Article

A Probabilistic Structural Equation Model to Evaluate Links between Gut Microbiota and Body Weights of Chicken Fed or Not Fed Insect Larvae

Johann Detilleux *, Nassim Moula [ID], Edwin Dawans, Bernard Taminiau [ID], Georges Daube [ID] and Pascal Leroy

Fundamental and Applied Research for Animals, Veterinary Faculty, University of Liège, Quartier Vallée 2, Avenue de Cureghem 7A-7D, 4000 Liège, Belgium; nassim.moula@uliege.be (N.M.); edawans@uliege.be (E.D.); bernard.taminiau@uliege.be (B.T.); georges.daube@uliege.be (G.D.); pascal.leroy@uliege.be (P.L.)
* Correspondence: jdetilleux@uliege.be

Simple Summary: Feeding poultry with insects could reduce production costs, but the impact of this diet on their gut microbiota and growth is little known because the network of relationships between their weights, the composition of their microbiota and their diet is complex and potentially biased by confounding factors (such as the gut compartment, age and sex of the birds). In this study, we were able to unravel these relationships in local breed chickens fed or not fed with black soldier fly larvae thanks to a technique of artificial intelligence (the probabilistic structural equation model). Bacteria were grouped into few entities with distinctive metabolic attributes and were probably linked nutritionally. Birds' age influenced body weights and bacterial composition. The proposed methodology was thus able to simplify the complex dependencies among bacteria present in the gut and to highlight links potentially important in the response of chicken to insect feed.

Abstract: Feeding chicken with black soldier fly larvae (BSF) may influence their rates of growth via effects on the composition of their gut microbiota. To verify this hypothesis, we aim to evaluate a probabilistic structural equation model because it can unravel the complex web of relationships that exist between the bacteria involved in digestion and evaluate whether these influence bird growth. We followed 90 chickens fed diets supplemented with 0%, 5% or 10% BSF and measured the strength of the relationship between their weight and the relative abundance of bacteria (OTU) present in their cecum or cloaca at 16, 28, 39, 67 or 73 days of age, while adjusting for potential confounding effects of their age and sex. Results showed that OTUs (62 genera) could be combined into ten latent constructs with distinctive metabolic attributes. Links were discovered between these constructs that suggest nutritional relationships. Age directly influenced weights and microbiotal composition, and three constructs indirectly influenced weights via their dependencies on age. The proposed methodology was able to simplify dependencies among OTUs into knowledgeable constructs and to highlight links potentially important to understand the role of insect feed and of microbiota in chicken growth.

1. Introduction

The use of insects in poultry feed has generated great interest based on the recognition that they are powerful bio-converters with a low environmental footprint [1]. Insects are naturally consumed by poultry and, when supplemented in basic diets, may improve poultry growth performances [2]. These performances depend on the multiple interactions between feed components, host cells and gut microbiota: When gut microbiota is modified, digestion and nutrient absorption are affected. These, in turn, can influence feed conversion, leading to altered growth and health [3].

The microbiota in gut mucus and lumen samples plays a role in the metabolism of chickens [4]. Indeed, bacterial fermentation of foodstuff entering the caeca produces up to 10% of the metabolizable energy [5]. The caecum is also the gastrointestinal tract compartment where the number and variety of bacteria are the highest, with up to 10^{11} bacteria per gram [6]. The composition of the chicken caecal bacterial flora has been studied by numerous authors, and significant variations were found according to the age and genetics of the birds [7], their surrounding environment [8] and diet composition [9]. Regarding the impact of insects, researchers found differences in the microbiota of broilers fed basic feed supplemented with 0% to 15% of partially defatted black soldier fly (BSF) larvae [10], in the microbiota of hens fed meals in which soybean was completely or not replaced by defatted BSF larvae [11] and in the microbiota of broiler Ross 708 chickens fed meals supplemented with 0% to 15% of Tenebrio Molitor [12]. Inversely, no differences were found in the composition of the caecal microbiota of broilers fed a basic diet supplemented with 0% to 8% of defrosted BSF larvae [13].

The different nutrients in BSF larvae can influence the composition of the broiler gut microbiota because these microbes use feed-derived compounds as growth substrates [14]. Although the content of BSF larvae varies considerably depending on their rearing substrate, the developmental life stage at which they are harvested and the method used to kill them, it is generally rich in protein (dry matter > 30%) and fat (dry matter > 20%), with an abundance of lauric acid known for its antimicrobial activity against Gram-positive bacteria [15]. It was also reported that chitin, the main component of the exoskeleton, serves as a fermentation source for caecal microbiota [16].

The analysis of the gut microbiotal composition may also vary with the methods used for their quantification, whether it is in the sequencing [17] or in the statistical methods of analysis. Many classical statistical tests are available to analyze the gut microbiome [18] but it is important to recognize that the evaluation of processed metagenomic data have several characteristics that necessitate specific statistical tools. Typically, these data are normalized into non-negative relative abundances of operational taxonomic units (OTU) [19], and because they are fundamentally discrete, these can only be approximately modeled by continuous variables [20]. Relative abundances are by nature the estimates of the multinomial probabilities for the OTU counts. These have compositional characteristics [19,21] and must be log-transformed to resolve the constant sum constraint [22]. Moreover, distributions of relative abundances are often highly skewed (over-dispersed) with a lot of zero values [23]. Datasets are usually very large, with the number of sampled animals being much lower than the number of bacterial species ("big-p, little-n" problem). In such a case, the probability of finding an OTU that correlates with the variable of interest is high, even if no real correlation exists in the domain. Another issue comes from the interaction within and between microbial communities [24] and the potential dependencies between OTU relative abundances (multicollinearity), which prevents the use of traditional regression methods to test whether an association exists between relative abundances and a variable of interest [25].

To address these issues, networks have been proposed to describe the relationships between OTU relative abundances [26,27], including Bayesian networks [28]. Bayesian networks are represented by directed acyclic graphs in which variables are symbolized by connected nodes: If a node depends on another node then a directed edge is drawn between them (direct probabilistic dependencies). A node is interpreted as the child node if it has the tip of the arrow attached to it, and as the parent node if it has the base of the arrow attached to it. Each node is associated with a probability distribution that indicates how the probability that a child node will take a certain value depends on the values taken by its parent node. Different algorithms exist to construct the structure of a network (i.e., the description of the dependence/independence relationships among nodes) from the data [29]. One method is to select the network that has the best fit to the data. This fit can be measured by various parameters, including the minimum description length score function (MDL) as a stopping rule. The MDL function is characterized by a preference for simpler

Bayesian networks than complex ones: a network with a lower MDL score indicates that it optimizes the tradeoff between complexity and fit more effectively than one with a higher score [30]. Once the structure of the network is known, conditional probability tables can be estimated by maximizing their likelihoods [31]. The MDL function can also be used to estimate the number of classes when discretizing OTU relative abundances with minimal information loss and maximum accuracy in representing the abundances, as it is required in most methods available to learn networks from data (e.g., BayesiaLab, Hugin Expert, Netica, R).

Besides Bayesian networks, structural equation models (SEMs) are another type of graphical model, which are useful to analyze the potential relationships between OTU relative abundances and a phenotype of interest [32,33]. These models allow for grouping relative abundances into a limited number of latent constructs (LC), which represent the underlying unobservable causes of variations in the observed abundances. Commonly, SEMs are composed of measurement and structural models. The measurement model describes the relationships between observed data and LCs while the structural model imputes the relationships between LCs and a variable of interest [34,35]. In the measurement part of the SEM, relationships are often based on known or expected underlying mechanisms. However, in probabilistic SEM (pSEM), they are inferred assuming no theory but based on the structure of a network [36]. In the structural part of the SEM, it is also possible to break down the relationship between a LC and a variable of interest into its direct and indirect components. In the terminology of causal mediation analysis, a "direct effect" is an effect that is not mediated by other variables and an "indirect effect" is the portion of the effect that can be explained by mediation alone. Their sum is sometimes called the 'total effect' [37,38].

The goals of this study are, using the pSEM methodology: (1) to summarize into a few LCs all relationships observed between relative abundances of OTUs found in the caeca and cloacal swabs of chickens fed or not fed BSF larvae, and (2) to estimate the effects (adjusted for potential confounders) of the presence of these larvae on the weights measured on the day of OTU collection, with effects mediated or not by changes in the microbiota in the LCs.

2. Materials and Methods

2.1. Bird and Insect Management

Management of insect and birds was kept intentionally simple, in semi-artificial conditions. The BSF larvae were raised on beet pulp, which was watered manually so the substrate remained more or less "sludgy", and under ambient temperature maintained between 20 and 30 °C. The larvae were collected by hand before the pupal stage and kept at −20 °C directly after collection and until bird feeding time. A portion of the larvae was allowed to complete development to adulthood. Adult flies were transferred into cages with natural and artificial lighting for mating and egg-laying [39].

