Bacterial meta-analysis of chicken cecal microbiota

Luis Alberto Chica Cardenas 1, 2, Viviana Clavijo 3, Martha Vives 3, Alejandro Reyes 1, 2, 4

1 Research Group on Computational Biology and Microbial Ecology, Department of Biological Sciences, Universidad de Los Andes, Bogotá, Colombia
2 Max Planck Tandem Group in Computational Biology, Universidad de Los Andes, Bogotá, Colombia
3 Centro de Investigaciones Microbiológicas, Universidad de Los Andes, Bogotá, Colombia
4 The Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, Missouri, United States

Corresponding Author: Alejandro Reyes
Email address: a.reyes@uniandes.edu.co

Poultry production is an industry that generates 90,000 metric tons of chicken meat worldwide. Thus, optimizing chicken growth and sustainable production is of great importance. A central factor determining not only production parameters, but also stability of the immune system and chicken health, is the diversity and variability of the microbiota present throughout the gastrointestinal tract. To date, several studies have investigated the relationship between bacterial communities and the gut microbiome, with limited data to compare. This study aims to create a bacterial meta-analysis based on studies using amplicon sequencing with Illumina sequencing technologies in order to build a baseline for comparison in future analyses of the cecal bacterial composition in chicken. A systematic literature review was performed (SYRF ID: e84f0468-e418-4eec-9da4-b517f1b4809d. Full project URL: https://app.syrf.org.uk/projects/e84f0468-e418-4eec-9da4-b517f1b4809d/detail ). From the all available and analyzed manuscripts only 9 contained full raw-sequence data available and the corresponding metadata. A total of 324 samples, comprising three different regions within the 16S rRNA gene, were analyzed. Due to the heterogeneity of the data, each region was analyzed independently and an effort for a joint analysis was performed as well. Taxonomic profiling revealed 11 phyla, with Firmicutes as the most prevalent phylum, followed by Bacteroidetes and Proteobacteria. At genus level, 109 genera were found. Shannon metric for alpha diversity showed that factors like type of chickens (Commercial or experimental) and 16S rRNA gene subregion have negligible effect on diversity. Despite the large number of parameters that were taken into account, the identification of common bacteria showed 5 genera to be common for all sets in at least 50% of the samples. These genera are highly associated to cellulose degradation and short chain fatty acids synthesis. In general, it was possible to identify some commonalities in the bacterial cecal microbial community despite the extensive variability and factors
differing from one study to another.
Bacterial meta-analysis of chicken cecal microbiota

Luis Alberto Chica Cardenas¹,², Viviana Clavijo³, Martha Josefina Vives Flóres³, Alejandro Reyes Muñoz¹,²,⁴

¹ Research Group on Computational Biology and Microbial Ecology, Department of Biological Sciences, Universidad de los Andes, Bogotá, Colombia.
² Max Planck Tandem Group in Computational Biology, Universidad de los Andes, Bogotá, Colombia.
³ Centro de Investigaciones Microbiológicas, Universidad de los Andes, Bogotá, Colombia.
⁴ The Edilson Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine, Saint Louis, MO, U.S.A

Corresponding Author:
Alejandro Reyes Muñoz¹,²,⁴
Carrera 1 #18A-12 Of. Ch201, Bogotá, Cundinamarca, 111711, Colombia
Email address: a.reyes@uniandes.edu.co
Abstract

Poultry production is an industry that generates 90,000 metric tons of chicken meat worldwide. Thus, optimizing chicken growth and sustainable production is of great importance. A central factor determining not only production parameters, but also stability of the immune system and chicken health, is the diversity and variability of the microbiota present throughout the gastrointestinal tract. To date, several studies have investigated the relationship between bacterial communities and the gut microbiome, with limited data to compare. This study aims to create a bacterial meta-analysis based on studies using amplicon sequencing with Illumina sequencing technologies in order to build a baseline for comparison in future analyses of the cecal bacterial composition in chicken. A systematic literature review was performed (SYRF ID: e84f0468-e418-4eec-9da4-b517f1b4809d. Full project URL: https://app.syrf.org.uk/projects/e84f0468-e418-4eec-9da4-b517f1b4809d/detail). From the all available and analyzed manuscripts only 9 contained full raw-sequence data available and the corresponding metadata. A total of 324 samples, comprising three different regions within the 16S rRNA gene, were analyzed. Due to the heterogeneity of the data, each region was analyzed independently and an effort for a joint analysis was performed as well. Taxonomic profiling revealed 11 phyla, with Firmicutes as the most prevalent phylum, followed by Bacteroidetes and Proteobacteria. At genus level, 109 genera were found. Shannon metric for alpha diversity showed that factors like type of chickens (Commercial or experimental) and 16S rRNA gene subregion have negligible effect on diversity. Despite the large number of parameters that were taken into account, the identification of common bacteria showed 5 genera to be common for all sets in at least 50% of the samples. These genera are highly associated to cellulose degradation and short chain fatty acids synthesis. In general, it was possible to identify some commonalities in the bacterial cecal microbial community despite the extensive variability and factors differing from one study to another.

Introduction

Chickens are considered to be one of the main sources of food production worldwide (Godfray et al., 2010). Chicken meat production accounts for more than 128 million tons per year worldwide (USDA Foreign Agricultural Service, 2019). It is also considered as the most efficient production of animal protein, due to its property of doubling the weight of food ingested into meat weight at the end of the production cycle. Furthermore, chicken meat has been reported as a source of highly digestible proteins, minerals and vitamins with low levels of saturated fatty acids, which accompanied with a balanced diet, might help reduce the risk of having cardiovascular and endocrine problems (Marangoni et al., 2015). Likewise, the efficiency of poultry meat production has had an impact on the price of the product, which make it more accessible for a broader number of social classes than other sources of meat protein (Scanes, 2007).

