Assembly of hundreds of novel bacterial genomes from the chicken caecum

Authors: Laura Glendinning[1], Robert D. Stewart[1], Mark J. Pallen[2, 3, 4], Kellie A. Watson[1], Mick Watson[1]

[1]Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, Midlothian, United Kingdom

[2]Microbes in the Food Chain, Quadram Institute Bioscience, Norwich, UK

[3]School of Biological Sciences, University of East Anglia, Norwich, Norfolk, UK

[4]School of Veterinary Medicine, University of Surrey, Guildford, Surrey, UK

Abstract:

Chickens are a highly important source of protein for a large proportion of the human population. The cecal microbiota plays a crucial role in chicken nutrition through the production of short chain fatty acids, nitrogen recycling and amino acid production. In this study we sequenced DNA from cecal contents samples taken from 24 chickens belonging to either a fast or slower growing breed consuming either a vegetable-only diet or a diet containing fish meal. We utilised 1.6T of Illumina data to construct 469 draft metagenome-assembled bacterial genomes, including 460 novel strains, 283 novel species and 42 novel genera. We also compared the abundance of these genomes, and the carbohydrate active enzymes they produce, between our chicken groups and demonstrate that there are both breed- and diet- specific microbiomes, as well as an overlapping core microbiome. This data will form the basis for future studies examining the composition and function of the chicken cecal microbiota.

Background:

There are an estimated 23 billion live chickens on the planet at any one time (1), out-numbering humans by over 3:1. As most of these are reared for food, the actual number of chickens produced per year is even higher, at almost 65 billion, leading some to speculate that the accumulation of chicken bones in the fossil record will be used by future archaeologists as a unique marker for the Anthropocene (2).

Since the 1960s, worldwide chicken meat production has increased by over ten times (3). Global meat production is predicted to be 16% higher in 2025 vs. 2015, with most of this increase originating from poultry meat production (4). Part of the popularity of chicken meat is that, due to intensive selection, chickens have been developed which are highly productive in terms of their growth rate with efficient feed conversion ratios (the rate at which chickens convert feed into muscle), decreasing from 3.0 in the 1960s to 1.7 in 2005 (5), making them a cheap source of protein in comparison to other livestock. Chickens also produce less

greenhouse gasses per kg of meat than pigs, cattle and sheep (6). The potential to manipulate the microbiota in chickens to gain further increases in productivity is of great commercial and scientific interest, leading to the use of probiotics across the poultry industry (7).

As well as playing an important role in pathogen protection (8) and immune system development (9), the microbiota of the chicken also plays a crucial nutritional role. The largest concentration of microbial cells in the chicken gastrointestinal tract can be found in the ceca and thereby the majority of chicken microbiota studies focus primarily on these microbial communities. Members of the cecal microbiota are able to produce short chain fatty acids (SCFAs) such as acetate, butyrate, lactate and propionate from carbohydrate sources which have passed through the small intestine; these SCFAs are then able to be absorbed by the bird and used as an energy source (10). Members of the chicken cecal microbiota have also been implicated in the recycling of nitrogen by the degradation of nitrogenous compounds (11) and the synthesis of amino acids (12). One study demonstrated that 21% of the variation in chicken abdominal fat mass could be attributed to the cecal microbiota composition, when controlling for host genetic effects (13). Differences have also been observed between birds with high and low feed efficiency (14, 15). However, despite extensive research over many decades, the quantitative importance of the ceca in chicken nutrition remains unclear (16), and relatively few microbes commensal in the chicken gut have been sequenced and deposited in public repositories.

The emergence of cheaper DNA sequencing technologies (17, 18) has led to an explosion in studies which have sought to characterise the chicken gastrointestinal microbiota, particularly using 16S rRNA gene based methods. Whilst valuable, marker-gene studies do not enable an in-depth functional and genomic characterisation of the microbiome. Some microbes from the chicken ceca have been successfully cultured and sequenced, including 133 gut anaerobes with a wide range of metabolic potentials (19); however it is highly unlikely that these microbes represent the entire diversity of the chicken cecal microbiota, due to the difficulty in culturing many anaerobic gut microorganisms. One method which avoids this issue of culturability is the construction of metagenome assembled genomes (MAGs). Due to improvements in computational power and sequencing technologies, and the development of new computational approaches (20, 21), it is now possible to accurately bin short-read metagenomic data into high-quality genomes. Using this technique thousands of MAGs have been generated from various environments, including humans (22, 23), the rumen (24, 25), pig faeces (26), marine surface waters (27, 28), an underground aquifer system (29) and other public datasets (30).

In this study we sought to use metagenomic sequencing, assembly and binning to investigate the chicken cecal microbiota. In order to maximise diversity, we chose two commercial bird genotypes with different growth phenotypes, fed two different diets. This also allows us to look at the effects of breed and diet on strain level microbial abundance. The lines chosen for the study are Ross 308, a fast growing broiler breed, and the Ranger Classic, a slower growing broiler aimed at free-range, organic farms. All birds were fed either

a vegetable-only diet or a diet based on fish meal as the protein source. The inclusion of fish meal in chicken diets has previously been linked to changes in the cecal microbiota and is correlated with an increased risk of necrotic enteritis (31, 32). We assemble 460 novel microbial strains, predicted to represent 283 novel microbial species and 42 novel microbial genera from the chicken microbiome, and go on to demonstrate both a breed- and diet- specific microbiota. Whilst we show that large numbers of strains are shared between birds, it is their relative abundance that largely drives breed and diet effects. This is the first large-scale binning of the chicken cecal microbiota, and we believe these data will form the basis for future studies of the structure and function of the chicken gut microbiome.

Methods:

Ethical statement

Animals were housed in premises licensed under a UK Home Office Establishment License within the terms of the UK Home Office Animals (Scientific Procedures) Act 1986. Housing and husbandry complied with the Code of Practice for Housing and Care of Animals Bred, Supplied or Used for Scientific Purposes and were overseen by the Roslin Institute Animal Welfare and Ethical Review Board. Animals were culled by schedule one methods authorized by the Animals (Scientific Procedures) Act 1986.

Study design

Ross 308 (Aviagen, UK) (n=12) and Ranger Classic (Aviagen, UK) (n=12) chickens were hatched and housed at the National Avian Research Facility in Edinburgh (UK). Birds were fed either a vegetable only diet or a diet supplemented with fish meal (**Table 1, Supplementary table 1**) (Diet formulation: **Supplementary tables 2 and 3,** nutritional info: **Supplementary table 4**). Birds received Mareks-Rispins vaccinations (Merial, France) at 1-2 days of age and were housed by group in separate floor pens (within the same room) with wood shaving bedding, and receiving food and water ad libitum. Stocking densities were based on UK Home Office Animals (Scientific Procedures) Act 1986, resulting in a floor area per bird of 0.133 m^2 at 5 weeks of age. Birds were euthanized by cervical dislocation at 5 weeks of age and cecal content samples were collected. Contents from both ceca were pooled to make one sample per bird. Samples were stored at 4°C for a maximum of 24 hours until DNA extraction, except for those from DNA extraction batch 2 which were frozen at -20°C for 9 days prior to DNA extraction (**Supplementary table 5**). DNA extraction was performed as described previously using the DNeasy PowerLyzer PowerSoil Kit (Qiagen, UK) (33). Shotgun sequencing was performed on a NovaSeq (Illumina) producing 150bp paired-end reads.

98 **Table 1: Chicken details**

Line	Diet	N	Mean body weight (kg)
Ranger Classic	Fish meal diet	6 (3 male, 3 female)	Female: 1.89 Male: 2.13
	Vegetable diet	6 (3 male, 3 female)	Female: 1.57 Male: 1.94
Ross 308	Fish meal diet	6 (3 male, 3 female)	Female: 2.33 Male: 2.70
	Vegetable diet	6 (4 male, 2 female)	Female: 2.23 Male: 2.89

99

100 **Bioinformatics**

101 Assembly and binning was carried out as previously described (24, 25). Illumina adaptors were removed using
102 trimmomatic (34). Single sample assemblies were performed using IDBA-UD (35) with the options --
103 num_threads 16 --pre_correction --min_contig 300. BWA MEM (36) was used to separately map reads from
104 every sample back to every assembly. SAMtools (37) was used to create BAM files and the command
105 jgi_summarize_bam_contig_depths was run on all BAM files for each assembly to calculate coverage. A
106 coassembly was also carried out on all 24 samples using MEGAHIT (options: --continue --kmin-1pass -m
107 100e+10 --k-list 27,37,47,57,67,77,87 --min-contig-len 1000 -t 16) (38). Contigs were filtered to a minimum
108 length of 2kb, then indexed and mapped as for single assemblies.