The experiment consisted in raising thirty one-day-old chicks of the local breed Ardennaise indoors at the Veterinary Faculty of Liege, all together for the first days of life to ensure microbiota exchange through typical bird behavior [40]. At the end of this "pre-experimental" period, we collected fecal samples via cloacal swabs and randomly allocated birds to three groups with ten birds per group and allowed them to adjust over two days. During this "experimental period", birds had ad libitum access to water and to commercial feed supplemented with either 0%, 5% or 10% BSF larvae. Supplementary Table S1 provides information about the ingredients and chemical composition of the diets, which were formulated to be of equal energy and protein concentrations. Birds were weighed on a weekly basis until the end of the experimental period, when fecal samples were collected from the caecum after birds were euthanized by cervical dislocation. The physical environment was identical for all three groups and the whole experiment was replicated three times. The sampling frame and age at slaughter were a little different across replicates because of external conditions related to the sanitary containment of COVID-19: In the first replicate,

we collected 5 cloacal (28 days of age) and 5 caecal (67 days of age) contents per group. In the second replicate, we collected 4 cloacal (16 days of age) and 8 caecal (73 days of age) contents per group. In the last replicate, we collected 8 caecal contents per group at 39 days of age. Contents of the caecum and cloacae were placed individually in sterile tubes with phosphate-buffered saline and stored immediately after collection at −80 °C until DNA extraction.

2.2. 16S rDNA Detection and Analyses

Total bacterial DNA was extracted from the samples with the PSP Spin Stool DNA Plus Kit 00310 (Invitek, Berlin, Germany), following the manufacturer's recommendations. The PCR amplification of the 16S rDNA V1–V3 hypervariable region and library preparation was performed with the following primers (with Illumina overhand adapters): forward (50-GAGAGTTTGATYMTGGCTCAG-30) and reverse (50-ACCGCGGCTGCTGGCAC-30). Each PCR product was purified with the Agencourt AMPure XP bead kit (Beckman Coulter, Pasadena, CA, USA) and submitted to a second PCR round for indexing, using Nextera XT index primers 1 and 2. After purification, PCR products were quantified using the Quant-IT PicoGreen (ThermoFisher Scientific; Waltham, MA, USA) and diluted to 10 ng/µL. A final quantification of each library was performed using the KAPA SYBR® FAST qPCR Kit (KapaBiosystems; Wilmington, MA, USA) before normalization, pooling and sequencing on a MiSeq sequencer using V3 reagents (Illumina; San Diego, CA, USA).

Sequence read processing was performed as previously described using the MOTHUR software package v1.39.5 [41] and the VSEARCH algorithm for chimera detection [42]. A clustering distance of 0.03 was used for OTU generation. The 16S reference alignment and taxonomical assignment from phylum to genus were performed with MOTHUR and were based upon the SILVA database (v1.32) of full-length 16S rDNA sequences [43]. Subsample datasets with 10,000 cleaned reads per sample were obtained and used to evaluate all OTUs' coverage (Good's coverage), which was adequate.

2.3. Data Processing and Modeling

For each individual, relative abundances were computed as the number of OTUs of one genus divided by the total number of OTUs. As an editing step, the dataset was filtered and only relative abundances higher than 0.01% in at least one replicate were kept in the analyses. This was carried out to eliminate counts that may represent a sequencing artefact and to minimize the number of tests that are unlikely to result in significant findings. Relative abundances greater than null were log-transformed because of the sum constraint [22].

To create the pSEM and identify genera associated or not associated with the weights measured on the day of OTU collection, we used the BayesiaLab (v4.6) software package [44]. Following their recommendations, log-transformed relative abundances were discretized with the R2-GenOpt* algorithm and null relative abundances were grouped into one specific category. Hereafter, these discretized values are called "RA". To discretize weights, a K-means classification algorithm was chosen after visually inspecting their distribution and they were grouped into five distinct and non-overlapping classes.

Modeling started with the creation of several network structures linking all RAs. To do so, we used all unsupervised learning algorithms proposed in BayesiaLab to increase the probability of finding a solution close to the global optimum. We evaluated the complexity of the networks with respect to the fit to the data via the structural coefficient (SC) analysis of the MDL score [45]. The strength of each relationship in the networks was measured by computing the Kullback–Leibler (KL) divergence between variables included in the relationship [46]. The best network structure was selected as the one with the lowest MDL score and for which all KL values were significantly different from null (G test; $p < 0.05$).

Once this best network was found, we created the measurement part of the pSEM by clustering RAs in accordance with the structure of the network and allowing a maximum number of ten RAs per cluster. The robustness of the resulting clusters was evaluated

on ten subsets with the jackknife cross-validation method [47], and the percentages of correspondence of each cluster with respect to the one obtained in the initial network were computed. The quality of the representation of each cluster was measured by its purity (i.e., mean of the RA assignment probabilities in the cluster) and the contingency table fit (which compares the entropy of the current naive Bayes structure to the entropy of a fully connected structure). The LCs were then assigned to each cluster using naive Bayes structures. The value assigned to a LC is the weighted average of all RAs included in the cluster, and the weights are the ratio of the mutual information of the RA with respect to its LC divided by the highest mutual information in the LC.

Finally, the strength of the relationships between the birds' weight categories, birds' age at sampling (16, 28, 39, 67 or 73 days), birds' sex (male or female), origin of the OTU sample (cloaca or caecum), amount of dietary BSF (0%, 5% or 10%) and LCs was assessed in the structural part of the pSEM. The algorithm recommended by the BayesiaLab's team to construct this part of the pSEM is called the "Structural Priors Learning algorithm", which is based on a heuristic search of the best network via a bootstrap aggregating approach [44]. Here, again we examined SC and KL values to choose the final model and evaluated its accuracy in estimating weights with the jackknife cross-validation procedure. The results were expressed as the area under the ROC curves, which are plots of the true against the false positive rates in classifying estimated weights. After all validations, total and direct effects of each LC on the weight categories were computed as the change in weights per unit change in the mean of the latent construct before (total effect) and after (direct effect) controlling for the other nodes.

3. Results

3.1. Descriptive Analytics

Weight means and standard errors are presented in Figure 1: weights were higher in males than females and increased with the age of the birds. After editing, a total of 62 genera were available to create the network and 47 of them were from the phylum *Firmicutes*. Details on relative abundances per genus can be found in Supplementary Table S2 for each diet (0%, 5%, 10% of insect incorporation), age (16, 28, 39, 67 and 73 days) and sex (male or female) of the birds, and origin of the fecal samples (cloaca or caecum). Mean relative abundances were highest for the *Lactobacillus* genus, with an average of 25.24% and 66.85% in samples taken from the cloaca and caecum, respectively. Besides *Lactobacillus*, fecal samples of the caecum contained on average 11.07%, 8.07%, 6.22% and 6.08% of bacteria of the genera *Lachnospiraceae*, *Ruminococcaceae*, *Alistipes* and *Faecalibacterium*, respectively. In the fecal samples from the cloacal swabs, the mean relative abundance for the genus *Candidatus Arthromicus* was 8.24%. The mean relative abundances of all other bacteria were lower than 5%.
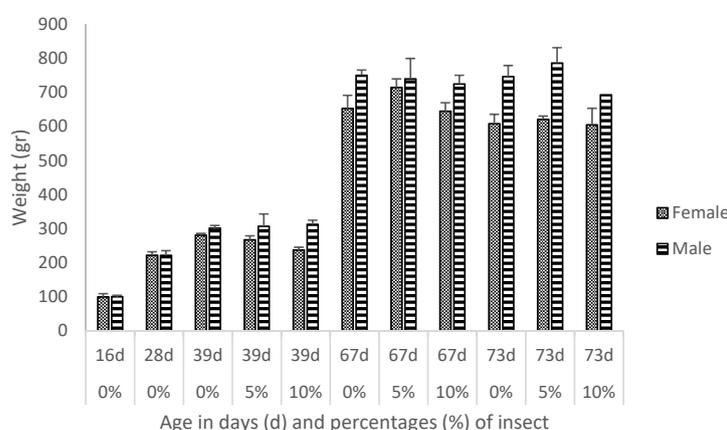
Figure 1. Weight means and standard errors (g) per sex, age of the birds in days (d) and percentages of *Hermitia Illucens* (%) in the diets.

3.2. Discretization

Discretization of weights with the K-Means algorithm led to 5 classes: 69.35–169.36 g (14.19%), 169.36–271.03 g (27.10%), 271.03–469.8 g (17.42%), 469.8–698.8 g (24.95%) and 698.8–873.0 g (16.34%). Concerning the RAs, the discretization obtained with the R2-GenOpt* algorithm are provided in Supplementary Table S3; in most cases, three RA categories were enough to describe these distributions.