The microbial community (microbiota) present in the gastrointestinal tract (gut) has been widely associated with factors involving the health of chickens such as the immune system, the physiology of the digestive system and exclusion of pathogens, as well as the performance in production (Clavijo & Vives, 2017). Commensal bacteria have been positively associated with the generation and regulation of the mucus layer, which is involved in the protection of epithelial cells against pathogenic bacteria (Clavijo & Vives, 2017). For instance, chicken gut microbiota...
might have a direct effect on controlling the prevalence of pathogenic bacteria by competitive
exclusion (La Ragione & Woodward, 2003). Additionally, commensal bacterial metabolites,
such as short chain fatty acids (SCFA), are responsible for the expansion of the absorption
surface in the gut by increasing the number of its proliferating cells (Kien et al., 2007).

Among the different segments that comprise the chicken’s gut, the cecum is the place
where the food is retained the longest time. It presents the highest rate of water absorption and
bacterial diversity (Xiao et al., 2016 & Shaufi et al., 2015). Of all phyla that colonize the cecum,
*Firmicutes* and *Bacteroidetes* are reported to be the most abundant. Their prevalence has been
associated to their capacity for digesting cellulose and non-starch polysaccharides, which cannot
be digested in the small intestine and leads to SCFA production (Clavijo & Vives, 2017 &
Sergeant et al., 2014). Moreover, the *Firmicutes* that colonize the cecal region are also
responsible for nitrogen cycling, which is highly associated with the efficiency of chickens to
take energy from food (Mancabelli et al., 2016 & Oakley et al., 2014). Hence, most of the
information available on chicken gut microbiota has been focused on this organ.

For microbial diversity analyses based on 16S rRNA gene data it has been proven that the
use of different sequencing technologies may alter species richness and estimates of microbial
diversity for the same sample (Allali et al., 2017). Therefore, comparisons between samples
sequenced by different methods or bacterial consensus using data mixtures might result in
significant biases for further analysis. However, a bacterial consensus or meta-analysis might
provide useful information in order to evaluate the effect of different factors in the normal
composition of cecal microbiota. To the best of our knowledge, three meta-analysis of chicken
cecal composition has been performed. (Wei, Morrison & Yu, 2013; Waite & Taylor, 2014; Zou,
Sharik & Parkinson, 2018). Even though these studies were proposed as a good model for new
projects, the data used was retrieved from samples sequenced by Sanger technology (Wei,
Morrison & Yu, 2013), 454 pyrosequencing (Waite & Taylor, 2014) or by a mixture of data from
different methods (Zou, Sharik & Parkinson, 2018). Nowadays, Illumina sequence platforms are
the predominant methods to obtain biological sequence information. In consideration of the
biases that a comparison of reads obtained by different sequencing strategies (Plummer et al.,
2015), we performed a bacterial meta-analysis based on sequences generated by Illumina
technology with the aim of identifying bacteria, that despite the variations associated with each
study, are prevalent in the cecum and therefore, could be crucial in chicken gut modulation.

**Materials & Methods**

**Data collection and quality filters**

In order to retrieve the data, an extensive literature search using “cecal microbiota
composition in chickens” as search term was performed. Moreover, keywords such as “cecum”,
“broiler chicken”, “microbiota composition”, “16S rRNA” and “Illumina” were also taken into
account. Different search engines (PubMed, Google Scholar, Elsevier) were used to conduct the
search. Literature search was performed exclusively for english written papers and conducted
until February 25, 2020. For the systematic search developed in this study, authors Luis Alberto
Chica and Alejandro Reyes serve as screeners of the evaluated studies, whereas Martha Vives took the place of referee. However, there were no disagreements between screeners. A reverse search strategy, applying the same search terms described above, was made in SRA database in order to find data that might have been excluded in the initial search. Review articles were also checked for additional studies.

With the aim of choosing the most relevant studies, the following filters were used as inclusion parameters: i) data must be available in public databases, ii) metadata must be provided within the manuscript or in the database where the data was published and iii) the number of samples reported in the article must coincide with the data submitted to the database. Those three steps are encompassed within a two-step exclusion criteria according to PRISMA guidelines (Moher et al., 2009), in which the screen criteria encompasses all studies that did not have available data and the eligibility criteria states for all studies in which the metadata was absent or the number of samples reported was different to the data submitted in public databases (Fig. 1).

The entire set of 16S rRNA gene sequences used for this analysis was retrieved from the Sequence Read Archive (SRA) database (Leinonen et al., 2010) and downloaded using the SRA Toolkit software (Leinonen et al., 2010). Trimmomatic software (Bolger, Lohse & Usadel, 2014) was used to filter low quality reads using a sliding window of 4 and a minimum Phred score of 20. Due to the lack of reported information regarding barcodes, primers and adapters for the different samples, a headcrop of 15 bp was performed for all samples. In order to validate trimming procedures, we conducted a primer search in raw and clean reads by using the specific primers reported in each study and the software Seqkit v0.14 (Shen et al., 2016) (Table S1)

**ASVs assignment**

Due to the diversity of the hypervariable regions sequenced by the different studies, we split the entire dataset in 3 subsets. Each subset comprises samples generated with varying primers and corresponding to different hypervariable regions (V3, V4, V3V4). Deblur software (Amir et al., 2017) implemented in Qiime2 (Bolyen et al., 2018) was used for performing denoising and dereplication of the reads by applying a truncation length of 130 bp. ASV tables generated by deblur were filtered by a minimum frequency of 10, where singletons and doubletons were also excluded. In order to compare the taxa identified by the different amplicon regions, ASV tables were summarized at the genus level using sklearn classifier for taxonomic classification. GreenGenes database v13.18 (DeSantis et al., 2006) was used for training the classifier. Subsequently, all taxa that did not reach genus level were removed from collapsed tables. Finally, filtered tables resulting from the 3 datasets were merged together.

