrMedicated, BroodCleanFrequency, AvgAgeToPasture, PastureHousing, FreqHousingMove, AlwaysNewPasture, PastureFeed, PaGMOFree, PaSoyFree, PaMedicated, LayersOnFarm, CattleOnFarm, SwineOn-Farm, GoatsOnFarm, SheepOnFarm, WaterSource, Freq-BirdHandling, AnyABXUse, LengthFeedRestrixProcess, Seasons, FlockAgeDays, Breed, FlockSize, AnimalSource) as previously described [14, 15].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42523-023-00260-w.

Supplementary Material 1	
Supplementary Material 2	
Supplementary Material 3	
Supplementary Material 4	
Supplementary Material 5	
Supplementary Material 6	
Supplementary Material 7	
Supplementary Material 8	

Acknowledgements

Data used in this study was generated as part of the Agricultural Research Service, USDA CRIS Project "Reduction of Invasive *Salmonella enterica* in Poultry through Genomics, Phenomics and Field Investigations of Small Multi-Species Farm Environments" #6040-32000-011-00-D. Mr. Leslie A. Shack of Mississippi State University College of Veterinary Medicine provided additional revision of the manuscript.

Author contributions

MR conceived and conducted data collection. MBA and NP performed the data analysis and visualization, and MBA drafted the manuscript. MR, BN, and MJR assisted with data interpretation and finalized the draft. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Agricultural Research Service, USDA NACA project entitled "Advancing Agricultural Research through High Performance Computing" #58-0200-0-002.

Data Availability

All data analyzed during this study are included in this published article (and its supplementary information files).

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

We did not handle any live birds or sample anything from live birds. Live animals were managed and handled solely by the collaborating farmers. We only sampled soil and fecal droppings during live production. Therefore, no IACUC was required.

Consent for publication

Not Applicable.

Author details

¹Geosystems Research Institute, Mississippi State University, Starkville, MS 39762, USA

²Department of Comparative Biomedical Sciences, College of Veterinary Medicine, Mississippi State University, Starkville, MS 39762, USA ³Department of Computer Science and Engineering, Mississippi State University, Starkville, MS 39762, USA ⁴Egg Safety and Quality Research Unit, USDA-ARS U.S. National Poultry Research Center, Athens GA 30605, USA

Received: 18 February 2023 / Accepted: 23 August 2023 Published online: 15 November 2023