3.3. Network Construction

The analysis of the SC curves revealed that SCs between 0.7 and 1 were adequate. All algorithms available in BayesiaLab for the unsupervised learning phase in the construction of the network provided similar MDL values, with the lowest one provided by the EQ algorithm with post-processed Taboo. The resulting network (Figure 2) contained 59 arcs and 62 nodes. Corresponding values for the KL and Pearson's correlation coefficients are provided in Table 1. All KL values were significantly different from null and most correlations were positive, with the exception of the correlation coefficients between *Lactobacillus* and those of the genera *Alistipes*, *Parabacteroides* and *Ruminococcaceae NK4A214*, between *Ruminococcaceae UCG-010* and *Alcaligenes* and between *Candidatus Saccharimonas* and *Eschrichia-Shigella*, which were all close to −0.30. Except for two nodes that were too weakly related to the others (i.e., *Clostridium sensu stricto 1* and 7), most nodes had one incoming and one outcoming arrow. A few 'dominant' nodes had more outcoming than incoming arrows: *Lachnospiraceae* (0 in, 6 out), *Ruminiclostridium_5* (1 in, 3 out), *Lactobacillus* (1 in, 5 out), *Ruminococcaccee NKA214* (1 in, 4 out), *Clostridiales vadin BB160* (1 in, 4 out), *Defluvitaleaceae_UCG-011* (1 in, 3 out), *Rombutsoa* (1 in, 3 out), *Tyzzerella* (1 in, 2 out) and *Ruminococcacceae_UCG-010* (1 in, 2 out).
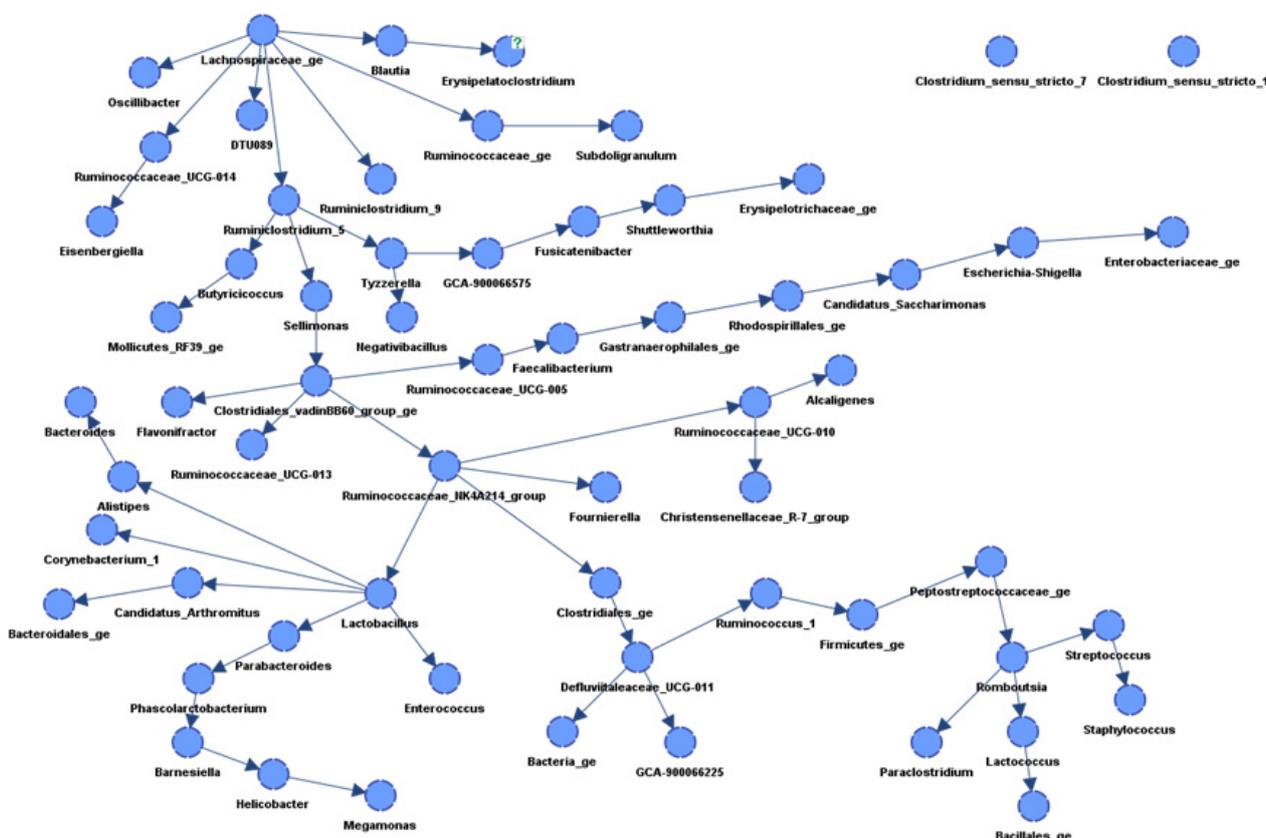

Figure 2. Network showing the arcs linking the bacterial relative abundances.

Table 1. Relationship analysis of the network depicted in Figure 1: Kullback–Leibler divergence (KL) and Pearson's correlation (Corr) between parent and child nodes.

Parent Node	Child Node	KL	Corr
Ruminococcaceae NK4A214 group	*Ruminococcaceae UCG-010*	0.77	0.83
Lachnospiraceae	*Ruminococcaceae*	0.76	0.94
Ruminiclostridium 5	*Butyricicoccus*	0.76	0.88
Tyzzerella	*Negativibacillus*	0.75	0.86
Ruminococcaceae UCG-010	*Christensenellaceae R-7 group*	0.73	0.81
Fusicatenibacter	*Shuttleworthia*	0.73	0.71
Tyzzerella	*GCA-900066575*	0.72	0.85
Clostridiales vadinBB60 group	*Flavonifractor*	0.72	0.76
Clostridiales vadinBB60 group	*Ruminococcaceae NK4A214 group*	0.72	0.79
Ruminiclostridium 5	*Sellimonas*	0.69	0.83
Ruminococcaceae UCG-014	*Eisenbergiella*	0.69	0.84
GCA-900066575	*Fusicatenibacter*	0.68	0.83
Ruminiclostridium 5	*Tyzzerella*	0.68	0.82
Lachnospiraceae	*Ruminococcaceae UCG-014*	0.67	0.88
Lachnospiraceae	*Ruminiclostridium 5*	0.67	0.86
Lachnospiraceae	*Ruminiclostridium 9*	0.66	0.86
Ruminococcaceae NK4A214 group	*Clostridiales*	0.65	0.77
Lachnospiraceae	*Oscillibacter*	0.65	0.84
Clostridiales	*Defluviitaleaceae UCG-011*	0.65	0.71
Escherichia-Shigella	*Enterobacteriaceae*	0.63	0.73
Sellimonas	*Clostridiales vadinBB60 group*	0.62	0.75
Ruminococcaceae	*Subdoligranulum*	0.61	0.85
Alistipes	*Bacteroides*	0.59	0.69
Clostridiales vadinBB60 group	*Ruminococcaceae UCG-013*	0.58	0.70
Ruminococcaceae NK4A214 group	*Lactobacillus*	0.57	−0.81
Ruminococcaceae UCG-005	*Faecalibacterium*	0.57	0.78
Lachnospiraceae	*DTU089*	0.56	0.78
Rhodospirillales	*Candidatus Saccharimonas*	0.54	0.58
Defluviitaleaceae UCG-011	*GCA-900066225*	0.54	0.70
Lactobacillus	*Alistipes*	0.54	−0.74
Shuttleworthia	*Erysipelotrichaceae*	0.53	0.64
Clostridiales vadinBB60 group	*Ruminococcaceae UCG-005*	0.51	0.72
Lachnospiraceae	*Blautia*	0.48	0.57
Barnesiella	*Helicobacter*	0.48	0.63
Ruminococcus 1	*Firmicutes*	0.48	0.26
Firmicutes	*Peptostreptococcaceae*	0.46	0.36
Lactobacillus	*Candidatus Arthromitus*	0.45	0.75
Faecalibacterium	*Gastranaerophilales*	0.45	0.62
Defluviitaleaceae UCG-011	*Ruminococcus 1*	0.45	0.69
Butyricicoccus	*Mollicutes RF39*	0.44	0.66
Ruminococcaceae NK4A214 group	*Fournierella*	0.43	0.51
Defluviitaleaceae UCG-011	*Bacteria*	0.40	0.53
Romboutsia	*Paraclostridium*	0.39	0.53
Peptostreptococcaceae	*Romboutsia*	0.39	0.73
Blautia	*Erysipelatoclostridium*	0.38	0.59
Helicobacter	*Megamonas*	0.36	0.68
Romboutsia	*Lactococcus*	0.36	0.64
Lactobacillus	*Corynebacterium 1*	0.36	0.28
Parabacteroides	*Phascolarctobacterium*	0.35	0.64
Lactobacillus	*Enterococcus*	0.33	0.39
Lactococcus	*Bacillales*	0.33	0.63
Gastranaerophilales	*Rhodospirillales*	0.32	0.52
Ruminococcaceae UCG-010	*Alcaligenes*	0.30	−0.34
Streptococcus	*Staphylococcus*	0.30	0.62