109 METABAT2 (21) was used on both single-sample assemblies and co-assemblies to carry out metagenomic
110 binning, taking into account coverage values and with the options --minContigLength 2000, --
111 minContigDepth 2. All bins were dereplicated using dRep (39) with the options dereplicate_wf -p 16 -comp
112 80 -con 10 -str 100 -strW. Bins were dereplicated at 99% average nucleotide identity (ANI), resulting in each
113 MAG being taxonomically equivalent to a microbial strain. Bins were also dereplicated at 95% ANI to calculate
114 the number of species represented within our MAGs. CompareM was used to calculate average amino acid
115 identity (AAI) (40).

116 The completeness and contamination of all bins was assessed using CheckM (41) with the options lineage_wf,
117 -t 16, -x fa and filtering for completeness ≥80% and contamination ≤10%. GTDB-Tk (42) was used to assign
118 taxonomy to MAGs, except for CMAG_333 which upon visual inspection of taxonomic trees was identified
119 more accurately as *Clostridia*. For submission of our MAGs to NCBI, MAGs were named based on the following
120 rule: if the lowest taxonomy assigned by GTDB-Tk did not correlate with an NCBI classification at the correct
121 taxonomic level then MAGs were named after the lowest taxonomic level at which NCBI and GTDB-Tk

122 matched. Comparative genomics between the MAGS and public datasets was carried out using MAGpy (43).
123 The taxonomic tree produced by MAGpy was re-rooted manually using Figtree (44) at the branch between
124 Firmicutes and the other bacterial phyla, and subsequently visualised using Graphlan (45). The novelty of
125 genomes in comparison to those present in public databases was also determined. Genomes were defined
126 as novel strains if the ANI output by GTDB-Tk was <99%. Genomes were determined as novel species if the
127 ANI output by GTDB-Tk was <95% or if an ANI was not output by GTDB-Tk then the average protein similarity
128 output by MAGpy was <95%. Genera were defined as novel if all MAGs which clustered at 60% AAI (46) were
129 not assigned a genus by GTDB-Tk. Proposed names for new genera and species belonging to these genera
130 were formulated based on the International Code of Nomenclature of Prokaryotes (47).

131 Carbohydrate active enzymes (CAZymes) were identified by comparing MAG proteins to the CAZy database
132 (48) using dbcan2 (version 7, 24th August 2018). The abundance of CAZyme groups was then calculated as
133 the sum of reads mapping to MAG proteins within each group after using DIAMOND (49) to align reads to
134 the MAG proteins.

135

136 **Statistics and graphs**

137 Univariate general linear models (GLMs) were performed in SPSS Statistics 21 (IBM) with gender, line and
138 diet as fixed factors. All other statistical analyses were carried out in R (50) (version 3.5.1.). NMDS graphs
139 were constructed using the Vegan package (51) and ggplot2 (52), using the Bray–Curtis dissimilarity. Boxplots
140 were constructed using the ggplot2 package. UpSet graphs were constructed using the UpSetR package (53).
141 Correlation coefficients, using R's hclust function, were used to cluster samples and MAGs within heatmaps.
142 PERMANOVA analyses were performed using the Adonis function from the Vegan package. The package
143 DESeq2 (54) was used to calculate differences in abundance for individual MAGs, taxonomies and CAZymes.
144 For MAGs, subsampling to the lowest sample coverage was performed prior to analysis by PERMANOVA and
145 NMDS and before calculating the 1X and 10X coverage of MAGs in samples.

146

147 **Data availability**

148 Paired-read fastq files have been submitted to the European Nucleotide Archive under project PRJEB33338.
149 MAG fasta files have been submitted to Edinburgh DataShare (https://doi.org/10.7488/ds/2584).

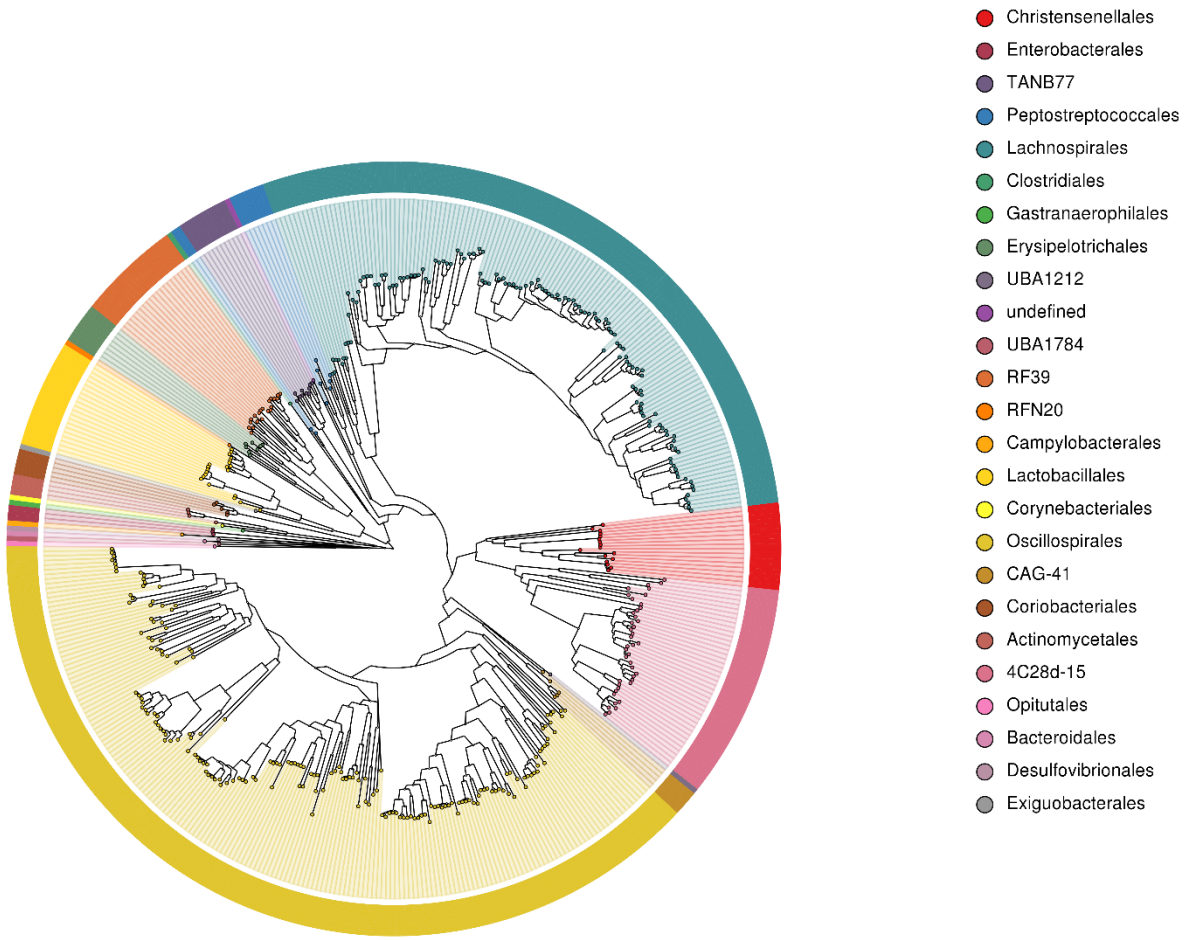
Results

Assembly of 469 draft microbial genomes from chicken ceca

We produced 1.6T of Illumina data from 24 chicken samples, carried out a metagenomic assembly of single samples and also a co-assembly of all samples. 4524 metagenomic bins were created from the single-sample binning and 576 more were created from co-assembly binning. We were left with a total of 469 dereplicated genomes (99% ANI) with estimated completeness of ≥80% and estimated contamination ≤10% (**Supplementary figure 1**), 377 of which originated from the single-sample binning and 92 from the co-assembly. Of these, 349 had completeness >90% and contamination <5% (high-quality draft genomes as defined by Bower *et al.* (55)), 210 were >95% complete with <5% contamination and 47 MAGs were >97% complete with 0% contamination. The distribution of these MAGs (based on coverage) between the 24 samples can be found in **Dataset 1**. After dereplication to 95% ANI, 335 MAGs remained, representing putative novel species identified in our samples. Our dataset therefore contains 469 microbial strains from 335 species. 283 of these species and 460 of these strains were novel when compared to public databases **(Dataset 2)**.