**Core microbiome generation, taxonomic evaluation and diversity estimates**

Independent core microbiome analysis was performed for all ASVs tables that were collapsed at genus level. Two approaches to define the core microbiome were assessed. i) a strict approach closer to the general definition of core, where a genus should be present in at least 80% of the samples and ii) a more relaxed search requiring a genus to be present in more than
50% of the samples. By using an in-house script, Venn diagrams were constructed with the aim of relating common genera present on the 3 hypervariable regions. Moreover, a core microbiome analysis per study was performed, selecting genera that were present in at least 80% of the samples of each given study. A comparison of the study specific core microbiomes was then performed to identify taxa that were present in the nine cores microbiome. The merged ASV table was transformed into relative frequency for computing taxonomic heatmaps at phylum and genus levels. For diversity estimates, rarefaction was performed at a sequence depth of 500. Alpha diversity was estimated using Shannon index and Observed species metrics. Bray-Curtis metric was employed to calculate Beta diversity. The significance of alfa diversity differences was evaluated by a Kruskall Wallis test. Meanwhile, the influence of different hypervariable regions on beta diversity estimates, was evaluated by a permutational analysis of variance (PERMANOVA), both tests as implemented in Qiime2.

**Results**

The literature search resulted in 27 studies that contained information fulfilling the search criteria ([Table 1](#)). After applying the filters described previously (see Methods), nine studies were selected to perform this meta-analysis ([Fig. 1](#)). Those studies contained a total of 324 samples and 36,743,185 of single end sequences, with a number of sequences per sample varying between 15,181 and 1,548,296. After applying sequencing quality filters, the number of sequences was reduced to 8,843,573. Among the different hypervariable regions, 4,178,126 sequences covered the V4 16S rRNA gene region, 1,897,781 the V3 gene and 3,280,185 covered both the V3 and V4 region (V3V4) ([Table S2](#)).

After denoising the reads, filtering the resulting ASV tables, collapsing all taxa at genus level and merging the ASVs, 109 genera and 11 phyla were identified. **Firmicutes** was the most prevalent phylum, followed by **Bacteroidetes** and **Proteobacteria**. In some cases, a given taxonomic profile was highly related to a single study, however, a similar abundance profile of **Firmicutes** and **Proteobacteria** was seen in more than 95% of the samples ([Fig. 2A](#)). At genus level **Oscillospira**, **Bacteroides**, **Helicobacter** and **Lactobacillus** were found as the most representative genera ([Fig. 2B](#)).

For the evaluation of common bacterial taxa, core microbiome of the samples was assessed by two different approaches. Common bacteria in (i) at least 50% and (ii) at least 80% of the samples were evaluated. The traditional cut-off of 80% abundance for core definition showed **Oscillospira** as the only genus that was identified in all 3 regions ([Fig. 3A](#)). After relaxing the threshold and evaluating the 50% core microbiome at genus level, 5 genera were found to be common for all sets (**Oscillospira**, **Lactobacillus**, **Faecalibacterium**, **Clostridium** and **Ruminococcus**). In addition, 6 genera were shared by V3 and V4 regions, 3 were exclusively shared between V3 and V3V4 and 2 genera were common for V4 and V3V4 regions ([Fig. 3B](#)). It is important to highlight that all genera that were exclusively found in a particular region were present in all studies concerning that specific region, but not necessarily in all samples ([Table](#)).
Finally, we evaluated the core microbiome per study, here, *Oscillospira* remained as the only genus shared by all the studies. Interestingly, multiple taxa were part of the core microbiome of several but one or two studies. Overall, *Oscillospira, Faecalibacterium, Lactobacillus, Clostridium, Bacteroides, Blautia, Ruminococcus* and *Coprobacillus* were present in the core of 6 studies or more. *(Fig. 3 C & Table S4)*, which includes all genera identified at the 50% overall core-microbiome.

Due to the high number of metadata categories considered in this analysis, alpha diversity variation was evaluated based on four categories: i) *16S rRNA* gene hypervariable region ii) study, iii) type of chickens (chickens under commercial rearing conditions or chickens rearing under controlled environment “Experimental”) and iv) chicken breed. Variables such as age, due to the number of studies with missing data make it impossible to consider it as a variable for comparison. Alpha rarefaction plot for all samples using a depth of 500 sequences per sample showed that all samples reached a saturation asymptote, which confirms that the samples can be used for further analyses and confirming a relatively low diversity in the chicken cecum *(Fig. S1)*. Rarefaction process discarded 33 samples that did not fulfill the minimum sampling depth, leaving 290 samples for the diversity analyses.

Most variation for alpha diversity was observed when comparing among the different studies. Interestingly, one of the groups of samples that presented low diversity values was a subset of the study of Varmuzoba et al. (2015), which contains chickens infected with *Salmonella enteriditis*, showing a significant reduction *(Wilcoxon, p-value = 0.009)*, compared with healthy chickens of the same study and healthy chickens from other studies *(Fig. 4A)*. Likewise, when analyzing the effect of the selected hypervariable region, we noticed a significant difference on Shannon values for all groups *(Kruskal Wallis, p-value = 1.9e-19)*, being V4 region with the higher values *(Fig. 4B)*. In order to prove if the difference presented in the V3V4 group was a consequence of the infected chickens that were previously described, we removed those samples from the dataset and performed again the analysis obtaining similar results. The comparison of filtered V3V4 group against V4 *(Wilcoxon, p-value = 1.2e-13)* and V3 *(Wilcoxon, p-value = 0.0067)* groups independently, still showed significant differences on both cases. On the other hand, comparisons between commercial and experimental chickens reveal no significant variation *(Wilcoxon, p-value = 0.188)* *(Fig. 4C)*. When using the breed of the chickens as variable, even though significant differences were observed, we cannot be sure of its relevance due to the lack of repeating breeds over different studies *(Fig. 4D)*.