Table 1. *Cont.*

Parent Node	Child Node	KL	Corr
Romboutsia	*Streptococcus*	0.30	0.42
Phascolarctobacterium	*Barnesiella*	0.27	−0.02
Candidatus Saccharimonas	*Escherichia-Shigella*	0.27	−0.30
Lactobacillus	*Parabacteroides*	0.23	−0.36
Candidatus Arthromitus	*Bacteroidales*	0.14	−0.24

3.4. Measurement Part of the pSEM

The RAs were clustered into 11 groups in accordance with the structure of the network shown in Figure 2. Purity and contingency table fit values of the clusters are presented in Table 2. All purity values were higher than 95% (overall mean = 98.27%) and all contingency table fit values (overall mean = 83.97%) were above the recommended threshold of 70%. Cross-validation with the jackknife method showed that nodes are often clustered into the same clusters. The average fit score was 65.40% and ranged from 53.55% to 76.32%.

Table 2. Mean purity and contingency table fit (CTF) of each cluster.

Cluster	Purity (%)	CTF (%)
0	99.65	79.32
1	98.91	61.79
2	98.46	80.41
3	98.23	88.48
4	97.98	76.14
5	99.57	80.59
6	98.98	87.11
7	98.66	87.84
8	96.82	87.85
9	97.28	96.10
10	96.43	98.02

The characteristics of the 11 LCs associated with these clusters are provided in Supplementary Table S4. High LC values usually corresponded to the high relative abundances of their components, with the exception of LC0, for which high values corresponded to high values of *Lactobacillus* and *Candidatus Arthromitus* but low values of the other components.

3.5. Structural Part of the pSEM

The structural part of the pSEM is represented in Figure 3 and corresponding metrics are provided in Table 3 for SC = 0.45, which was the network with the lowest structure/data ratio. All KL values were significantly different from null, with the exception of the arc between sex and weight. Results of the ten-fold cross-validation procedure indicated that areas under the ROC curves varied from 92.87% to 96.86% (mean = 95.86%). The structure was characterized by three notable features: the node "weight" is the child node of the nodes "age" and "sex", the node "age" is the parent of the nodes LC0, LC1, LC4, LC5 and LC6 and the node "BSF" is not linked to the node "weight". This structure was found in all networks constructed with SC values ranging from 0.45 to 1. It was also found in all of these networks that nodes LC2, LC3, LC7 and LC9 are child nodes of LC6.
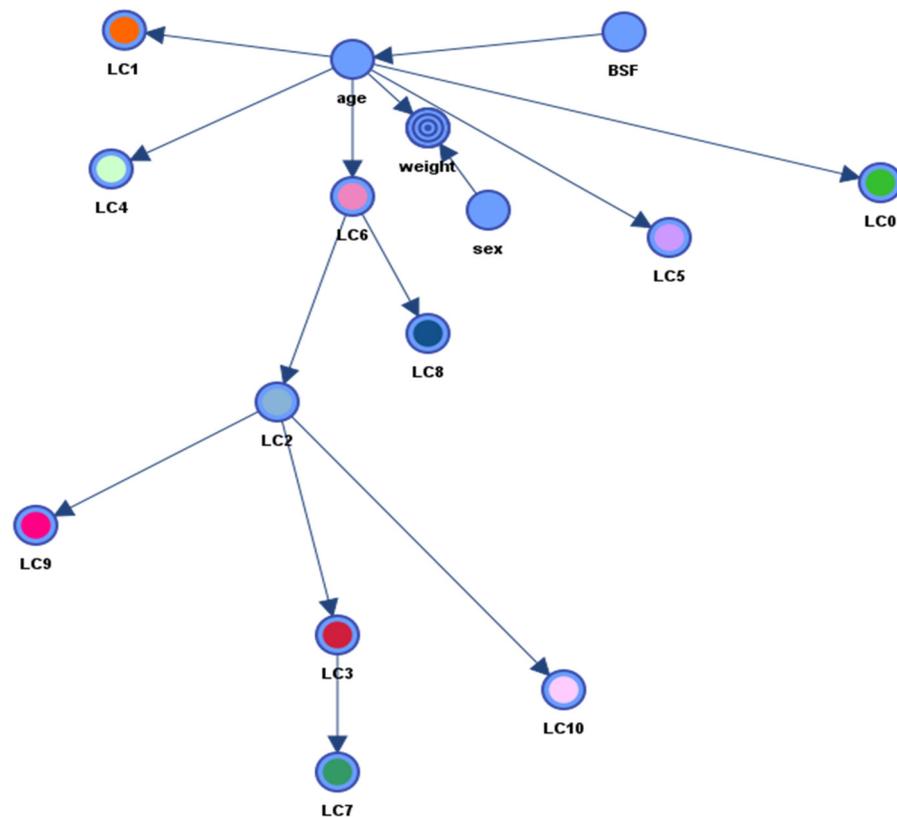
Figure 3. Network showing arcs between weight, latent constructs (LC0 to LC10), age, sex and BSF nodes.

Table 3. Relationship analysis of the network depicted in Figure 2: Kullback–Leibler (KL) divergence and Pearson's correlation between parent and child nodes.

Parent Node	Child Node	KL Divergence	Correlation
age	weight	1.51	0.94
age	LC1	1.34	−0.24
LC6	LC2	0.95	0.62
age	LC4	0.92	0.53
LC6	LC8	0.92	0.56
LC2	LC10	0.87	0.84
LC2	LC3	0.82	0.86
age	LC6	0.78	0.37
LC3	LC7	0.78	0.79
age	LC0	0.68	−0.70
age	LC5	0.68	0.53
LC2	LC9	0.41	0.68
BSF	age	0.34	0.44
sex	weight	0.18	0.07

Total effects on weights were significantly different from null for LC1, LC4, LC5, LC6, LC8, age and sex: weight increased by 10.59 g per day of age and male birds were heavier than female birds. Among LCs, the strongest total effect was for LC5, with a value of 285.66 (standardized value = 0.57). Total effects of LC1, LC4, LC6 and LC8 were −116.38, 108.96, 492.45 and 380.21, respectively. The other effects were not significantly different from null. Concerning the direct effects on weight, only the effects of age and sex were significantly different from null, with values of 10.52 and 34.50 g for age and sex, respectively.

4. Discussion

In this study, it was hypothesized that feeding chicken with different BSF amounts influences their rates of growth via an effect on their gut microbiota. To test this hypothesis, we evaluated the strength of the interrelationships between weight, diet and OTU abundances with a pSEM because this type of model combines the advantages of traditional SEMs with those of Bayesian networks. Age and sex of the birds, as well as the gut compartment where the microbiota was characterized, were also included in the pSEM because these variables may alter the estimated values of the strength of association between weights and microbial abundances.

4.1. Descriptive Analytics

Although bacterial relative abundances are known to vary across studies [48], *Firmicutes* was found to be the most abundant phylum and *Lactobacillus* the most abundant genus, as was reported in a recent meta-analysis [49], although in different percentages. The composition of fecal microbiota samples obtained via cloacal swabs or directly from the caecum were different, as observed by others [50–52]. Also observed by others, many species could not be classified, encouraging further studies to improve techniques for sequencing and annotating genomes in chicken microbiota [53].

Ardennaise weights (Figure 1) were similar to those found in previous studies: means at 11 weeks of age were 924.70 and 766.51 g for males and females, respectively [54], and 656 g at 8 weeks of age [55].