Dataset 2 contains the NCBI taxonomic assignment for each MAG along with the assembly characteristics and GTDB-Tk taxonomic assignments. **Dataset 3** contains comparative genomics information produced by MAGpy. **Figure 1** shows a phylogenetic tree of the MAGs. This was used to manually correct any errors in taxonomic identification. The tree is dominated by *Oscillospirales* (179) and *Lachnospirales* (134), with smaller numbers of genomes representing other bacterial orders, including *4C28d-15* (42), *Lactobacillales* (21), *RF39* (20), *Christensenellales* (17), *TANB77* (10), *Peptostreptococcales* (9), *Erysipelotrichales* (8), etc. 97 MAGS were identified to species, 246 identified to genus, 115 identified to family, 10 identified to order and 1 identified to class. No MAGs were identified as Archaea.

Figure 1:



Phylogenetic tree of the 469 draft microbial genomes from the chicken ceca, labelled by taxonomic order, as defined by GTDB-Tk. Draft genomes labelled as "undefined" were only able to be assigned taxonomy at a higher level than order.

Of the MAGs that show greater than 95% ANI with an existing sequenced genome, several of these genomes have previously been identified in chickens. Our MAGs include 6 novel strains of *Anaeromassilibacillus sp. An250* (19), a novel strain of *Anaerotignum lactatifermentans* (56), a novel strain of *Blautia sp. An81* (19), 3 novel strains of *Drancourtella sp. An57* (19), a novel strain of *Enterococcus cecorum* (57), 2 novel strains of *Escherichia coli* (14, 58, 59), 3 novel strains of *Eubacteriaceae bacterium CHKCI004* (60), a novel strain of *Eubacterium sp. An11* (19), two novel strains of *Faecalibacterium* spp. (19, 30), 7 novel strains of *Flavonifactor* spp. (19), 3 novel strains of *Gordonibacter* spp. (19), 1 novel strain of *Helicobacter pullorum* (61), 15 novel strains of *Lachnoclostridium* spp. (19), 6 novel strains of *Lachnospiraceae bacterium UBA1818* (30), 2 novel strains of *Massiliomicrobiota sp. An134* (19) and 5 novel strains of *Pseudoflavonifractor sp. An184* (19).

We also identified several Lactobacillli which have previously been isolated from the chicken gastrointestinal tract and have been suggested as potential probiotics in chickens, including 5 novel strains of *Lactobacillus*

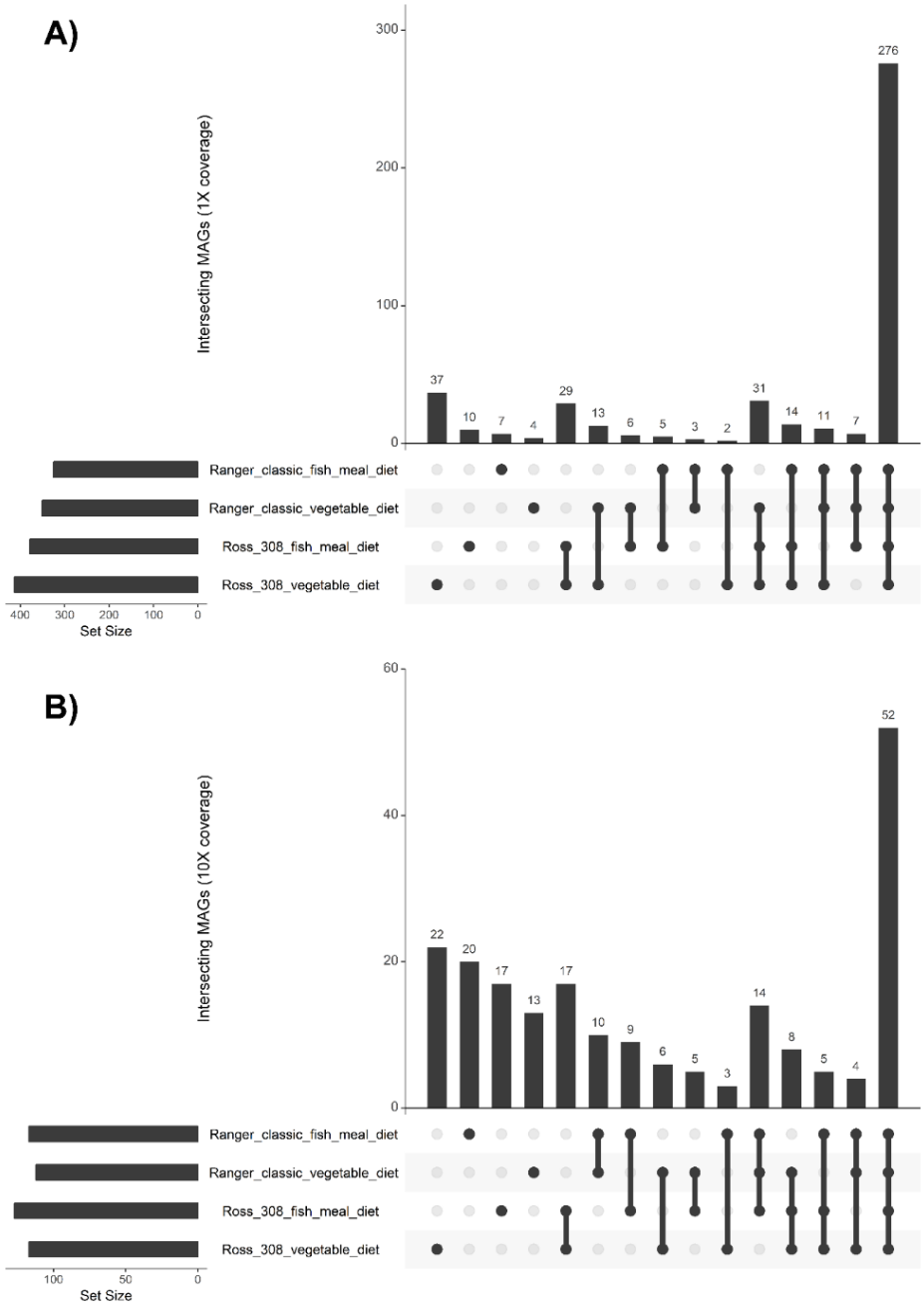
crispatus (62-64), 2 novel strains of *Lactobacillus gallinarum* (65), a novel strain of *Lactobacillus johnsonii* (66, 67), a novel strain of *Lactobacillus oris* (68), a novel strain of *Lactobacillus reuteri* (59, 62, 69) and a novel strain of *Lactobacillus salivarius* (59, 67, 70).

Our MAGs represent several putative novel species from 7 taxonomic classes: including 25 species of *Bacilli*, 252 species of *Clostridia*, 2 species of *Coriobacteriia*, 1 species of *Desulfovibrionia*, 1 species of *Lentisphaeria*, 1 species of *Vampirovibrionia* and 1 species of *Verrucomicrobiae*. These include 5 novel species of *Lactobacillus*. Our MAGs also contain 42 putative novel genera which contain 69 of our MAGs. We defined a genus as novel if all MAGs which clustered at 60% AAI were not assigned a genus by GTDB-Tk (**Dataset 4**). This is a conservative method of defining genera, as genera within different taxonomies may cluster at higher AAIs (46, 71, 72). 40 of these novel genera belong to the class *Clostridia*, with over half belonging to the order *Oscillospirales*. One of the remaining novel genera contains one MAG which belongs to the *Bacilli* class (order *Exiguobacterales*) while the remaining genus belongs to the *Cyanobacteria*, within the order *Gastranaerophilales*. GTDB-Tk was unable to assign taxonomy to either of these genera at lower than order level, indicating that they may belong to novel bacterial families. It should also be noted that several genus-level MAG clusters do not contain any MAGs which were assigned a valid NCBI genus label but instead only received names defined by GTDB-Tk. For example, Group 16 (**Dataset 4**) is entirely constituted by MAGs of the genus UBA7102.

Presence of a core chicken cecal microbiota

125 MAGs were found to be present in at least 1X coverage in all samples and 4 of these MAGs were found to be ≥X10 in all samples: *Alistipes sp. CHKCI003* CMAG_6, uncultured *Bifidobacterium* sp. CMAG_55, uncultured *Bifidobacterium* sp. CMAG_59 and *Firmicutes bacterium CAG_94* CMAG_438. Only one MAG was found to be uniquely present in only one sample at ≥1X coverage: uncultured *Clostridia* sp. CMAG_391 in Chicken 16 (Ross 308: Vegetable diet). The distribution of MAGs between groups can be seen in **Figure 2.** 276 MAGs were on average present at at least 1X coverage in all groups and could therefore be described as a core microbiota shared amongst the chickens in our study.