References

- Whitton C, Bogueva D, Marinova D, Phillips CJC. Are we approaching Peak Meat Consumption? Analysis of meat consumption from 2000 to 2019 in 35 countries and its relationship to gross domestic product. Animals: An open Access Journal from MDPI 2021, 11(12).
- Diaz Carrasco JM, Casanova NA, Fernández Miyakawa ME. Microbiota, Gut Health and Chicken Productivity: What Is the Connection? *Microorganisms* 2019, 7(10).
- Berrang ME, Cox NA, Frank JF, Buhr RJ. Bacterial penetration of the Eggshell and Shell membranes of the Chicken hatching egg: a review. J Appl Poult Res. 1999;8(4):499–504.
- Ding J, Dai R, Yang L, He C, Xu K, Liu S, Zhao W, Xiao L, Luo L, Zhang Y, et al. Inheritance and establishment of gut microbiota in chickens. Front Microbiol. 2017;8:1967.
- Pan D, Yu Z. Intestinal microbiome of poultry and its interaction with host and diet. Gut Microbes. 2014;5(1):108–19.
- Roth N, Käsbohrer A, Mayrhofer S, Zitz U, Hofacre C, Domig KJ. The application of antibiotics in broiler production and the resulting antibiotic resistance in Escherichia coli: a global overview. Poult Sci. 2019;98(4):1791–804.
- Stanley D, Geier MS, Denman SE, Haring VR, Crowley TM, Hughes RJ, Moore RJ. Identification of chicken intestinal microbiota correlated with the efficiency of energy extraction from feed. Vet Microbiol. 2013;164(1–2):85–92.
- Kers JG, Velkers FC, Fischer EAJ, Hermes GDA, Stegeman JA, Smidt H. Host and environmental factors affecting the intestinal microbiota in chickens. Front Microbiol. 2018;9:235.
- Alagawany M, Abd El-Hack ME, Farag MR, Sachan S, Karthik K, Dhama K. The use of probiotics as eco-friendly alternatives for antibiotics in poultry nutrition. Environ Sci Pollut Res. 2018;25(11):10611–8.
- Jeni RE, Dittoe DK, Olson EG, Lourenco J, Corcionivoschi N, Ricke SC, Callaway TR. Probiotics and potential applications for alternative poultry production systems. Poult Sci. 2021;100(7):101156.
- 11. Yaqoob MU, Wang G, Wang M. An updated review on probiotics as an alternative of antibiotics in poultry - A review. Anim Biosci. 2022;35(8):1109–20.
- Gupta VK, Kim M, Bakshi U, Cunningham KY, Davis JM, Lazaridis KN, Nelson H, Chia N, Sung J. A predictive index for health status using species-level gut microbiome profiling. Nat Commun. 2020;11(1):4635.
- Ma C, Chen K, Wang Y, Cen C, Zhai Q, Zhang J. Establishing a novel colorectal cancer predictive model based on unique gut microbial single nucleotide variant markers. Gut Microbes. 2021;13(1):1–6.
- Ayoola MB, Pillai N, Nanduri B, Rothrock MJ, Ramkumar M. Preharvest Environmental and Management Drivers of Multidrug Resistance in Major Bacterial Zoonotic Pathogens in Pastured Poultry Flocks. Microorganisms. 2022;10(9):1703.
- Hwang D, Rothrock M, Pang H, Dev Kumar G, Mishra A. Farm management practices that affect the prevalence of Salmonella in pastured poultry farms. LWT- Food Science and Technology. 2020;127:109423.
- Saba T. Recent advancement in cancer detection using machine learning: systematic survey of decades, comparisons and challenges. J Infect Public Health. 2020;13(9):1274–89.
- Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak. 2019;19(1):1–6.
- Lian X, Yang S, Li H, Fu C, Zhang Z. Machine-learning-based predictor of human-Bacteria protein-protein interactions by incorporating comprehensive host-network Properties. J Proteome Res. 2019;18(5):2195–205.
- Ghyselinck J, Verstrepen L, Moens F, Van den Abbeele P, Said J, Smith B, Bjarnason I, Basit AW, Gaisford S. A 4-strain probiotic supplement influences gut microbiota composition and gut wall function in patients with ulcerative colitis. Int J Pharm. 2020;587:119648.
- Barlow GM, Yu A, Mathur R. Role of the gut microbiome in obesity and diabetes Mellitus. Nutr Clin Practice: Official Publication Am Soc Parenter Enter Nutr. 2015;30(6):787–97.