4.2. Measurement Part of the pSEM

Gut bacteria co-exist in complex networks, so the first step in testing our hypothesis was to explore OTU-to-OTU associations and to summarize them into a few representative LCs so the effect of diets on them could be quantified. Results of the pSEM analysis showed that RAs could indeed be combined in distinct unobservable LCs (Supplementary Table S4). The composition of these LCs can be compared to the four enterotypes defined by Kaakoush et al. [56]. In their study, they used principal component analysis to aggregate the relative abundances found in 56-day-old chicken feces and discovered 4 enterotypes whose compositions may be related to that of the LCs in this study. For example, enterotype 1 was dominated by *Lactobacillus* (LC0) and *Peptostreptococcaceae* (LC1). Besides genera in enterotype 1, enterotype 2 was dominated by *Streptococcus* (LC1) and all Proteobacteria of LC4 (*Escherichia*, *Shigella* and *Enterobacter*), enterotype 3 was dominated by the only Actinobacteria found in this study (*Corynebacterium 1*) and enterotype 4 was dominated by *Ruminococcaceae* (mainly in LC2, LC3, LC6 and LC7) and Bacteroidetes (*Alistipes* and *Bacteroides*, both in LC0). Conceivably, several of the LCs could have been merged together to better match the composition of these enterotypes, but purity values (>95%) and contingency tables' fits (>70%) indicated a good representation of the RA joint posterior densities, which is not in favor of further clustering in this study. In another study (layer hens of various ages), Xiao et al. [57] proposed a network of 30 genera connected by 478 links corresponding to absolute Spearman's rank correlation values between OTU relative abundances higher than 0.5, but available information was not sufficient to support or contradict our findings. The clustering of *Lactobacillus* species with members of the *Bacteroides*, *Ruminococcaceae* and *Bacteroidales* genera in LC0 was similarly observed by Zou et al. [58], and this is another argument in favor of the structure of our network.

Metabolic dependencies among bacteria may partly explain the composition of these LCs [59]. Indeed, cecal microbiota plays important roles in the metabolic pathways leading to the production of short-chain fatty acids [60,61] and compounds from the fermentation of ileal bypass proteins [62]. For example, RAs of *Lactobacilli* (LC0) were inversely associated with RAs of many other genera (Table 1), and this may be substantiated by the ability of *Lactobacilli* to produce bacteriocin-like substances potentially toxic for other bacteria and to convert carbohydrates into lactic acid, which decreases the pH, which usually restricts bacterial growth [48]. Similarly, bacteria belonging to four out of the six genera of LC3

(i.e., *Oscillibacter*, *Lachnospiraceae Ruminococcacea* and *Subdoligranulum*, all with high loadings) were shown to be involved in butyrate synthesis [63–65]. Bacteria belonging to four out of the five genera of LC5 (i.e., *Megamonas*, *Phascolarctobacterium*, *Barnesiella* and *Parabacteroides*) expressed enzymes for propionate production pathways [64]. Members of *Alistipes* and *Bacteroides*, both with high loadings in LC0, were shown to convert carbohydrates, including cellulose, through the succinate pathway [66,67]. Relative abundances of *Eisenbergiella* and *Ruminococcaceae UCG-014*, the only components of LC10, were positively correlated with the amount of butyric acid in the caecum of chicks infected with the porcine delta-coronavirus [68]. In humans, *Shuttleworthia*, with the highest loading in LC7, mainly produce butyrate via fermentation of non-digestible carbohydrates [69] and were associated with male broilers of high body weights [70].

Besides carbohydrates, bacteria may also use proteins as fermentable substrates, as do bacteria of genera *Peptostreptococcaceae*, *Romboutsia* and *Paraclostridium* [71,72] with the highest abundances among the OTUs of LC1. Bacteria also play a role in chicken immunity. For example, LC2 is dominated by bacteria of *Clostridiales vadinBB60* and *Mollicutes RF39* genera. Both were positively correlated with levels of regulatory T cells and CD11C$^+$CD103$^+$ dendritic cells in mesenteric lymph nodes of mice sensitized with ovalbumin [70]. Similarly, *Candidatus Arthromitus* and *Bacteroides* (both in LC0) have been shown to activate the innate and adaptive immunity [56,73,74] and *Eisenbergiella* (LC10) was negatively correlated with TNF-α and IFN-γ expression levels in chicks infected with the porcine delta-coronavirus [68].

4.3. Structural Part of the pSEM

The second step to test our hypothesis was to explore links across the LCs and diet, while controlling for variables such as sex, age and gut compartment, which may also influence weights. We observed that constructs LC2, LC3, LC7, LC9 and LC10 were child nodes of LC6 in all networks created with different SC values. Nutritional interactions may explain these relationships because metabolic cross-feeding is frequent in these microbial communities, as was reviewed by Flint et al. [75]. One may then imagine that bacteria of LC6 convert substrates into products subsequently used by bacteria of LC2, LC3, LC7, LC9 and LC10, but further analyses are necessary to validate or reject this hypothesis.

Age was shown to influence body weights and microbiotal composition in this (Figure 3 and Table 3) and many other studies [7,8,76,77]. The KL values were highest between the "age" node and the nodes of "weight", LC1 and LC4 (Table 3). In support of this observation, the genus *Faecalibacterium*, the most abundant genus of LC4 (with relative abundances going from 0.06% at 16 days of age to 6.95% at 73 days of age) was previously associated with mature microbiota in humans and chicken [78]. Richards et al. [4] also observed that the genus *Faecalibacterium* colonizes the cecum later in life of broilers of three different breeds, as well as other members of the *Ruminococcaceae*, *Firmicutes* and *Mollicutes RF39*, but this last observation was not supported in our study. A link also connected the nodes "age" and "BSF", which can be explained by the design of the study. Indeed, birds at ages 16 and 28 days were sampled before BSF larvae were included in their diets, while older birds were fed on a diet supplemented or not with BSF larvae.

No link was found between the node "BSF" and any of the LCs. Similarly, little or no differences were found in the composition of cecal microbiota of Ross broilers fed a commercial feed with either 0% or 8% of fresh BSF larvae [13], or up to 20% [79], while others observed differences in the relative abundances of several genera in Ross broilers fed meals including 0% to 15% of partially defatted BSF larvae [10] or meal in which soybean oil was partially substituted by natural or modified BSF larvae fat [80]. It is possible that the substances in BSF fat explain these findings. Indeed, fat of BSF larvae is rich in medium-chain fatty acids known to influence microbiota and intestinal villi morphology [15,80,81].

Likewise, no link was created between the node "BSF" and the node "weight". This suggests that BSF larvae incorporation in the diets of broilers would not affect the weight of birds of a similar age. This was found by many, but not all studies, as reviewed by Abd El-

Hack [82]. Discrepancies were also observed in meta-analyses of the growth of chicken [2] and fish [82] fed or not fed BSF larvae. No main effect was observed in both meta-analyses, but replaced protein sources and the BSF inclusion level were found responsible for the discrepancies between the studies included in the meta-analyses. Rearing substrates, environmental conditions and harvesting and processing methods of BSF larvae were also shown to explain the differences between studies [83,84]. In this study, soybean meal was replaced by up to 5% of fresh BSF larvae and diets were formulated to have equal energy and protein concentrations. The low incorporation rates of the BSF larvae [2] and the similarity of the amino acid profiles of these larvae and soybean meal [85] could explain the small difference observed between the weight means of birds fed meal supplemented or not with BSF larvae. On the other hand, the growth rate was shown to be reduced when broilers were fed higher levels of BSF larvae than those used in this study [2,16,86].

Finally, we observed that the direct effects on weights of all LCs were null, while the total effects of LC1, LC4 and LC6 were different from null—negative for LC1 and positive for LC4 and LC6. Therefore, because the node "age" was intermediate within the pathway linking the nodes for these LCs and the node "weight" (Figure 3), it can be deduced that LC1 indirectly decreased and LC4 and LC6 indirectly increased body weights via their effect on age. These findings point to the importance of mediation analysis when exploring the mechanisms underlying the relationships between microbiota, age and weight.

4.4. Limitations

The proposed methodology obviously has several limitations, which should be taken into account in the interpretation of our data. For example, factors such as the genetic background, immune system activity and behavior of the birds were not considered, although they are known to influence microbiotal composition [87]. Feed intake was not measured and may have affected growth performances. However, results reported in the literature are conflicting [16]. For example, feed consumption was not statistically different between Ardennaise chickens fed a commercial diet, of which 8% was replaced by fresh BSF larvae, and chickens fed the control diet [88]. The environment was kept as constant as possible across replications, but it is still possible that the effect of age on some LCs is rather due to environmental conditions than age by itself, especially since the effects of age, location and replication are partially confounded in this study. It may also be argued that the sample size was small and limits the statistical power of the network. On the other hand, the MDL principle that we used is known to be particularly relevant for small samples [89], although it is possible that more significant differences would be observed with a larger sample size.

Discretization of the observed relative abundances into RAs yielded only a few classes, which may be interpreted as a loss of information, but Lehmann and Hurlbert [90] observed that two and three ordered categories are often appropriate when discretizing continuous data. Additionally, discretization allows evaluating non-linear relationships [91]. This is particularly relevant in this study as quite a few relationships were non-linear, as suggested by the differences between KL divergence and Pearson's correlation coefficients (Tables 1 and 3). Lastly, results of the cross-validation procedures provided high values for the areas under the ROC curves, higher than the threshold of 80%, suggesting good classification performance [92].