Figure 2:



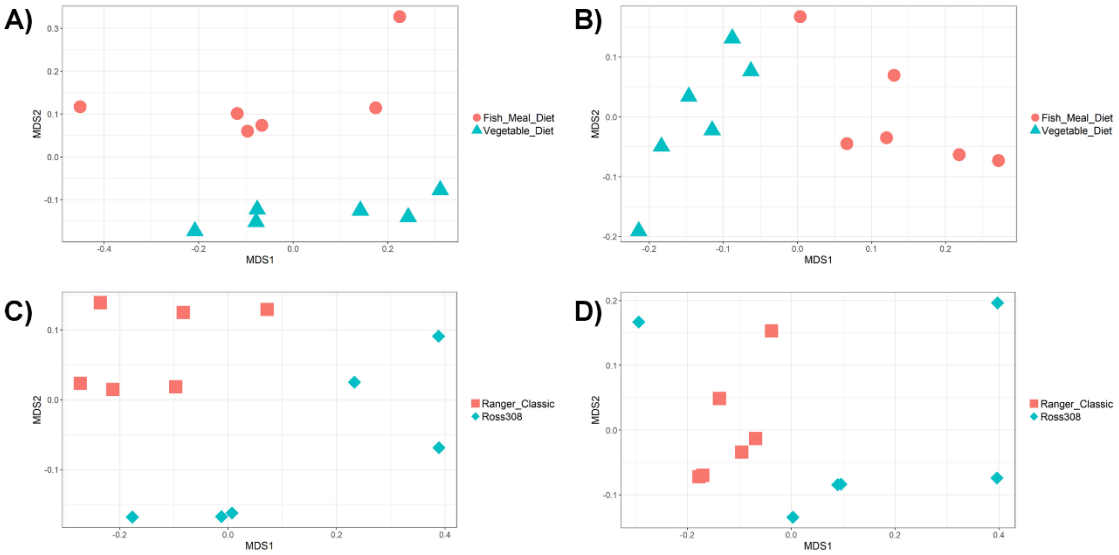
UpSet graphs showing the number of shared MAGs at A) average 1X coverage and B) average 10X coverage in the four chicken groups

Differences in cecal MAGs based on chicken line and diet.

When comparing samples based on the coverage of MAGs, significant clustering of samples by group can be observed when comparing all groups (PERMANOVA: $P < 0.001$); between chicken lines (All samples: PERMANOVA: $P < 0.001$; Within vegetable diet: PERMANOVA: $P = 0.015$, Within fish meal diet PERMANOVA: $P = 0.0082$)(**Figure 3**) and between diets (All samples: PERMANOVA: $P = 0.008$; Within Ross 308 line:

PERMANOVA: P = 0.018; Within Ranger Classic line: PERMANOVA: P = 0.0043) (**Figure 3**). A significant interaction was also observed between line and diet (Line*Diet PERMANOVA: P = 0.038). Gender and DNA extraction batch were not found to have significantly affected the abundance of MAGs (PERMANOVA: P>0.05).

Figure 3:



NMDS of chicken cecal samples clustered by proportion of MAGs (Bray-Curtis dissimilarity). A) Ross 308 birds clustered by diet (PERMANOVA: P = 0.018) B) Ranger Classic birds clustered by diet (PERMANOVA: P = 0.0043) C) Birds on a vegetable diet clustered by line (PERMANOVA: P = 0.015) D) Birds on a fish meal diet clustered by line (PERMANOVA: P = 0.0082).

MAGs which were significantly more abundant by coverage between groups were identified by DESeq2 (**Figure 4**); a full list of these MAGs can be found in **Dataset 5**. In Ross 308 birds, 43 MAGs were found to be differentially abundant between the two diets, while in Ranger Classic birds 45 MAGs were found to be differentially abundant. Several MAGs were found to be differentially abundant between the two lines when birds were consuming a vegetable diet (61 MAGs) or a fish meal diet (69 MAGs). 98 MAGs were found to be differentially abundant between lines when controlling for diet and 64 MAGs were found to be differentially abundant between diets when controlling for line.

246 **Figure 4:**

247

248 **Heatmap showing the proportional coverage of MAGs which were significantly differently abundant**
249 **between groups (Deseq2, P ≤ 0.05). Euclidean clustering was used to cluster MAGs and samples.**

250

No MAGs were found to be significantly more abundant in both Ross 308 and Ranger Classic birds fed a fish meal diet, whilst four MAGs were found to be significantly more abundant in both Ross 308 and Ranger Classic birds fed a solely vegetable diet: uncultured *Lachnospiraceae* sp. CMAG_102, *Lachnoclostridium sp. An76* CMAG_121, *Faecalibacterium sp. An121* CMAG_31 and uncultured *Clostridia* sp. CMAG_357.

Eight MAGs were found to be significantly more abundant in Ross 308 chickens on both diets: uncultured *Pseudoflavonifractor* sp. CMAG_226, uncultured *Oscillospiraceae* sp. CMAG_257, uncultured *Clostridia* sp. CMAG_273 and uncultured *Clostridia* sp. CMAG_331, *Clostridia* sp. CMAG_194, *Lactobacillus gallinarum* CMAG_28, uncultured *Faecalibacterium* sp. CMAG_33 and *Lactobacillus gallinarum* CMAG_35. In contrast, only one MAG was found to be consistently more abundant in Ranger Classic birds on both diets (uncultured *Lachnospiraceae* sp. CMAG_229).

Lactobacilli are of particular interest to probiotic manufacturers. We found that both MAGs identified as *L.gallinarum* were more abundant in Ross 308 birds when controlling for diet, and four of the five MAGs identified as *L.crispatus* were more abundant in birds fed a diet with fish meal when controlling for chicken line.

One notable observation is the high amount of *Helicobacter pullorum* observed in the Ross 308: Vegetable diet group. While *H. pullorum* is often thought of as a pathogen, it has previously been isolated from the ceca of asymptomatic chickens (61) and carriage of *Helicobacter* by chickens is common in commercial flocks (73-75).

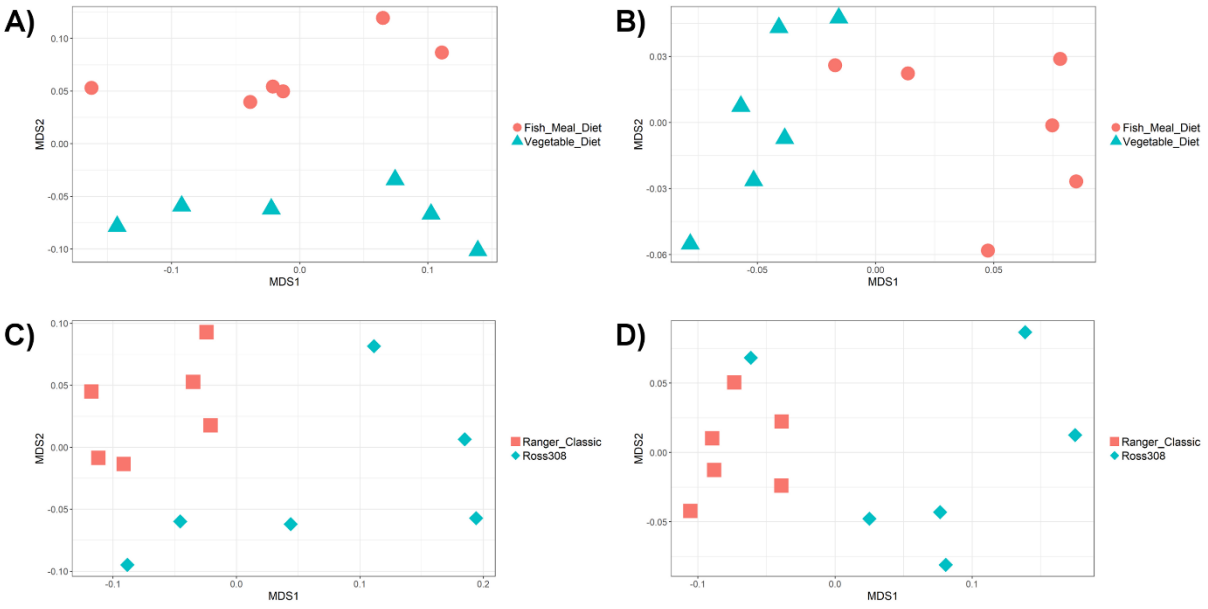
Differences in CAZymes between lines and diets

Carbohydrate-active enzymes (CAZymes) are enzymes involved in the metabolism, synthesis and binding of carbohydrates. They are grouped by the CAZy database (48) into the following major groups: the auxiliary activities (AAs) class, carbohydrate-binding modules (CBMs), carbohydrate esterases (CEs), glycoside hydrolases (GHs), glycosyltransferases (GTs) and polysaccharide lyases (PLs). As their names suggest, CEs are responsible for the hydrolysis of carbohydrate esters while CBMs are responsible for binding carbohydrates. GHs and PLs are both responsible for cleaving glycosidic bonds, hydrolytically or non-hydrolytically respectively, while GTs are able to catalyse the formation of glycosidic bonds. The AA class are not themselves CAZymes but instead act in conjunction with them as redox enzymes. We compared the predicted proteins from our MAGs with the CAZy database using dbcan with the cut-offs E-value < 1e-18 and coverage > 0.35.