By using Bray Curtis metric, distance matrices were built, and beta diversity was estimated for the same categories evaluated above. No strong clustering was observed by any of the variables analyzed *(Fig. 5)*, however, some recognizable clustering was observed for the samples belonging to Mancabelli et al. (2016) and Costa et al. (2017). Further statistical analysis using Permanova pairwise method was applied on the variable Study and hypervariable region, both analyses showed significant differences. Whereas all pairwise comparisons performed to the variable study reported q-values ranging between 0.00125 and 0.423, being the comparison of two groups of the same study (control and infected chickens) the only with no significant
differences, the comparisons applied to the hypervariable region variable report all q-values of 0.001.

**Discussion**

The aim of this study was to create a baseline for future research in order to give a higher insight into the general composition of the microbial communities in the cecum of chickens. This analysis aimed to collect the largest possible dataset of studies describing bacterial cecal content using Illumina technologies and amplicon sequencing.

Unfortunately, the study was limited since many of the published studies did not have their data available in public databases or did not share the metadata related with each project, making impossible to use the data or evaluate and replicate the results. Problems such as the lack of data available and the high number of sequences that were excluded after the quality filters were applied can be avoided by implementing common quality standards in the experimental design, data generation and data deposition. Only with serious commitment to high standards it is possible to achieve reproducible and comparable results among different studies. Another important variable is that there is still no consensus on the specific hypervariable region of the 16S rRNA gene to focus. In order to compare the different results obtained it was needed to collapse the ASVs tables at genus level and merge them.

After taxonomic classification of the 3 sets of samples (different hypervariable regions) at phylum and genus level, the results showed the predominance of Firmicutes and Bacteroidetes in most of the samples. These results are concordant with previously reported consensus based on sequences generated by different sequencing platforms (Wei, Morrison & Yu, 2013). When looking at the genus level, the number of genera reported were much higher in this meta-analysis than what was reported for individual studies (Xu et al, 2016; Zhou et al., 2016 & Biasalto et al., 2018) and comparable with the meta-analysis of Sanger generated sequences, performed by Wei, Morrison & Yu (2013), in which 117 genera were found to be present in the chicken gastrointestinal tract. Thus, the increase in sensitivity, achieved by the higher number of sequences generated using Illumina sequencing technology, might lead to comparable results, even when fewer samples were used.

The behavior of diversity estimates for 16S rRNA gene regions showed that the V4 region presents the highest diversity values at ASV level when comparing with V3 and V3V4. This behavior is supported by experimental results (Sperling et al., 2017). On the other hand, when the V3V4 data set is separated in groups of control chickens and chickens infected with Salmonella enteriditis, control chickens show the same diversity behavior than animals belonging to the other studies, however differences when comparing V3V4 region against V3 and V4 regions were still present. Those differences are likely due to factors such as biases on the primer’s specificity, although the V3V4 primers used on the different studies did not have the same sequence, or the resolution obtained by the inter-specific variability of the amplified region. When comparing diversity estimates between commercial and experimental chickens no difference was found, leading to assume that variables like temperature, food sources, water
supply, housing, hygiene, and the number of animals per cage may not have as significant effects as other variables on bacterial diversity, although it is important to highlight that only two studies encompass commercial samples, from which a single study groups 92% of the samples, leading to a potential bias. Despite the large number of variables inherent to each study, the collapse of ASVs tables to genus level allowed to diminish the influence of those variables in beta diversity analysis and led to the clusterization of several samples corresponding to different studies. Although, neither of the metadata categories considered in this metanalysis could group the samples, leading to assume that beta diversity grouping is driven by multiple categories.

The evaluation of common bacteria, in different percentages of the samples at the genus level, revealed the presence of several bacterial genera with an important role in food conversion. The genera *Faecalibacterium*, *Oscillospira*, *Lactobacillus*, *Clostridium* and *Ruminococcus* were found to be present in more than 50% of the samples. The importance of *Faecalibacterium* has been attributed to their capacity to express acetyl-CoA acetyltransferase and several enzymes involved in the production of butyrate (Polansky et al., 2016). Production of butyrate could also be accomplished by *Clostridium* and *Oscillospira* species and due to the role of butyrate on anti-inflammatory responses and grow performance, the remarkable prevalence of this genera could be explained (Biasato et al., 2018). On the other hand, *Ruminococcus* and *Lactobacillus* have been extensively described as major colonizers of cecum and ileum, and as part of *Ruminococcaceae* family, their role in SCFA production has been proposed (Wang, Lilburn & Yu, 2016; Rinttilä & Apajalahti, 2013; Wei, Morrison & Yu, 2013; Gophna, Konikoff & Nielsen, 2017). Moreover, the ability of *Ruminococcus* to degrade cellulose, can explain its major presence and abundance in the cecum (Devillard et al., 2004).

Beta diversity results using different studies as variable showed no clear differentiation of samples from the same study, only two studies showed this behavior. These findings differ with the analysis performed by Zou, Sharik & Parkinson (2018), in which the clustering of samples from the same study is strong. Is not clear the potential source of the variations in the results observed, but they might come from the different efforts in cleaning and normalizing the datasets in order to make them as comparable as possible thus reducing the biases inherent to experimental procedures. However, differences in the software used for the analysis and other technical factors could also have an influence in this variation.