- 21. Rowin J, Xia Y, Jung B, Sun J. Gut inflammation and dysbiosis in human motor neuron disease. Physiological Rep 2017, 5(18).
- 22. Pasolli E, Truong DT, Malik F, Waldron L, Segata N. Machine learning Metaanalysis of large metagenomic datasets: Tools and Biological Insights. PLoS Comput Biol. 2016;12(7):e1004977.
- Lam KL, Cheng WY, Su Y, Li X, Wu X, Wong KH, Kwan HS, Cheung PC. Use of random forest analysis to quantify the importance of the structural characteristics of beta-glucans for prebiotic development. Food Hydrocoll. 2020;108:106001.
- Plaza-Diaz J, Ruiz-Ojeda FJ, Gil-Campos M, Gil A. Mechanisms of action of Probiotics. Adv Nutr. 2019;10(suppl1):49–566.
- Kulkarni RR, Gaghan C, Gorrell K, Sharif S, Taha-Abdelaziz K. Probiotics as Alternatives to Antibiotics for the Prevention and Control of necrotic enteritis in chickens. Pathogens (Basel Switzerland) 2022, 11(6).
- 26. Markowiak P, Śliżewska K. The role of probiotics, prebiotics and synbiotics in animal nutrition. Gut Pathogens. 2018;10:21.
- Arsène MMJ, Davares AKL, Andreevna SL, Vladimirovich EA, Carime BZ, Marouf R, Khelifi I. The use of probiotics in animal feeding for safe production and as potential alternatives to antibiotics. Veterinary World. 2021;14(2):319–28.
- Jha R, Das R, Oak S, Mishra P. Probiotics (Direct-Fed Microbials) in Poultry Nutrition and their Effects on nutrient utilization, growth and laying performance, and Gut Health: a systematic review. Animals: An open Access Journal from MDPI 2020, 10(10).
- 29. Reuben RC, Sarkar SL, Ibnat H, Roy PC, Jahid IK. Novel mono- and multi-strain probiotics supplementation modulates growth, intestinal microflora composition and haemato-biochemical parameters in broiler chickens. Veterinary Med Sci. 2022;8(2):668–80.
- Mani SR, Vijayan K, Jacob JP, Vijayakumar S, Kandhasamy S. Evaluation of probiotic properties of Lysinibacillus macroides under in vitro conditions and culture of Cyprinus carpio on growth parameters. Arch Microbiol. 2021;203(7):4705–14.
- Ocejo M, Oporto B, Hurtado A. 16S rRNA amplicon sequencing characterization of caecal microbiome composition of broilers and free-range slow-growing chickens throughout their productive lifespan. Sci Rep. 2019;9(1):2506.
- Rothrock MJ Jr, Locatelli A, Feye KM, Caudill AJ, Guard J, Hiett K, Ricke SC. A microbiomic analysis of a pasture-raised broiler flock elucidates Foodborne Pathogen Ecology along the Farm-To-Fork continuum. Front Vet Sci. 2019;6:260.
- van Gerwe T, Miflin JK, Templeton JM, Bouma A, Wagenaar JA, Jacobs-Reitsma WF, Stegeman A, Klinkenberg D. Quantifying transmission of Campylobacter jejuni in commercial broiler flocks. Appl Environ Microbiol. 2009;75(3):625–8.
- Sibanda N, McKenna A, Richmond A, Ricke SC, Callaway T, Stratakos AC, Gundogdu O, Corcionivoschi N. A review of the Effect of Management Practices on Campylobacter Prevalence in Poultry Farms. Front Microbiol. 2018;9:2002.
- Soro AB, Whyte P, Bolton DJ, Tiwari BK. Strategies and novel technologies to control Campylobacter in the poultry chain: a review. Compr Rev Food Sci Food Saf. 2020;19(4):1353–77.
- Śmiałek M, Kowalczyk J, Koncicki A. The Use of Probiotics in the reduction of Campylobacter spp. Prevalence in Poultry. Animals. 2021;11(5):1355.
- Balta I, Butucel E, Stef L, Pet I, Gradisteanu-Pircalabioru G, Chifiriuc C, Gundogdu O, McCleery D, Corcionivoschi N. Anti-Campylobacter Probiotics: latest mechanistic insights. Foodborne Pathog Dis. 2022;19(10):693–703.
- Asakura H, Nakayama T, Yamamoto S, Izawa K, Kawase J, Torii Y, Murakami S. Long-term grow-out affects Campylobacter jejuni colonization fitness in coincidence with altered microbiota and lipid composition in the Cecum of laying hens. Front Vet Sci. 2021;8:675570.
- Babacan O, Harris SA, Pinho RM, Hedges A, Jørgensen F, Corry JEL. Factors affecting the species of Campylobacter colonizing chickens reared for meat. J Appl Microbiol. 2020;129(4):1071–8.
- Śmiałek M, Kowalczyk J, Koncicki A. The Use of Probiotics in the reduction of Campylobacter spp. Prevalence in Poultry. Animals: An open Access Journal from MDPI 2021, 11(5).
- Rossi F, Amadoro C, Colavita G. Members of the Lactobacillus Genus Complex (LGC) as opportunistic pathogens: a review. Microorganisms 2019, 7(5).
- 42. Tarique M, Abdalla A, Masad R, Al-Sbiei A, Kizhakkayil J, Osaili T, Olaimat A, Liu S-Q, Fernandez-Cabezudo M, al-Ramadi B, et al. Potential probiotics and postbiotic characteristics including immunomodulatory effects of lactic acid bacteria isolated from traditional yogurt-like products. LWT. 2022;159:113207.
- Khiabani SA, Haghighat S, Khosroshahi HT, Asgharzadeh M, Kafil HS. Clostridium species diversity in gut microbiota of patients with renal failure. Microb Pathog. 2022;169:105667.