When clustering groups by the abundance of MAG derived CAZymes, all groups separate visually (**Figure 5**) but only the following differences were significant: Ross 308 birds were shown to cluster significantly by diet (PERMANOVA, P=0.021), and birds receiving a fish meal diet clustered significantly by line (PERMANOVA, P=0.0065). A significant interaction was observed between line and diet (Line*Diet PERMANOVA: P = 0.0051).

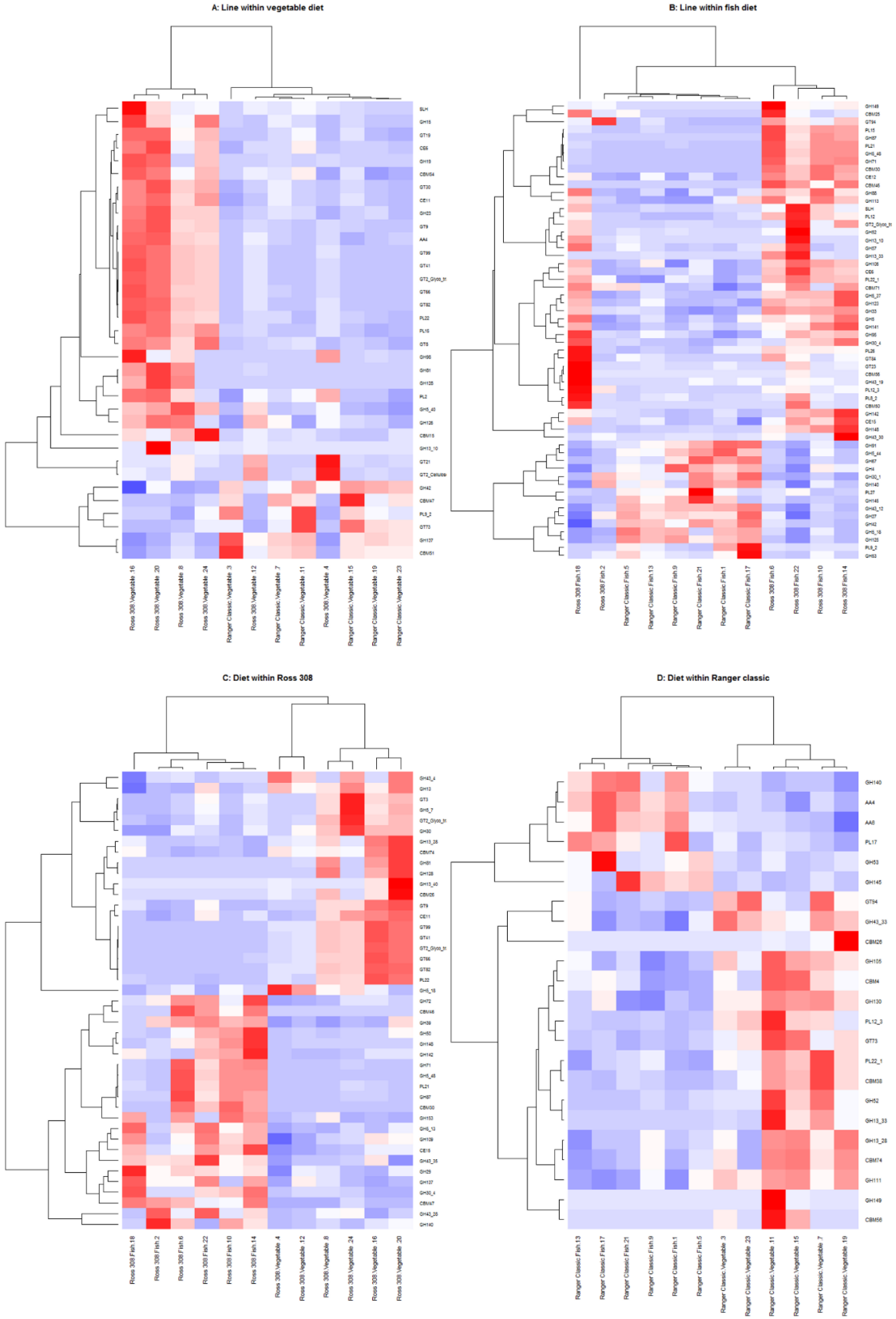
Using DESeq2 we also found that the abundances of specific CAZymes differed between groups (**Figure 6**), full lists of which can be found in **Dataset 6**. We found several starch degrading enzymes to be differentially abundant between lines when controlling for diet, including GH13 subfamily 10, GH15, GH57, GH4 and GH31, and between diets when controlling for line, including GH13, GH13 subfamily 28 and GH13 subfamily 33. We also found that several CAZymes involved in metabolising cellulose and hemi-cellulose were differentially abundant between lines when controlling for diet, including GH5 (subfamilies 19, 37, 48, 44, 18), CE6, GH43 (subfamilies 30, 19, 29, 12), GH115, CE2 and GH67, and between diets when controlling for line, including GH5 (subfamilies 7 and 48) and GH43 (subfamilies 33, 4 and 35). Gender and DNA extraction batch were not found to have significantly affected the abundance of CAZymes (PERMANOVA: P>0.05).

Figure 5:



NMDS of chicken cecal samples clustered by abundance of MAG CAZymes (Bray-Curtis dissimilarity). A) Ross 308 birds clustered significantly by diet (PERMANOVA: P = 0.021) B) Ranger Classic birds did not cluster significantly by diet (PERMANOVA: P = 0.095) C) Birds on a vegetable diet did not cluster significantly by line (PERMANOVA: P = 0.061) D) Birds on a fish meal diet clustered significantly by line (PERMANOVA: P = 0.0065).

Figure 6:



Heatmap showing the proportional coverage of MAGs which were significantly differently abundant between groups (Deseq2, P ≤ 0.05). Euclidean clustering was used to cluster MAGs and samples.

Line and gender impact the weight of the chicken

As we did not monitor individual feed intake, we cannot comment on the feed-conversion ratio of these birds; however, when housed and fed as a group, there are clear statistical differences between the birds in terms of weight (**Supplementary figure 2**). Univariate GLMs with fixed factors of gender, line and diet were performed, with bird weight as the dependent variable. Both gender ($P<0.001$) and line ($P<0.001$) were found to significantly impact weight, as expected. Diet was not found to significantly affect bird weight overall ($P=0.220$). We did observe a significant increase in bird weight in Ranger Classic birds ($P=0.007$), of both genders, fed a fish meal diet; which was not observed in the Ross 308 birds ($P=0.778$).

Discussion:

The function of the cecal microbiota in commercial chicken flocks remains somewhat unclear (16). However, several studies have implicated it as playing a role in nutrition, including in the production of SCFAs and the recycling of nitrogen from uric acid (10, 11, 76). Differences have also been observed in the cecal microbiota communities between chickens with differing feed efficiencies and abdominal fat accumulation (13-15). It may therefore be possible to increase chicken productivity by the manipulation of the chicken cecal microbiota. However, before this is possible we need to develop a good understanding of the types of bacteria present in the chicken and their nutritional function.

In this study we constructed 469 metagenome assembled genomes from chicken cecal contents, greatly expanding upon previous chicken cecal MAGs (77). 349 of our MAGs had completeness >90% and contamination <5% and can therefore be classed as high-quality draft genomes as defined by Bower *et al.* (55). Our MAGs include 460 novel strains and 283 novel species, including 5 novel *Lactobacillus* species. 97 MAGs were able to be identified to species level by GTDB-Tk and a further 246 could be identified to genus. We also identified 42 novel bacterial genera, 40 of which belonged to the class *Clostridia*. The remaining two genera belonged to the *Bacilli* class and the *Gastranaerophilales* order of *Cyanobacteria*, and may also belong to novel taxonomic families.