Even though this study may have potential biases related to the data availability and different experimental procedures performed by different groups in different countries, it presents a useful model of comparison and baseline for future studies when cecal microbiome composition wants to be assessed (*Table S5, Additional File1*). Despite the heterogeneity of the samples, several bacterial genera, with implications in food conversions rate, were reported to be present in an important proportion of the sample.

**Conclusions**

Microbial communities’ meta-analysis is a very valuable tool to identify commonalities as well as sources of variation that structures the communities in different environmental
settings. Our study showed that the first and most important limitation is the availability of the data and thorough and complete description of the metadata associated. Once, those factors were taken into consideration, source of variation can be both technical and experimental, such as the choice of primers to use or the presence of certain pathogen as part of the experimental design, however, no single factor was the major driver of inter sample variation although in some cases sample from a given study clearly clustered together.

Finding commonalities among the different datasets depends of the taxonomical level of resolution desired, at phylum level we were able to detect 11 different phyla and a general structure with dominance of Firmicutes, Bacteroidetes and Proteobacteria. At higher taxonomical levels such as genus, although some genera were identified in a majority (>50%) of the samples overall or within studies, few of them were consistently present suggesting a functional redundancy among closely related bacteria. Most of these common bacteria have been associated with growth and health of different hosts. As we improve our methods of generating and reporting metagenomic and metabarcoding studies, more insights into the ecological and functional role of the different members of bacterial communities associated to a given host will be elucidated.

Acknowledgements

The authors would like to acknowledge doctors Maryam Chaib, Angela Viviana Peña and Juan Manuel Anzola for their help in the revision of the manuscript. Likewise, we thank the High-Performance Computing (HPC) Service at Universidad de los Andes, that allow us to perform the computational analysis.

References

Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., Kightley, E. P., Thompson, L. R., Hyde, E. R., Gonzalez, A., & Knight, R. (2017). Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *MSystems*, 2(2), e00191-16. [https://doi.org/10.1128/mSystems.00191-16](https://doi.org/10.1128/mSystems.00191-16)

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodriquez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8), 852–857. [https://doi.org/10.1038/s41587-019-0209-9](https://doi.org/10.1038/s41587-019-0209-9)

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu17
DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., & Andersen, G. L. (2006). Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*, 72(7), 5069 LP – 5072. https://doi.org/10.1128/AEM.03006-0.

Leinonen, R., Sugawara, H., Shumway, M., & International Nucleotide Sequence Database Collaboration. (2010). The sequence read archive. *Nucleic acids research*, 39(suppl_1), D19-D21.

Sperling, J. L., Silva-Brandão, K. L., Brandão, M. M., Lloyd, V. K., Dang, S., Davis, C. S., Sperling, F. A. H., & Magor, K. E. (2017). Comparison of bacterial 16S rRNA variable regions for microbiome surveys of ticks. *Ticks and Tick-Borne Diseases*, 8(4), 453–461. https://doi.org/https://doi.org/10.1016/j.ttbdis.2017.02.002.

Wei, S., Morrison, M., & Yu, Z. (2013). Bacterial census of poultry intestinal microbiome. *Poultry science*, 92(3), 671-683.

Waite, D. W., & Taylor, M. W. (2014). Characterizing the avian gut microbiota: membership, driving influences, and potential function. *Frontiers in microbiology*, 5, 223.

Polansky, O., Sekelova, Z., Faldynova, M., Sebkova, A., Sisak, F., & Rychlik, I. (2016). Important metabolic pathways and biological processes expressed by chicken cecal microbiota. *Applied and environmental microbiology*, 82(5), 1569-1576.

Wang, L., Lilburn, M., & Yu, Z. (2016). Intestinal microbiota of broiler chickens as affected by litter management regimens. *Frontiers in microbiology*, 7, 593.

Rinttilä, T., & Apajalahti, J. (2013). Intestinal microbiota and metabolites—Implications for broiler chicken health and performance1. *Journal of Applied Poultry Research*, 22(3), 647-658.

Devillard, E., Goodheart, D. B., Karnati, S. K. R., Bayer, E. A., Lamed, R., Miron, J., Nelson, K. E., & Morrison, M. (2004). Ruminococcus albus 8 Mutants Defective in Cellulose Degradation Are Deficient in Two Processive Endocellulases, Cel48A and Cel9B, Both of Which Possess a Novel Modular Architecture. *Journal of Bacteriology*, 186(1), 136 LP – 145. https://doi.org/10.1128/JB.186.1.136-145.2004.

Gophna, U., Konikoff, T., & Nielsen, H. B. (2017). Oscillospira and related bacteria—From metagenomic species to metabolic features. *Environmental microbiology*, 19(3), 835-841.
Marangoni, F., Corsello, G., Cricelli, C., Ferrara, N., Ghiselli, A., Lucchin, L., & Poli, A. (2015). Role of poultry meat in a balanced diet aimed at maintaining health and wellbeing: an Italian consensus document. Food & nutrition research, 59(1), 27606.

Scanes, C. G. (2007). The Global Importance of Poultry. Poultry Science, 86(6), 1057–1058. [https://doi.org/10.1093/ps/86.6.1057]

La Ragione, R. M., & Woodward, M. J. (2003). Competitive exclusion by Bacillus subtilis spores of Salmonella enterica serotype Enteritidis and Clostridium perfringens in young chickens. Veterinary microbiology, 94(3), 245-256.

Clavijo, V., & Flórez, M. J. V. (2017). The gastrointestinal microbiome and its association with the control of pathogens in broiler chicken production: A review. Poultry science.