5. Conclusions

In this study, we have developed a new method to assess the direct and indirect roles of gut bacterial microbiota in the growth of chickens fed or not fed BSF larvae. In this first application, it was shown that the pSEM reduced the complex web of relationships among the identified bacterial genera into a few unobservable latent constructs, which showed distinctive metabolic attributes and nutritional dependencies. Direct links were found between two of these constructs, one containing the genus *Faecalibacterium*, with the birds' age and weight. No direct link was found between the level of BSF supplementation

and growth. Application of this method to microbiota data from other sources may provide valuable insight into how it may influence chickens' growth.

These results collectively provide an improved understanding, although further studies are necessary to validate causal associations.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/biology11030357/s1, Table S1: Ingredients (%) and chemical (g/kg DM) of the control experimental diets. Table S2: Mean relative abundances (%) of each genus per sex (Female or Male) and age (in days) of the birds, per origin of the sample (Caecum or Cloaca), and per percent of *Hermetia Illucens* in the diet (0%, 5% or 10%). Table S3: For each genus, mean (standard deviation) of OTU relative abundances, limits of the classes after discretization with the R2-GenOpt* algorithm and percentages of each class (%). Table S4: Mutual information (MI) and mean relative abundances of the genera in the lowest, middle and highest categories of their latent construct (LC) and the corresponding LC means (in italics). The LC9 has two categories.

Institutional Review Board Statement: All procedures and experiments complied with the guidelines and were approved by the Local Ethics Commission of the University of Liège (Dossier 14/1666—Commission Ethique Animale ULg, Liège, Belgium), with respect to animal experimentation and care of animals under study, and all efforts were made to minimize suffering.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available within the article and the supplementary materials.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. van Huis, A.; Oonincx, D.G.A.B. The environmental sustainability of insects as food and feed. *A review. Agron. Sustain. Dev.* **2017**, *37*, 43. [CrossRef]
2. Moula, N.; Detilleux, J. A meta-analysis of the effects of insects in feed on poultry growth performances. *Animals* **2019**, *9*, 201. [CrossRef] [PubMed]
3. Diaz Carrasco, J.M.; Casanova, N.A.; Fernández Miyakawa, M.E. Microbiota, gut health and chicken productivity: What is the connection? *Microorganisms* **2019**, *7*, 374. [CrossRef] [PubMed]
4. Richards, P.; Fothergill, J.; Bernardeau, M.; Wigley, P. Development of the caecal microbiota in three broiler breeds. *Front. Vet. Sci.* **2019**, *6*, 201. [CrossRef]
5. Jozefiak, D.; Rutkowski, A.; Martin, S.A. Carbohydrate fermentation in the avian ceca: A review. *Anim. Feed Sci. Technol.* **2004**, *133*, 1–15. [CrossRef]
6. Yeoman, C.J.; Chia, N.; Jeraldo, P.; Sipos, M.; Goldenfeld, N.D.; White, B.A. The microbiome of the chicken gastrointestinal tract. *Anim. Health Res. Rev.* **2012**, *13*, 89–99. [CrossRef]
7. Ocejo, M.; Oporto, B.; Hurtado, A. 16S rRNA amplicon sequencing characterization of caecal microbiome composition of broilers and free-range slow-growing chickens throughout their productive lifespan. *Sci. Rep.* **2019**, *9*, 2506. [CrossRef]
8. Kers, J.G.; Velkers, F.C.; Fischer, E.A.J.; Hermes, G.D.A.; Stegeman, J.A.; Smidt, H. Host and environmental factors affecting the intestinal microbiota in chickens. *Front. Microbiol.* **2018**, *9*, 235. [CrossRef]
9. Zhang, Y.; Liu, Y.; Li, J.; Xing, T.; Jiang, Y.; Zhang, L.; Gao, F. Dietary resistant starch modifies the composition and function of caecal microbiota of broilers. *J. Sci. Food Agric.* **2020**, *100*, 1274–1284. [CrossRef]

10. Biasato, I.; Ferrocino, I.; Dabbou, S.; Evangelista, R.; Gai, F.; Gasco, L.; Cocolin, L.; Capucchio, M.T.; Schiavone, A. Black soldier fly and gut health in broiler chickens: Insights into the relationship between cecal microbiota and intestinal mucin composition. *J. Anim. Sci. Biotechnol.* **2020**, *11*, 11. [CrossRef]

11. Borrelli, L.; Coretti, L.; Dipineto, L.; Bovera, F.; Menna, F.; Chiariotti, L.; Nizza, A.; Lembo, F.; Fioretti, A. Insect-based diet, a promising nutritional source, modulates gut microbiota composition and SCFAs production in laying hens. *Sci. Rep.* **2017**, *7*, 16269. [CrossRef] [PubMed]

12. Biasato, I.; Ferrocino, I.; Grego, E.; Dabbou, S.; Gai, F.; Gasco, L.; Cocolin, L.; Capucchio, M.T.; Schiavone, A. Yellow mealworm inclusion in diets for heavy-size broiler chickens: Implications for intestinal microbiota and mucin dynamics. *Animals* **2020**, *10*, 1909. [CrossRef] [PubMed]

13. Moula, N.; Hornick, J.L.; Cabaraux, J.F.; Korsak, N.; Dawans, E.; Antoine, N.; Taminiau, B.; Detilleux, J. Effects of dietary black soldier fly larvae on performance of broilers mediated or not through changes in microbiota. *J. Insects Food Feed.* **2017**, *4*, 31–42. [CrossRef]

14. Mahmood, T.; Guo, Y. Dietary fiber and chicken microbiome interaction: Where will it lead to? *Anim. Nutr.* **2020**, *202*, 1–8. [CrossRef]

15. Tran, G.; Gnaedinger, C.; Mélin, C. Feedipedia, an Online Encyclopedia of Animla Feeds. Available online: https://www.feedipedia.org/node/16388 (accessed on 25 January 2022).

16. Dörper, A.; Veldkamp, T.; Dicke, M. Use of black soldier fly and house fly in feed to promote sustainable poultry production. *JIFF* **2020**, *7*, 761–780. [CrossRef]

17. Allali, I.; Arnold, J.W.; Roach, J.; Cadenas, M.B.; Butz, N.; Hassan, H.M.; Koci, M.; Ballou, A.; Mendoza, M.; Ali, R.; et al. A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. *BMC Microbiol.* **2017**, *17*, 194. [CrossRef]

18. Xia, Y.; Sun, J. Hypothesis testing and statistical analysis of microbiome. *Genes Dis.* **2017**, *4*, 138–148. [CrossRef]

19. Gloor, G.B.; Macklaim, J.M.; Pawlowsky-Glahn, V.; Egozcue, J.J. Microbiome datasets are compositional: And this is not optional. *Front Microbiol.* **2017**, *8*, 2224. [CrossRef]

20. Holmes, I.; Harris, K.; Quince, C. Dirichlet multinomial mixtures: Generative models for microbial metagenomics. *PLoS ONE* **2012**, *7*, e30126.

21. Egozcue, J.J.; Graffelman, J.; Ortego, M.I.; Pawlowsky-Glahn, V. Some thoughts on counts in sequencing studies. *NAR Genom. Bioinform.* **2020**, *2*, lqaa094.

22. Quinn, T.P.; Erb, I.; Gloor, G.; Notredame, C.; Richardson, M.F.; Crowley, T.M. A field guide for the compositional analysis of any-omics data. *Gigascience* **2019**, *8*, giz107. [CrossRef] [PubMed]

23. Weiss, S.; Xu, Z.Z.; Peddada, S.; Amir, A.; Bittinger, K.; Gonzalez, A.; Lozupone, C.; Zaneveld, J.R.; Vázquez-Baeza, Y.; Birmingham, A.; et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **2017**, *5*, 27. [CrossRef] [PubMed]

24. Berg, G.; Rybakova, D.; Fischer, D.; Cernava, T.; Vergès, M.C.; Charles, T.; Chen, X.; Cocolin, L.; Eversole, K.; Corral, G.H.; et al. Microbiome definition re-visited: Old concepts and new challenges. *Microbiome* **2020**, *8*, 103. [CrossRef] [PubMed]

25. Vatcheva, K.P.; Lee, M.; McCormick, J.B.; Rahbar, M.H. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology* **2016**, *6*, 227. [CrossRef] [PubMed]

26. Layeghifard, M.; Hwang, D.M.; Guttman, D.S. Disentangling interactions in the microbiome: A network perspective. *Trends Microbiol.* **2017**, *25*, 217–228. [CrossRef]

27. Dohlman, A.B.; Shen, X. Mapping the microbial interactome: Statistical and experimental approaches for microbiome network inference. *Exp. Biol Med.* **2019**, *244*, 445–458. [CrossRef]

28. Pearl, J. *Causality: Models, Reasoning, and Inference*, 1st ed.; Cambridge University Press: Cambridge, UK, 2000; pp. 1–384.