The majority of our MAGs belonged to the orders *Oscillospirales* and *Lachnospirales*, members of the *Clostridia* class. The high abundance of *Clostridia* observed during our study correlates with several previous studies examining the chicken cecal microbiota (19, 78-83). This is likely the product of chicks being raised in an environment where they are not exposed to a maternal microbiota as feral hens and chicks exposed to an adult hen have microbiotas which are far less dominated by *Firmicutes* and contain higher abundances of *Bacteroidetes* (84, 85).

Within our dataset we found 276 microbes which were on average present at a minimum 1X coverage in all four of our groups, potentially indicating a core chicken microbiota. However caution must be taken as all of our chickens were raised in the same facility and samples were all taken at the same time-point, which will have limited the variability in microbes present. Chicken microbiota can vary across flocks (86), at different times in the bird's life (87) and between free-range and intensively-reared chickens (88). To provide a truly representative dataset of chicken microbial genomes it would be necessary to sequence cecal samples from birds from multiple lines and raised under a variety of conditions. However, we do think it is likely that there is a core chicken cecal microbiota which is shared across sites and is irrespective of management conditions having assembled bacterial species that have been found in chickens in previous studies (14, 19, 30, 60). We also identified several novel *Lactobacillus* strains which have previously been posited as potential chicken probiotics, including *L.crispatus* (62-64), *L.gallinarum* (65), *L.johnsonii* (66, 67), *L.oris* (68), *L.reuteri* (59, 62, 69) and *L.salivarius* (59, 67, 70).

When analysing the abundance of MAGs between birds from different lines, consuming either a vegetable diet or a diet containing fish meal, we found significant differences in the microbial communities based on both line and diet. This agrees with previous studies where significant differences have been described in the intestinal microbiota of chickens from different lines, including those from faster and slower growing lines (89-91). Differences have also previously been observed in the microbiota when feeding chickens a diet supplemented with fish meal (31, 32). This correlates with differences observed in the weights of birds fed the fish meal diet. Ranger Classic birds fed a fish meal diet weighed significantly more than those fed a vegetable-only diet, whereas those was no significant difference between the weight of the Ross 308 birds fed on these two diets.

Examining those bacteria which were consistently significantly increased in a specific line regardless of diet or a specific diet regardless of line, the majority of these bacteria are novel species, therefore it is difficult to hypothesise why they are more abundant in particular bird lines or when birds are fed certain diets. Of those species that had previously been identified, the two *L.galinarum* strains were both consistently found to be more abundant in Ross 308 birds, while *Lachnoclostridium sp. An76* CMAG_121 and *Faecalibacterium sp. An121* CMAG_31 were found to be more abundant in birds on the vegetable diet. *L.gallinarum*, is a homofermentative and thermotolerant (65, 92) species which has previously been suggested as a potential chicken probiotic (63, 93, 94), while *Lachnoclostridium sp. An76* and *Faecalibacterium sp. An121* (19) have only very recently been discovered and are therefore not well characterised.

We are unsure why *H.pullorum* was observed in such high levels in the Ross 308: Vegetable diet group. We are unable to rule out contamination from the environment as our groups were housed in separate pens within the same room. We did not observe any negative health effects in this group, and the bacterium is very common in some flocks (61, 73-75, 95).

We wondered whether the differences in microbiota we observed between groups were associated with changes in the metabolic potential of the cecal microbial communities. Microbes isolated from the chicken ceca have previously been shown to have highly variable metabolic pathways (96, 97). We found that the abundances of certain MAG derived CAZymes involved in starch and cellulose degradation were significantly differently abundant between lines and diets. These molecules are highly abundant in the predominantly grain based diets fed to chicken. However, energy from starches and celluloses are not available to the chicken host unless these are first degraded into smaller carbohydrates by the gut microbiota, therefore differences between the ability of the cecal microbiota to degrade these molecules may lead to greater efficiency of energy extraction from feed (81).

It is also interesting to note that when analysing the abundance of MAG derived CAZymes in the chicken ceca, we only observed significantly separate clustering of birds by diet in the Ross 308 birds and by line in animals that were consuming the fish meal diet. This indicates that the differences in MAG abundances for these groups resulted in significantly different pools of metabolic genes. However, significant differences in MAG abundances were also observed for Ranger Classics on the two diets and for chickens of different lines consuming the vegetable diet, but this did not result in a significant difference in the total abundance of the CAZyme. This finding serves to highlight that changes in microbiota community composition do not necessarily lead to significant changes in the total metabolic potential of that community, although it is possible more significant differences would be observed with a larger sample size. It is worth noting that while our Ross 308 vegetable diet group contained 4 males and 2 females and the other groups contained 3 males and 3 females, gender was found to have no impact on the abundance of CAZymes or MAGs and this therefore should not have impacted our results.

One outlier was observed in our data: Chicken 2 appeared to cluster separately by the abundance of its MAGs in comparison to other Ross 308 birds consuming a fish meal diet, supporting the idea that while diet and line are associated with differences in the microbiota, variation will still exist between birds of the same line consuming similar diets. It should also be noted that the individual feed intake of each bird was not measured, meaning that some birds may have consumed different quantities of food, which could lead to variation in their microbiota compositions.

In conclusion, through the construction of metagenome assembled genomes we have greatly increased the quantity of chicken derived microbial genomes present in public databases and our data can be used as a reference dataset in future metagenomic studies. While previous studies have demonstrated that *Clostridia* are very common in the chicken ceca, our study shows that within this class there is a wide diversity of species present, something which has perhaps been underestimated by culture based studies. To gain a mechanistic insight into the function of these bacteria and to capture the wide-diversity of bacteria present in chickens, large-scale culture based studies will be necessary.

Acknowledgements:

We would like to thank the staff at the Greenwood Building, Roslin Institute for the care of our animals. We would also like to thank Denny Gorman for his help with sample preparations. The Roslin Institute forms part of the Royal (Dick) School of Veterinary Studies, University of Edinburgh. This project was supported by the Biotechnology and Biological Sciences Research Council, including institute strategic programme and national capability awards to The Roslin Institute (BBSRC: BB/P013759/1, BB/P013732/1, BB/J004235/1, BB/J004243/1). MJP is supported by the Quadram Institute Bioscience BBSRC-funded Strategic Program: Microbes in the Food Chain (Project No. BB/R012504/1) and its constituent project BBS/E/F/000PR10351 (Theme 3, Microbial Communities in the Food Chain) and by the Medical Research Council CLIMB grant (MR/L015080/1)

References:

1. Food, Nations AOotU. FAOSTAT database. FAO Rome, Italy; 2019.
2. Bennett CE, Thomas R, Williams M, Zalasiewicz J, Edgeworth M, Miller H, et al. The broiler chicken as a signal of a human reconfigured biosphere. R Soc Open Sci. 2018;5(12):11.
3. Ritchie H, Roser M. Meat and Seafood Production & Consumption 2019 [
4. OECD/FAO. Meat. OECD-FAO Agricultural Outlook 2016-2025: OECD Publishing, Paris; 2016.
5. Prall GFW, van der Steen HAM, Plastow GS. Application of genomics to the pork industry. Journal of Animal Science. 2005;83(suppl_13):E1-E8.
6. Gill M, Smith P, Wilkinson JM. Mitigating climate change: The role of domestic livestock. Animal. 2010;4(3):323-33.
7. Kabir SML. The role of probiotics in the poultry industry. Int J Mol Sci. 2009;10(8):3531-46.
8. Clavijo V, Florez MJV. The gastrointestinal microbiome and its association with the control of pathogens in broiler chicken production: A review. Poultry Science. 2018;97(3):1006-21.
9. Crhanova M, Hradecka H, Faldynova M, Matulova M, Havlickova H, Sisak F, et al. Immune response of chicken gut to natural colonization by gut microflora and to *Salmonella enterica* Serovar *Enteritidis* infection. Infection and Immunity. 2011;79(7):2755-63.
10. Jozefiak D, Rutkowski A, Martin SA. Carbohydrate fermentation in the avian ceca: A review. Animal Feed Science and Technology. 2004;113(1-4):1-15.
11. Karasawa Y. Significant role of the nitrogen recycling system through the ceca occurs in protein-depleted chickens. J Exp Zool. 1999;283(4-5):418-25.
12. Parsons CM, Potter LM, Brown RD. Effects of dietary carbohydrate and of intestinal microflora on excretion of endogenous amino-acids by poultry. Poultry Science. 1983;62(3):483-9.
13. Wen C, Yan W, Sun C, Ji C, Zhou Q, Zhang D, et al. The gut microbiota is largely independent of host genetics in regulating fat deposition in chickens. The ISME Journal. 2019.
14. Mignon-Grasteau S, Narcy A, Rideau N, Chantry-Darmon C, Boscher M-Y, Sellier N, et al. Impact of selection for digestive efficiency on microbiota composition in the chicken. PLOS ONE. 2015;10(8):e0135488.
15. Stanley D, Denman SE, Hughes RJ, Geier MS, Crowley TM, Chen HL, et al. Intestinal microbiota associated with differential feed conversion efficiency in chickens. Applied Microbiology and Biotechnology. 2012;96(5):1361-9.
16. Svihus B, Choct M, Classen HL. Function and nutritional roles of the avian caeca: A review. Worlds Poult Sci J. 2013;69(2):249-63.