Allali, I., Arnold, J. W., Roach, J., Cadenas, M. B., Butz, N., Hassan, H. M., Koci, M., Ballou, A., Mendoza, M., Ali, R., & Azcarate-Peril, M. A. (2017). A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. BMC Microbiology, 17(1), 194. [https://doi.org/10.1186/s12866-017-1101-8]

Xiao, Y., Xiang, Y., Zhou, W., Chen, J., Li, K., & Yang, H. (2016). Microbial community mapping in intestinal tract of broiler chicken. Poultry science, 96(5), 1387-1393.

Shaufi, M. A. M., Sieo, C. C., Chong, C. W., Gan, H. M., & Ho, Y. W. (2015). Deciphering chicken gut microbial dynamics based on high-throughput 16S rRNA metagenomics analyses. Gut pathogens, 7(1), 4.

Sergeant, M. J., Constantinidou, C., Cogan, T. A., Bedford, M. R., Penn, C. W., & Pallen, M. J. (2014). Extensive microbial and functional diversity within the chicken cecal microbiome. PloS one, 9(3), e91941.

Mancabelli, L., Ferrario, C., Milani, C., Mangifesta, M., Turroni, F., Duranti, S., Lugli, G. A., Viappiani, A., Ossiprandi, M. C., van Sinderen, D., & Ventura, M. (2016). Insights into the biodiversity of the gut microbiota of broiler chickens. Environmental Microbiology, 18(12), 4727–4738. [https://doi.org/10.1111/1462-2920.13363]

Oakley, B. B., Lilleshøj, H. S., Kogut, M. H., Kim, W. K., Maurer, J. J., Pedroso, A., Lee, M. D., Collett, S. R., Johnson, T. J., & Cox, N. A. (2014). The chicken gastrointestinal microbiome. FEMS Microbiology Letters, 360(2), 100–112. [https://doi.org/10.1111/1574-6968.12608]
Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food Security: The Challenge of Feeding 9 Billion People. *Science, 327*(5967), 812 LP – 818. https://doi.org/10.1126/science.1185383.

Varmuzova, K., Matulova, M. E., Gerzova, L., Cejkova, D., Gardan-Salmon, D., Panhéleux, M., Robert, F., Sisak, F., Havlickova, H., & Rychlik, I. (2015). Curcuma and Scutellaria plant extracts protect chickens against inflammation and Salmonella Enteritidis infection. *Poultry Science, 94*(9), 2049–2058. https://doi.org/https://doi.org/10.3382/ps/pev190.

Plummer, E., Twin, J., Bulach, D. M., Garland, S. M., & Tabrizi, S. N. (2015). A comparison of three bioinformatics pipelines for the analysis of preterm gut microbiota using 16S rRNA gene sequencing data. *Journal of Proteomics & Bioinformatics, 8*(12), 283.

Wang, J., Nesengani, L. T., Gong, Y., Yang, Y., & Lu, W. (2018). 16S rRNA gene sequencing reveals effects of photoperiod on cecal microbiota of broiler roosters. *PeerJ, 6*, e4390.

Kim, S., Rubinelli, P. M., Park, S. H., & Ricke, S. C. (2018). Ability of Arkansas LaKast and LaKast hybrid rice bran to reduce Salmonella Typhimurium in chicken cecal incubations and effects on cecal microbiota. *Frontiers in Microbiology, 9*, 134.

Xia, Y., Kong, J., Zhang, G., Zhang, X., Seviour, R., & Kong, Y. (2019). Effects of dietary supplementation with lysozyme on the structure and function of the cecal microbiota in broiler chickens. *PloS one, 14*(6).

Wu, S., Li, T., Niu, H., Zhu, Y., Liu, Y., Duan, Y., Sun, Q., & Yang, X. (2019). Effects of glucose oxidase on growth performance, gut function, and cecal microbiota of broiler chickens. *Poultry Science, 98*(2), 828–841. https://doi.org/https://doi.org/10.3382/ps/pey393.

Marimuthu, S. K., Balasubramanian, B., Selvam, R., & D'Souza, P. (2019). Modulation of chicken cecal microbiota by a phytogenic feed additive, Stodi®: A metagenomic analysis. *Pharmacognosy Research, 11*(3), 201.

Biasato, I., Ferrocino, I., Biasibetti, E., Grego, E., Dabbou, S., Sereno, A., Gai, F., Gasco, L., Schiavone, A., Cocolin, L., & Capucchio, M. T. (2018). Modulation of intestinal microbiota, morphology and mucin composition by dietary insect meal inclusion in free-range chickens. *BMC Veterinary Research, 14*(1), 383. https://doi.org/10.1186/s12917-018-1690-y.

Kollarcikova, M., Kubasova, T., Karasova, D., Crhanova, M., Cejkova, D., Sisak, F., & Rychlik, I. (2019). Use of 16S rRNA gene sequencing for prediction of new opportunistic pathogens in chicken ileal and cecal microbiota. *Poultry science, 98*(6), 2347-23.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS med, 6*(7), e1000097.

Xu, Y., Yang, H., Zhang, L., Su, Y., Shi, D., Xiao, H., & Tian, Y. (2016). High-throughput sequencing technology to reveal the composition and function of cecal microbiota in Dagu chicken. *BMC microbiology, 16*(1), 1-9.

Zhou, X., Jiang, X., Yang, C., Ma, B., Lei, C., Xu, C., ... & Men, S. (2016). Cecal microbiota of Tibetan Chickens from five geographic regions were determined by 16S rRNA sequencing. *Microbiologyopen, 5*(5), 753-762.

Zou, A., Sharif, S., & Parkinson, J. (2018). Lactobacillus elicits a 'Marmite effect' on the chicken cecal microbiome. *npj Biofilms and Microbiomes, 4*(1), 1-5.