29. Scanagatta, M.; Salmerón, A.; Stella, F. A survey on Bayesian network structure learning from data. *Prog. Artif. Intell.* **2019**, *8*, 425–439. [CrossRef]

30. Grunwald, P. Minimum description length tutorial. In *Advances in Minimum Description Length. Theory and Applications*, 1st ed.; Grunwald, P.D., Myung, J.I., Pitt, M.A., Eds.; MIT Press: Cambridge, MA, USA, 2005; pp. 23–81.

31. Ji, Z.; Xia, Q.; Meng, G. A review of parameter learning methods in Bayesian network. In *Advanced Intelligent Computing Theories and Applications*, 1st ed.; Huang, D.S., Han, K., Eds.; Springer: London, UK, 2015; Volume 3, pp. 3–12.

32. Saborío-Montero, A.; Gutiérrez-Rivas, M.; García-Rodríguez, A.; Atxaerandio, R.; Goiri, I.; López de Maturana, E.; Jiménez-Montero, J.A.; Alenda, R.; González-Recio, O. Structural equation models to disentangle the biological relationship between microbiota and complex traits: Methane production in dairy cattle as a case of study. *J. Anim Breed. Genet.* **2020**, *137*, 36–48. [CrossRef]

33. Tiezzi, F.; Fix, J.; Schwab, C.; Shull, C.; Maltecca, C. Gut microbiome mediates host genomic effects on phenotypes: A case study with fat deposition in pigs. *Comput. Struct. Biotechnol. J.* **2020**, *19*, 530–544. [CrossRef]

34. Haenlein, M.; Kaplan, A.M. A beginner's guide to partial least squares analysis. *Underst. Statistics* **2004**, *3*, 283–297. [CrossRef]

35. Detilleux, J.; Theron, L.; Duprez, J.N.; Reding, E.; Humblet, M.F.; Planchon, V.; Delfosse, C.; Bertozzi, C.; Mainil, J.; Hanzen, C. Structural equation models to estimate risk of infection and tolerance to bovine mastitis. *Genet. Sel. Evol.* **2013**, *45*, 6. [CrossRef] [PubMed]

36. de Waal, A.; Yoo, K. Latent variable Bayesian networks constructed using structural equation modelling. In Proceedings of the 21st International Conference on Information Fusion, Cambridge, UK, 10–13 July 2018; pp. 688–695.

37. Detilleux, J. Tolerance to bovine clinical mastitis: Total, direct, and indirect milk losses. *J. Dairy Sci.* **2018**, *10*, 3334–3343. [CrossRef] [PubMed]

38. Gunzler, D.; Chen, T.; Wu, P.; Zhang, H. Introduction to mediation analysis with structural equation modeling. *Shanghai Arch. Psychiatry* **2013**, *25*, 390–394. [PubMed]

39. Hoc, B.; Noël, G.; Carpentier, J.; Francis, F.; Caparros, M.; Megido, R. Optimization of black soldier fly (Hermetia illucens) artificial reproduction. *PLoS ONE* **2019**, *30*, e0216160. [CrossRef]

40. Stanley, D.; Hughes, R.J.; Geier, M.S.; Moore, R.J. Bacteria within the gastrointestinal tract microbiota correlated with improved growth and feed conversion: Challenges presented for the identification of performance enhancing probiotic bacteria. *Front. Microbiol.* **2016**, *7*, 187. [CrossRef] [PubMed]

41. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541. [CrossRef]

42. Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **2016**, *4*, e2584. [CrossRef]

43. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2013**, *41*, D590–D596. [CrossRef]

44. Conrady, S.; Jouffe, L. *Bayesian Networks & BayesiaLab: A Practical Introduction for Researchers*, 1st ed.; Bayesia: Franklin, TN, USA, 2015; pp. 1–371.

45. Ryynänen, O.P.; Leppänen, T.; Kekolahti, P.; Mervaala, E.; Töyräs, J. Bayesian network model to evaluate the effectiveness of continuous positive airway pressure treatment of sleep apnea. *Healthcare Inform. Res.* **2018**, *24*, 346–358. [CrossRef]

46. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]

47. Chernick, M.R. The jackknife: A resampling method with connections to the bootstrap. *Wiley Interdisc. Rev. Comput. Stat.* **2012**, *4*, 224–226. [CrossRef]

48. Rychlik, I. Composition and function of chicken gut microbiota. *Animals* **2020**, *10*, 103. [CrossRef] [PubMed]

49. Chica Cardenas, L.A.; Clavijo, V.; Vives, M.; Reyes, A. Bacterial meta-analysis of chicken cecal microbiota. *PeerJ* **2021**, *9*, e10571. [CrossRef] [PubMed]

50. Stanley, D.; Geier, M.S.; Chen, H.; Hughes, R.J.; Moore, R.J. Comparison of fecal and cecal microbiotas reveals qualitative similarities but quantitative differences. *BMC Microbiol.* **2015**, *15*, 51. [CrossRef] [PubMed]

51. Videvall, E.; Strandh, M.; Engelbrecht, A.; Cloete, S.; Cornwallis, C.K. Measuring the gut microbiome in birds: Comparison of faecal and cloacal sampling. *Mol. Ecol. Resour.* **2018**, *18*, 424–434. [CrossRef]

52. Lee, S.J.; Cho, S.; La, T.M.; Lee, H.J.; Lee, J.B.; Park, S.Y.; Song, C.S.; Choi, I.S.; Lee, S.W. Comparison of microbiota in the cloaca, colon, and magnum of layer chicken. *PLoS ONE* **2020**, *15*, e0237108. [CrossRef]

53. Johnson, J.S.; Spakowicz, D.J.; Hong, B.Y.; Petersen, L.M.; Demkowicz, P.; Chen, L.; Leopold, S.R.; Hanson, B.M.; Agresta, H.O.; Gerstein, M.; et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **2019**, *10*, 5029. [CrossRef]

54. Larivière, J.M.; Michaux, C.; Farnir, F.; Detilleux, J.; Verleyen, V.; Leroy, P. Reproductive performance of the Ardennaise chicken breed under traditional and modern breeding management systems. *Int. J. Poultry Sci.* **2009**, *8*, 446–451. [CrossRef]

55. Moula, N.; Antoine-Moussiaux, N.; Farnir, F.; Philippart de Foy, M.; Leroy, P. Performances zootechniques de la poule Ardennaise, une race ancienne pour le futur? *Ann. Méd. Vét.* **2009**, *153*, 66–75.

56. Kaakoush, N.O.; Sodhi, N.; Chenu, J.W.; Cox, J.M.; Riordan, S.M.; Mitchell, H.M. The interplay between Campylobacter and Helicobacter species and other gastrointestinal microbiota of commercial broiler chickens. *Gut Pathog.* **2014**, *6*, 18. [CrossRef]

57. Xiao, S.S.; Mi, J.D.; Mei, L.; Liang, J.; Feng, K.X.; Wu, Y.B.; Liao, X.D.; Wang, Y. Microbial diversity and community variation in the intestines of layer chickens. *Animals* **2021**, *11*, 840. [CrossRef] [PubMed]

58. Zou, A.; Sharif, S.; Parkinson, J. Lactobacillus elicits a 'Marmite effect' on the chicken cecal microbiome. *Jpn. Biofilms. Microbiomes.* **2018**, *4*, 27. [CrossRef]

59. Zelezniak, A.; Andrejev, S.; Ponomarova, O.; Mende, D.R.; Bork, P.; Patil, K.R. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6449–6454. [CrossRef] [PubMed]

60. Sergeant, M.J.; Constantinidou, C.; Cogan, T.A.; Bedford, M.R.; Penn, C.W.; Pallen, M.J. Extensive microbial and functional diversity within the chicken cecal microbiome. *PLoS ONE* **2014**, *9*, e91941. [CrossRef] [PubMed]

61. Svihus, B.; Choct, M.; Classen, H.L. Function and nutritional roles of the avian caeca: A review. *World Poultry Sci. J.* **2013**, *69*, 249–263. [CrossRef]

62. Karasawa, Y. Significant role of the nitrogen recycling system through the ceca occurs in protein-depleted chickens. *J. Exp. Zool.* **1999**, *283*, 418–425. [CrossRef]

63. Eeckhaut, V.; Van Immerseel, F.; Croubels, S.; De Baere, S.; Haesebrouck, F.; Ducatelle, R.; Louis, P.; Vandamme, P. Butyrate production in phylogenetically diverse Firmicutes isolated from the chicken caecum. *Microb. Biotechnol.* **2011**, *4*, 503–512. [CrossRef]