17. Watson M. Illuminating the future of DNA sequencing. Genome Biol. 2014;15(2):2.
18. Loman NJ, Watson M. Successful test launch for nanopore sequencing. Nat Methods. 2015;12(4):303-4.
19. Medvecky M, Cejkova D, Polansky O, Karasova D, Kubasova T, Cizek A, et al. Whole genome sequencing and function prediction of 133 gut anaerobes isolated from chicken caecum in pure cultures. BMC Genomics. 2018;19(1):561.
20. Sangwan N, Xia FF, Gilbert JA. Recovering complete and draft population genomes from metagenome datasets. Microbiome. 2016;4:11.
21. Kang DWD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015;3:15.
22. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell. 2019;176(3):649-+.
23. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, et al. A new genomic blueprint of the human gut microbiota. Nature. 2019;568(7753):499-+.
24. Stewart RD, Auffret MD, Warr A, Walker AW, Roehe R, Watson M. The genomic and proteomic landscape of the rumen microbiome revealed by comprehensive genome-resolved metagenomics. bioRxiv. 2018:489443.
25. Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, et al. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. Nat Commun. 2018;9:11.
26. Wang W, Hu H, Zijlstra RT, Zheng J, Gänzle MG. Metagenomic reconstructions of gut microbial metabolism in weanling pigs. Microbiome. 2019;7(1):48.
27. Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. Untangling genomes from metagenomes: Revealing an uncultured class of marine euryarchaeota. Science. 2012;335(6068):587-90.
28. Hugerth LW, Larsson J, Alneberg J, Lindh MV, Legrand C, Pinhassi J, et al. Metagenome-assembled genomes uncover a global brackish microbiome. Genome Biol. 2015;16:18.
29. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat Commun. 2016;7:11.
30. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2017;2(11):1533-42.
31. Wu SB, Stanley D, Rodgers N, Swick RA, Moore RJ. Two necrotic enteritis predisposing factors, dietary fishmeal and Eimeria infection, induce large changes in the caecal microbiota of broiler chickens. Vet Microbiol. 2014;169(3-4):188-97.
32. Stanley D, Wu SB, Rodgers N, Swick RA, Moore RJ. Differential responses of cecal microbiota to fishmeal, *Eimeria* and *Clostridium perfringens* in a necrotic enteritis challenge model in chickens. Plos One. 2014;9(8):10.
33. Glendinning L, Wright S, Pollock J, Tennant P, Collie D, McLachlan G. Variability of the sheep lung microbiota. Appl Environ Microbiol. 2016;82(11):3225-38.
34. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.
35. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics. 2012;28(11):1420-8.
36. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997. 2013.
37. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
38. Li DH, Liu CM, Luo RB, Sadakane K, Lam TW. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674-6.

39. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. Isme J. 2017;11(12):2864-8.
40. Parks D. CompareM. https://github.com/dparks1134/CompareM2019.
41. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research. 2015;25(7):1043-55.
42. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018;36(10):996-+.
43. Stewart RD, Auffret MD, Snelling TJ, Roehe R, Watson M. MAGpy: A reproducible pipeline for the downstream analysis of metagenome-assembled genomes (MAGs). Bioinformatics. 2018:bty905.
44. Rambaut A. FigTree v1. 4. http://treebioedacuk/software/figtree2012.
45. Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ. 2015;3:17.
46. Luo CW, Rodriguez LM, Konstantinidis KT. MyTaxa: An advanced taxonomic classifier for genomic and metagenomic sequences. Nucleic Acids Res. 2014;42(8):12.
47. Parker CT, Tindall BJ, Garrity GM. International code of nomenclature of prokaryotes: Prokaryotic code (2008 revision). Int J Syst Evol Microbiol. 2019;69(1A):S7-S111.
48. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. Nucleic Acids Res. 2009;37:D233-D8.
49. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59-60.
50. Team RC. R: A language and environment for statistical computing. 2013.
51. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. R package version 25-2 https://CRANR-projectorg/package=vegan2013.
52. Wickham H. ggplot2: Elegant graphics for data analysis. http://ggplot2org2016.
53. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: Visualization of intersecting sets. IEEE Trans Vis Comput Graph. 2014;20(12):1983-92.
54. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):38.
55. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome- assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2018;36(7):660-.
56. van der Wielen P, Rovers G, Scheepens JMA, Biesterveld S. *Clostridium lactatifermentans* sp nov., a lactate-fermenting anaerobe isolated from the caeca of a chicken. Int J Syst Evol Microbiol. 2002;52:921-5.
57. Boerlin P, Nicholson V, Brash M, Slavic D, Boyen F, Sanei B, et al. Diversity of *Enterococcus cecorum* from chickens. Vet Microbiol. 2012;157(3-4):405-11.
58. Awad WA, Mann E, Dzieciol M, Hess C, Schmitz-Esser S, Wagner M, et al. Age-related differences in the luminal and mucosa-associated gut microbiome of broiler chickens and shifts associated with *Campylobacter jejuni* infection. Frontiers in Cellular and Infection Microbiology. 2016;6(154).
59. Yu H, Si W, Gong J, Forster RJ, Yang C, Huang R, et al. 16S rRNA gene-based analysis of mucosa-associated bacterial community and phylogeny in the chicken gastrointestinal tracts: From crops to ceca. FEMS Microbiology Ecology. 2007;59(1):147-57.
60. Duggett NA, Kay GL, Sergeant MJ, Bedford M, Constantinidou CI, Penn CW, et al. Draft genome sequences of six novel bacterial isolates from chicken ceca. Genome Announcements. 2016;4(3):e00448-16.
61. Stanley J, Linton D, Burnens AP, Dewhirst FE, On SLW, Porter A, et al. *Helicobacter pullorum* sp. nov. - genotype and phenotype of a new species isolated from poultry and from human patients with gastroenteritis. Microbiol-UK. 1994;140:3441-9.
62. Lu J, Idris U, Harmon B, Hofacre C, Maurer JJ, Lee MD. Diversity and succession of the intestinal bacterial community of the maturing broiler chicken. Appl Environ Microbiol. 2003;69(11):6816.