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PloS one, 11*(10), e0163962.
Figure and Table legends:

Table 1. Characteristics of the manuscripts reviewed to perform the analyses. Each column represents one of the search parameters that were taken into account in order to select the most suitable studies. Articles that fulfilled each criterion were labeled with an X. The column titled “Used for this study” lists the studies selected for the analysis, whereas the last column contains the index assigned to each study, which is used as reference in different figures. NA refers to Not Applicable.

Figure 1. Flow chart showing the filtering steps, resulting in the 9 articles that were used for this meta-analysis. This flow chart was made according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al., 2009). Four steps are considered to be crucial for the correct evaluation of different records, in order to include them into the analysis. Whereas identification step only involves the finding of records and the elimination of redundant papers, in the screening step, all records in which data was not available are removed. After the screening process, the eligibility step involves the exclusion of all records that, even with available data, were removed for multiple reasons. For this article purpose, those reasons are considered as filters ii and iii (see methods: Data collection and quality filters). Finally, the inclusion step accounts for the final number of records used for the analysis.

Figure 2. Heat-map showing relative abundance of each sample from each study at different taxonomic assignments. (A) phylum and (B) genus level, for the merged ASV table. The most prevalent genera are shown (accounting for up to 90% of cumulative abundance). Index numbers at the bottom represent each study that was analyzed (Table 1). Relation of index numbers and metadata for each study is also shown on Table S1. In the y-axis, “L_Clostridium” refers to Clostridium genus assigned to Lachnospiraceae family and “E_Clostridium” refers to Clostridium genus assigned to Erysipelotrichaceae family.

Figure 3. Analysis of shared taxons in the three different hypervariable regions considered in the study. (A) Analysis at genus level, using a 80% threshold, which means that bacteria must be present in at least the 80% of the samples. (B) Genus level at 50% threshold. (C) Core microbiome per study using an 80% threshold. Number of Core genera per study are shown on
the horizontal bar plot, while the number of genera shared in the core of one or more studies are indicated in the vertical bar plots, the shared studies for each specific columns are shown by the connected dots in the lower panel.

**Figure 4. Evaluation of alpha diversity of the entire set of samples by using Shannon metric and Kruskal-Wallis statistics for significance.** Violin plots represent different groups for each category. (A) Study; (B) Hypervariable region amplified from the 16S rRNA gene. (C) Type of chickens (commercial, experimental). (D) Breed of the chickens, where Bh corresponds to Hubbard Hybrid breed and Broiler to all fattening breeds that are not specified. Boxplots represent the interquartile range and in red the median for each group. The number of samples in each group is shown right below the name. For Panel A, Indexes of metadata were used for visualization purposes and are related with the specific studies in Table S1. The colored violin plots correspond to Varmuzova study (2015), which is highlighted due to the inclusion of Salmonella enteriditis samples (green) that showed lower diversity than treated samples with plant extract (blue) and control samples (yellow) from the same study.

**Figure 5. PCoA showing the spatial dispersion of the samples. Distance matrix was constructed based on Bray-Curtis metric.** Coordinate plots represent (A) Study; The study of Varmuzova et al., 2015 was split due to the high intra-study variation obtained. Varmuzova_2015 = Control Chickens and uninfected Chickens treated with the plant extracts; Varmuzova_2015_I = Infected chickens; Varmuzova_2015_P_I = Infected chickens, treated with the plant extracts. (B) Hypervariable region of the 16S rRNA gene. (C) Type of chickens (commercial, experimental), when commercial chickens refer to animals raised on normal production environments and experimental refers to animals that were kept into environments with controlled conditions. (D) Breed of the animals, where Broiler corresponds to all fattening breeds that are not specified.

**SUPPLEMENTARY MATERIAL LEGENDS**

**Table S1: Number of reads per library containing primers before and after read cleanup.** SRA accession for each library is shown along with the counts of reads where 16S rRNA gene
amplification primers were found, along with the information related to the study from which was generated.

Table S2: Relation of indexes representing each set of samples in Table 1, Fig 2 and Fig 4A, with the relevant metadata belonging to the nine studies evaluated in the present meta-analysis. Information includes the 16S rRNA gene hypervariable region that was sequenced, the location where the study was developed, the breed of the animals, detailed information of the sequencing technology, amplification primers, amplicon size, extraction methodology, number of samples per study, number of reads before and after quality filters, a brief description of each project and the accession numbers for the raw data as well as the digital online identifier (doi) for each article.

Table S3: Proportion of samples from each study containing each genus that were found exclusively in a specific hypervariable 16S rRNA gene region at the different core thresholds (50% and 80%).

Table S4: List of all genera that are part of the core microbiome in one or more studies. An “X” was used to denote the presence of a specific taxa (rows) within the core microbiome of a given specific study (columns). The final column shows the count of the number of studies with a given taxon as part of their core microbiome.

Table S5: List of all ASVs that encompasses the taxonomic classification of all genera that were found to be part of the core microbiome per study and hypervariable region.

Figure S1: Rarefaction curves of all samples using different metrics. (A) Observed species and (B) Shannon index metrics.

Additional File1: Corresponding commands of Trimmomatic and Qiime2 that were employed in this analysis. The script is divided into three subsections (Trimmmomatic analysis, generation of ASVs tables and taxonomic classification for separated data sets and the analysis of the merged and genus collapsed table.
Table 1 (on next page)

Table 1. Characteristics of the manuscripts reviewed to perform the analyses.