64. Polansky, O.; Sekelova, Z.; Faldynova, M.; Sebkova, A.; Sisak, F.; Rychlik, I. Important metabolic pathways and biological processes expressed by chicken cecal microbiota. *Appl. Environ. Microbiol.* **2015**, *82*, 1569–1576. [CrossRef]

65. Biddle, A.; Stewart, L.; Blanchard, J.; Leschine, S. Untangling the genetic basis of fibrolytic specialization by *Lachnospiraceae* and *Ruminococcaceae* in diverse gut communities. *Diversity* **2013**, *5*, 627–640. [CrossRef]

66. De Maesschalck, C.; Eeckhaut, V.; Maertens, L.; De Lange, L.; Marchal, L.; Daube, G.; Dewulf, J.; Haesebrouck, F.; Ducatelle, R.; Taminau, B.; et al. Amorphous cellulose feed supplement alters the broiler caecal microbiome. *PoultSci* **2019**, *98*, 3811–3817. [CrossRef]

67. Louis, P.; Flint, H.J. Formation of propionate and butyrate by the human colonic microbiota. *Environ. Microbiol.* **2017**, *19*, 29–41. [CrossRef] [PubMed]

68. Li, H.Y.; Zhang, H.L.; Zhao, F.J.; Wang, S.Q.; Wang, Z.X.; Wei, Z.Y. Modulation of gut microbiota, short-chain fatty acid production, and inflammatory cytokine expression in the cecum of porcine Deltacoronavirus-infected chicks. *Front. Microbiol.* **2020**, *11*, 897. [CrossRef] [PubMed]

69. Rivière, A.; Selak, M.; Lantin, D.; Leroy, F.; De Vuyst, L. Bifidobacteria and butyrate-producing colon bacteria: Importance and strategies for their stimulation in the human gut. *Front. Microbiol.* **2016**, *7*, 979. [CrossRef] [PubMed]

70. Lee, K.C.; Kil, D.Y.; Sul, W.J. Cecal microbiome divergence of broiler chickens by sex and body weight. *J. Microbiol.* **2017**, *55*, 939–945. [CrossRef] [PubMed]

71. Amaretti, A.; Gozzoli, C.; Simone, M.; Raimondi, S.; Righini, L.; Pérez-Brocal, V.; García-López, R.; Moya, A.; Rossi, M. Profiling of protein degraders in cultures of human gut microbiota. *Front. Microbiol.* **2019**, *10*, 2614. [CrossRef]

72. Slobodkin, A. The Family *Peptostreptococcaceae*. In *The Prokaryotes*, 1st ed.; Rosenberg, E., DeLong, E.F., Lory, S., Stackebrand, E., Thompson, F., Eds.; Springer: Berlin, Germany, 2014; pp. 291–302.

73. Zhou, C.; Chen, L.L.; Lu, R.Q.; Ma, W.W.; Xiao, R. Alteration of intestinal microbiota composition in oral sensitized C3H/HeJ mice is associated with changes in dendritic cells and T cells in mesenteric lymph nodes. *Front. Immunol.* **2021**, *12*, 631494. [CrossRef]

74. Hedblom, G.A.; Reiland, H.A.; Sylte, M.J.; Johnson, T.J.; Baumler, D.J. Segmented filamentous bacteria - Metabolism meets immunity. *Front. Microbiol.* **2018**, *9*, 1991. [CrossRef]

75. Flint, H.J.; Scott, K.P.; Duncan, S.H.; Louis, P.; Forano, E. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes.* **2012**, *3*, 289–306. [CrossRef]

76. Awad, W.A.; Mann, E.; Dzieciol, M.; Hess, C.; Schmitz-Esser, S.; Wagner, M.; Hess, M. Age-related differences in the luminal and mucosa-associated gut microbiome of broiler chickens and shifts associated with *Campylobacter jejuni* infection. *Front. Cell Infect. Microbiol.* **2016**, *6*, 154. [CrossRef]

77. Stamilla, A.; Ruiz-Ruiz, S.; Artacho, A.; Pons, J.; Messina, A.; Lucia Randazzo, C.; Caggia, C.; Lanza, M.; Moya, A. Analysis of the microbial intestinal tract in broiler chickens during the rearing period. *Biology* **2021**, *10*, 942. [CrossRef]

78. Donaldson, E.E.; Stanley, D.; Hughes, R.J.; Moore, R.J. The time-course of broiler intestinal microbiota development after administration of cecal contents to incubating eggs. *PeerJ* **2017**, *5*, e3587. [CrossRef] [PubMed]

79. De Vilela Souza, J.V.; Kheravii, S.K.; Bajagai, Y.S.; Kolakshyapati, M.; Wu, S.; Ruhnke, I. Inclusion, of black soldier fly larvae in a meat chicken diet has minor effect on caeca microbiota. In Proceedings of the 32nd Annual Australian Poultry Science Symposium; University of Sydney: Sydney, Australia, 2021; Volume 32, p. 146.

80. Dabbou, S.; Lauwaerts, A.; Ferrocino, I.; Biasato, I.; Sirri, F.; Zampiga, M.; Bergagna, S.; Pagliasso, G.; Gariglio, M.; Colombino, E.; et al. Modified black soldier fly larva fat in broiler diet: Effects on performance, carcass traits, blood parameters, histomorphological features and gut microbiota. *Animals* **2021**, *11*, 1837. [CrossRef] [PubMed]

81. Wu, Y.; Zhang, H.; Zhang, R.; Cao, G.; Li, Q.; Zhang, B.; Wang, Y.; Yang, C. Serum metabolome and gut microbiome alterations in broiler chickens supplemented with lauric acid. *Poult. Sci.* **2021**, *100*, 101315. [CrossRef] [PubMed]

82. Abd El-Hack, M.E.; Shafi, M.E.; Alghamdi, W.Y.; Abdelnour, S.A.; Shehata, A.M.; Noreldin, A.E.; Ashour, E.A.; Swelum, A.A.; Al-Sagan, A.A.; Alkhateeb, M.; et al. Black soldier fly (*Hermetia illucens*) meal as a promising feed ingredient for poultry: A comprehensive review. *Agriculture* **2020**, *10*, 339. [CrossRef]

83. Weththasinghe, P.; Hansen, J.Ø.; Mydland, L.T.; Øverland, M. A systematic meta-analysis based review on black soldier fly (*Hermetia illucens*) as a novel protein source for salmonids. *Rev. Aquac.* **2021**. [CrossRef]

84. Shumo, M.; Osuga, I.M.; Khamis, F.M.; Tanga, C.M.; Fiaboe, K.K.M.; Subramanian, S.; Ekesi, S.; van Huis, A.; Borgemeister, C. The nutritive value of black soldier fly larvae reared on common organic waste streams in Kenya. *Sci. Rep.* **2019**, *9*, 10110. [CrossRef]

85. DiGiacomo, K.; Leury, B. Review: Insect meal: A future source of protein feed for pigs? *Animal* **2019**, *13*, 3022–3030. [CrossRef]

86. Dabbou, S.; Gai, F.; Biasato, I.; Capucchio, M.T.; Biasibetti, E.; Dezzutto, D.; Meneguz, M.; Plachà, I.; Gasco, L.; Schiavone, A. Black soldier fly defatted meal as a dietary protein source for broiler chickens: Effects on growth performance, blood traits, gut morphology and histological features. *J. Anim. Sci. Biotechnol.* **2018**, *9*, 49. [CrossRef]

87. Marmion, M.; Ferone, M.T.; Whyte, P.; Scannell, A.G.M. The changing microbiome of poultry meat; from farm to fridge. *Food Microbiol.* **2021**, *99*, 103823. [CrossRef]

88. Moula, N.; Scippo, M.L.; Douny, C.; Degand, G.; Dawans, E.; Cabaraux, J.F.; Hornick, J.L.; Medigo, R.C.; Leroy, P.; Francis, F.; et al. Performances of local poultry breed fed black soldier fly larvae reared on horse manure. *Anim Nutr.* **2018**, *4*, 73–78. [CrossRef]

89. Grünwald, P.D. Introducing the minimum description length principle. In *Advances in Minimum Description Length. Theory and Applications*, 1st ed.; Grünwald, P.D., Myung, I.J., Pitt, M.A., Eds.; MIT Press: London, UK, 2004; pp. 3–23.

90. Lehmann, D.R.; Hulbert, J. Are three-point scales always good enough? *J. Marketing Res.* **1972**, *9*, 444–446. [CrossRef]

91. Kim, S.K.; Frisby, C.L. Gaining from discretization of continuous data: The correspondence analysis biplot approach. *Behav. Res.* **2019**, *51*, 589–601. [CrossRef] [PubMed]

92. Safari, S.; Baratloo, A.; Elfil, M.; Negida, A. Evidence based emergency medicine; Part 5 Receiver Operating Curve and Area under the Curve. *Emergency* **2016**, *4*, 111–113. [PubMed]