552 63. Neal-McKinney JM, Lu X, Duong T, Larson CL, Call DR, Shah DH, et al. Production of organic acids by
553 probiotic Lactobacilli can be used to reduce pathogen load in poultry. PLOS ONE. 2012;7(9):e43928.
554 64. Beasley SS, Takala TM, Reunanen J, Apajalahti J, Saris PEJ. Characterization and
555 electrotransformation of *Lactobacillus Crispatus* isolated from chicken crop and intestine. Poultry Science.
556 2004;83(1):45-8.
557 65. Fujisawa T, Benno Y, Yaeshima T, Mitsuoka T. Taxonomic study of the lactobacillus-acidophilus
558 group, with recognition of *lactobacillus-gallinarum* sp-nov and *lactobacillus-johnsonii* sp-nov and synonymy
559 of *lactobacillus-acidophilus* group-a3 (johnson et-al 1980) with the type strain of *lactobacillus-amylovorus*
560 (nakamura 1981). Int J Syst Bacteriol. 1992;42(3):487-91.
561 66. Taheri HR, Moravej H, Tabandeh F, Zaghari M, Shivazad M. Efficacy of combined or single use of
562 *Lactobacillus crispatus* LT116 and *L. johnsonii* LT171 on broiler performance British Poultry Science.
563 2010;51(5):580-5.
564 67. Bjerrum L, Engberg RM, Leser TD, Jensen BB, Finster K, Pedersen K. Microbial community
565 composition of the ileum and cecum of broiler chickens as revealed by molecular and culture-based
566 techniques. Poultry Science. 2006;85(7):1151-64.
567 68. Dec M, Nowaczek A, Urban-Chmiel R, Stepien-Pysniak D, Wernicki A. Probiotic potential of
568 Lactobacillus isolates of chicken origin with anti-Campylobacter activity. J Vet Med Sci. 2018;80(8):1195-
569 203.
570 69. Yu B, Liu JR, Hsiao FS, Chiou PWS. Evaluation of *Lactobacillus reuteri* Pg4 strain expressing
571 heterologous β-glucanase as a probiotic in poultry diets based on barley. Animal Feed Science and
572 Technology. 2008;141(1):82-91.
573 70. Saint-Cyr MJ, Haddad N, Taminiau B, Poezevara T, Quesne S, Amelot M, et al. Use of the potential
574 probiotic strain *Lactobacillus salivarius* SMXD51 to control *Campylobacter jejuni* in broilers. International
575 Journal of Food Microbiology. 2017;247:9-17.
576 71. Orata FD, Meier-Kolthoff JP, Sauvageau D, Stein LY. Phylogenomic analysis of the
577 Gammaproteobacterial Methanotrophs (order Methylococcales) calls for the reclassification of members at
578 the genus and species levels. Front Microbiol. 2018;9:17.
579 72. Konstantinidis KT, Tiedje JM. Towards a genome-based taxonomy for prokaryotes. J Bacteriol.
580 2005;187(18):6258-64.
581 73. Zanoni RG, Rossi M, Giacomucci D, Sanguinetti V, Manfreda G. Occurrence and antibiotic
582 susceptibility of *Helicobacter pullorum* from broiler chickens and commercial laying hens in Italy.
583 International Journal of Food Microbiology. 2007;116(1):168-73.
584 74. Ceelen LM, Decostere A, Van den Buick K, On SLW, Baele M, Ducatelle R, et al. *Helicobacter*
585 *pullorum* in chickens, Belgium. Emerg Infect Dis. 2006;12(2):263-7.
586 75. Kaakoush NO, Sodhi N, Chenu JW, Cox JM, Riordan SM, Mitchell HM. The interplay between
587 *Campylobacter* and *Helicobacter* species and other gastrointestinal microbiota of commercial broiler
588 chickens. Gut Pathogens. 2014;6:10.
589 76. Lei F, Yin YS, Wang YZ, Deng B, Yu HD, Li LJ, et al. Higher-level production of volatile fatty acids in
590 vitro by chicken gut microbiotas than by human gut microbiotas as determined by functional analyses. Appl
591 Environ Microbiol. 2012;78(16):5763-72.
592 77. Sergeant MJ, Constantinidou C, Cogan TA, Bedford MR, Penn CW, Pallen MJ. Extensive microbial
593 and functional diversity within the chicken cecal microbiome. Plos One. 2014;9(3):13.
594 78. Ballou AL, Ali RA, Mendoza MA, Ellis JC, Hassan HM, Croom WJ, et al. Development of the chick
595 microbiome: How early exposure influences Future Microbial Diversity. Front Vet Sci. 2016;3:12.
596 79. Oakley BB, Buhr RJ, Ritz CW, Kiepper BH, Berrang ME, Seal BS, et al. Successional changes in the
597 chicken cecal microbiome during 42 days of growth are independent of organic acid feed additives. BMC
598 Vet Res. 2014;10:8.
599 80. Sekelja M, Rud I, Knutsen SH, Denstadli V, Westereng B, Naes T, et al. Abrupt temporal fluctuations
600 in the chicken fecal microbiota are explained by Its gastrointestinal origin. Appl Environ Microbiol.
601 2012;78(8):2941-8.
602 81. Stanley D, Geier MS, Denman SE, Haring VR, Crowley TM, Hughes RJ, et al. Identification of chicken
603 intestinal microbiota correlated with the efficiency of energy extraction from feed. Vet Microbiol.
604 2013;164(1-2):85-92.

605 82. Wei S, Morrison M, Yu Z. Bacterial census of poultry intestinal microbiome. Poultry Science.
606 2013;92(3):671-83.
607 83. Hieke ASC, Hubert SM, Athrey G. Circadian disruption and divergent microbiota acquisition under
608 extended photoperiod regimens in chicken. PeerJ. 2019;7:28.
609 84. Kubasova T, Kollarcikova M, Crhanova M, Karasova D, Cejkova D, Sebkova A, et al. Contact with
610 adult hen affects development of caecal microbiota in newly hatched chicks. Plos One. 2019;14(3):13.
611 85. Ferrario C, Alessandri G, Mancabelli L, Gering E, Mangifesta M, Milani C, et al. Untangling the cecal
612 microbiota of feral chickens by culturomic and metagenomic analyses. Environ Microbiol.
613 2017;19(11):4771-83.
614 86. Stanley D, Geier MS, Hughes RJ, Denman SE, Moore RJ. Highly variable microbiota development in
615 the chicken gastrointestinal tract. Plos One. 2013;8(12):7.
616 87. Videnska P, Sedlar K, Lukac M, Faldynova M, Gerzova L, Cejkova D, et al. Succession and
617 replacement of bacterial populations in the caecum of egg laying hens over their whole life. Plos One.
618 2014;9(12):14.
619 88. Mancabelli L, Ferrario C, Milani C, Mangifesta M, Turroni F, Duranti S, et al. Insights into the
620 biodiversity of the gut microbiota of broiler chickens. Environ Microbiol. 2016;18(12):4727-38.
621 89. Ding JM, Zhao LL, Wang LF, Zhao WJ, Zhai ZX, Leng L, et al. Divergent selection-induced obesity
622 alters the composition and functional pathways of chicken gut microbiota. Genet Sel Evol. 2016;48:9.
623 90. Ocejo M, Oporto B, Hurtado A. 16S rRNA amplicon sequencing characterization of caecal
624 microbiome composition of broilers and free-range slow-growing chickens throughout their productive
625 lifespan. Sci Rep. 2019;9:14.
626 91. Pandit RJ, Hinsu AT, Patel NV, Koringa PG, Jakhesara SJ, Thakkar JR, et al. Microbial diversity and
627 community composition of caecal microbiota in commercial and indigenous Indian chickens determined
628 using 16s rDNA amplicon sequencing. Microbiome. 2018;6:13.
629 92. Jebava I, Chuat V, Lortal S, Valence F. Peptidoglycan hydrolases as species-specific markers to
630 differentiate *Lactobacillus helveticus* from *Lactobacillus gallinarum* and other closely related
631 homofermentative Lactobacilli. Curr Microbiol. 2014;68(4):551-7.
632 93. Saminathan M, Sieo CC, Kalavathy R, Abdullah N, Ho YW. Effect of prebiotic oligosaccharides on
633 growth of *Lactobacillus* strains used as a probiotic for chickens. Afr J Microbiol Res. 2011;5(1):57-64.
634 94. Askelson TE, Campasino A, Lee JT, Duong T. Evaluation of phytate-degrading *Lactobacillus* culture
635 administration to broiler chickens. Appl Environ Microbiol. 2014;80(3):943-50.
636 95. Manfreda G, Parisi A, Lucchi A, Zanoni RG, De Cesare A. Prevalence of *Helicobacter pullorum* in
637 conventional, organic, and free-range broilers and typing of isolates. Appl Environ Microbiol.
638 2011;77(2):479-84.
639 96. Eeckhaut V, Van Immerseel F, Croubels S, De Baere S, Haesebrouck F, Ducatelle R, et al. Butyrate
640 production in phylogenetically diverse Firmicutes isolated from the chicken caecum. Microb Biotechnol.
641 2011;4(4):503-12.
642 97. Polansky O, Sekelova Z, Faldynova M, Sebkova A, Sisak F, Rychlik I. Important metabolic pathways
643 and biological processes expressed by chicken cecal microbiota. Appl Environ Microbiol. 2016;82(5):1569-
644 76.

645

646

647 **Dataset legends:**

648 **Dataset 1:** Average coverage of MAGs in all samples. Coverage was calculated by mapping MAG scaffolds to
649 the adaptor trimmed Illumina reads for each sample. The average coverage of the scaffolds from a mag within
650 a sample were taken as the average abundance of that mag in the sample.

651 **Dataset 2:** Description of each chicken MAG (metagenome-assembled genome), including novelty of species
652 or strain, NCBI_name, GTDB-Tk _taxonomy, CheckM completeness and contamination, assembly size (mb),
653 N50, number of contigs, the longest contig length (bp) and the GC content.

654 **Dataset 3**: Taxonomy assigned by MAGpy to MAGs.

655 **Dataset 4**: Clustering of samples at 60% AAI to form genus clusters. Novel genera were defined as clusters of
656 MAGs at 60% AAI which were not assigned a genus by GTDB-Tk

657 **Dataset 5:** MAGs which were identified as being significantly more abundant by DESeq2 between diets and
658 lines.

659 **Dataset 6:** CAZymes which were identified as being significantly more abundant by DESeq2 between diets
660 and lines.