Each column represents one of the search parameters that were taken into account in order to select the most suitable studies. Articles that fulfilled each criterion were labeled with an X. The column titled “Used for this study” lists the studies selected for the analysis, whereas the last column contains the index assigned to each study, which is used as reference in different figures. NA refers to Not Applicable.
| Study               | Available in public databases | Complete metadata | Information in databases according to article | Used for this study | Index |
|--------------------|-------------------------------|-------------------|-----------------------------------------------|---------------------|-------|
| Mancabelli et al., 2017 | x                              | x                 | x                                             | x                   | 1     |
| Shaufi et al., 2015   | x                              | x                 | x                                             | x                   | 2     |
| Ballou et al., 2016   | x                              | x                 | x                                             | x                   | 3     |
| Costa et al., 2017    | x                              | x                 | x                                             | x                   | 4     |
| Xu et al., 2016       | x                              | x                 | x                                             | x                   | 5     |
| Xia et al., 2019      | x                              | x                 | x                                             | x                   | 6     |
| Varmuzova et al., 2015| x                              | x                 | x                                             | x                   | 7     |
| Biasalto et al., 2018 | x                              | x                 | x                                             | x                   | 8     |
| Zhou et al., 2016     | x                              | x                 | x                                             | x                   | 9     |
| Awad et al., 2016     | x                              |                   |                                               | NA                  |       |
| Han et al., 2016      | x                              |                   |                                               | NA                  |       |
| Kim et al., 2018      |                               |                   |                                               | NA                  |       |
| Kollarckikova et al., 2019 |                   |                   |                                               | NA                  |       |
| Krueger et al., 2017  | x                              |                   |                                               | NA                  |       |
| Li et al., 2015       |                               |                   |                                               | NA                  |       |
| Ma et al., 2017       |                               |                   |                                               | NA                  |       |
| Marimuthu et al., 2020|                               |                   |                                               | NA                  |       |
| Mon et al., 2015      |                               |                   |                                               | NA                  |       |
| Park et al., 2016     |                               |                   |                                               | NA                  |       |
| Park et al., 2017     |                               |                   |                                               | NA                  |       |
| Polansky et al., 2016 |                               |                   |                                               | NA                  |       |
| Saxena et al., 2016   |                               |                   |                                               | NA                  |       |
| Varmuzova et al., 2016|                               |                   |                                               | NA                  |       |
| Wang et al., 2018     | x                              |                   |                                               | NA                  |       |
| Wu et al., 2019       |                               |                   |                                               | NA                  |       |
| Xiao et al., 2016     |                               |                   |                                               | NA                  |       |
| Yan et al., 2017      |                               |                   |                                               | NA                  |       |
Figure 1

Flow chart showing the filtering steps, resulting in the 9 articles that were used for this meta-analysis.

This flow chart was made according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al., 2009). Four steps are considered to be crucial for the correct evaluation of different records, in order to include them into the analysis. Whereas identification step only involves the finding of records and the elimination of redundant papers, in the screening step, all records in which data was not available are removed. After the screening process, the eligibility step involves the exclusion of all records that, even with available data, were removed for multiple reasons. For this article purpose, those reasons are considered as filters ii and iii (see methods: Data collection and quality filters). Finally, the inclusion step accounts for the final number of records used for the analysis.
Figure 2

Heat-map showing relative abundance of each sample from each study at different taxonomic assignments.

(A) phylum and (B) genus level, for the merged ASV table. The most prevalent genera are shown (accounting for up to 90% of cumulative abundance). Index numbers at the bottom represent each study that was analyzed (Table 1). Relation of index numbers and metadata for each study is also shown on Table S1. In the y-axis, “L_Clostridium” refers to Clostridium genus assigned to Lachnospiraceae family and “E_Clostridium” refers to Clostridium genus assigned to Erysipelotrichaceae family.
Figure 3

Analysis of shared taxons in the three different hypervariable regions considered in the study.

(A) Analysis at genus level, using a 80% threshold, which means that bacteria must be present in at least the 80% of the samples. (B) Genus level at 50% threshold. (C) Core microbiome per study using an 80% threshold. Number of Core genera per study are shown on the horizontal bar plot, while the number of genera shared in the core of one or more studies are indicated in the vertical bar plots, the shared studies for each specific columns are shown by the connected dots in the lower panel.
Figure 4

Evaluation of alpha diversity of the entire set of samples by using Shannon metric and Kruskal-Wallis statistics for significance.

Violin plots represent different groups for each category. (A) Study; (B) Hypervariable region amplified from the 16S rRNA gene. (C) Type of chickens (commercial, experimental). (D) Breed of the chickens, where Bh corresponds to Hubbard Hybrid breed and Broiler to all fattening breeds that are not specified. Boxplots represent the interquartile range and in red the median for each group. The number of samples in each group is shown right below the name. For Panel A, Indexes of metadata were used for visualization purposes and are related with the specific studies in Table S1. The colored violin plots correspond to Varmuzova study (2015), which is highlighted due to the inclusion of Salmonella enteriditis samples (green) that showed lower diversity than treated samples with plant extract (blue) and control samples (yellow) from the same study.
Figure 5

PCoA showing the spatial dispersion of the samples. Distance matrix was constructed based on Bray-Curtis metric.

Coordinate plots represent (A) Study; The study of Varmuzova et al., 2015 was split due to the high intra-study variation obtained. Varmuzova_2015 = Control Chickens and uninfected Chickens treated with the plant extracts; Varmuzova_2015_I = Infected chickens; Varmuzova_2015_P_I = Infected chickens, treated with the plant extracts. (B) Hypervariable region of the 16S rRNA gene. (C) Type of chickens (commercial, experimental), when commercial chickens refer to animals raised on normal production environments and experimental refers to animals that were kept into environments with controlled conditions. (D) Breed of the animals, where Broiler corresponds to all fattening breeds that are not specified.