- 44. Oakley BB, Kogut MH. Spatial and temporal changes in the broiler chicken cecal and fecal Microbiomes and Correlations of Bacterial Taxa with Cytokine Gene expression. Front Vet Sci. 2016;3:11.
- Li CL, Wang J, Zhang HJ, Wu SG, Hui QR, Yang CB, Fang RJ, Qi GH. Intestinal morphologic and microbiota responses to dietary Bacillus spp. in a broiler chicken model. Front Physiol. 2018;9:1968.
- Maioli TU, Borras-Nogues E, Torres L, Barbosa SC, Martins VD, Langella P, Azevedo VA, Chatel JM. Possible benefits of Faecalibacterium prausnitzii for Obesity-Associated Gut Disorders. Front Pharmacol. 2021;12:740636.
- Ezeji JC, Sarikonda DK, Hopperton A, Erkkila HL, Cohen DE, Martinez SP, Cominelli F, Kuwahara T, Dichosa AEK, Good CE, et al. Parabacteroides distasonis: intriguing aerotolerant gut anaerobe with emerging antimicrobial resistance and pathogenic and probiotic roles in human health. Gut Microbes. 2021;13(1):1922241.
- Mukherjee A, Dutta D, Banerjee S, Ringø E, Breines EM, Hareide E, Chandra G, Ghosh K. Potential probiotics from indian major carp, Cirrhinus mrigala. Characterization, pathogen inhibitory activity, partial characterization of bacteriocin and production of exoenzymes. Res Vet Sci. 2016;108:76–84.
- Hinton A Jr, Hume ME, Deloach JR. Role of metabolic Intermediates in the inhibition of Salmonella typhimurium and Salmonella enteritidis by Veillonella. J Food Prot. 1993;56(11):932–7.
- Lourenco JM, Rothrock MJ Jr, Sanad YM, Callaway TR. The effects of feeding a soybean-based or a soy-free diet on the gut microbiome of pasture-raised chickens throughout their lifecycle. Front Sustainable Food Syst. 2019;3:36.
- Zhai Q, Wang G, Zhao J, Liu X, Tian F, Zhang H, Chen W. Protective effects of Lactobacillus plantarum CCFM8610 against acute cadmium toxicity in mice. Appl Environ Microbiol. 2013;79(5):1508–15.
- 52. Dashtbanei S, Keshtmand Z. A mixture of Multi-Strain Probiotics (Lactobacillus Rhamnosus, Lactobacillus Helveticus, and Lactobacillus Casei) had Anti-Inflammatory, anti-apoptotic, and Anti-Oxidative Effects in oxidative Injuries Induced by Cadmium in Small Intestine and Lung. Probiotics and antimicrobial proteins; 2022.
- N GA, EM MA, Shabanna S, Abd-Elrahman E. Protective efficacy of Streptococcus Thermophilus against Acute Cadmium Toxicity in mice. Iran J Pharm Research: IJPR. 2018;17(2):695–707.
- Liu Y, Li Y, Liu K, Shen J. Exposing to cadmium stress cause profound toxic effect on microbiota of the mice intestinal tract. PLoS ONE. 2014;9(2):e85323.
- Zhang F, Zheng W, Guo R, Yao W. Effect of dietary copper level on the gut microbiota and its correlation with serum inflammatory cytokines in Sprague-Dawley rats. J Microbiol. 2017;55(9):694–702.

- Coryell M, McAlpine M, Pinkham NV, McDermott TR, Walk ST. The gut microbiome is required for full protection against acute arsenic toxicity in mouse models. Nat Commun. 2018;9(1):5424.
- Wu WY, Chou PL, Yang JC, Chien CT. Silicon-containing water intake confers antioxidant effect, gastrointestinal protection, and gut microbiota modulation in the rodents. PLoS ONE. 2021;16(3):e0248508.
- 58. Popoff MY, Le Minor L. Antigenic formulas of the Salmonella serovars. 1997.
- Stern NJ, Wojton B, Kwiatek K. A Differential-selective medium and dry icegenerated atmosphere for recovery of Campylobacter jejuni. J Food Prot. 1992;55(7):514–7.
- Rothrock MJ Jr, Hiett KL, Gamble J, Caudill AC, Cicconi-Hogan KM, Caporaso JG. A hybrid DNA extraction method for the qualitative and quantitative assessment of bacterial communities from poultry production samples. *Journal of visualized experiments: JoVE* 2014(94).
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the national academy of sciences* 2011, 108(supplement_1):4516–4522.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI. QIIME allows analysis of highthroughput community sequencing data. Nat Methods. 2010;7(5):335–6.
- 63. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460–1.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72(7):5069–72.
- 65. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2010;26(2):266–7.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.