Efficacy of fecal sampling as a gut proxy in the study of chicken gut microbiota

RUNNING TITLE: Efficacy of fecal sampling as a gut proxy

Wei Yan[a], Jiangxia Zheng[a], Chaoliang Wen[a], Congliang Ji[b], Dexiang Zhang[b], Yonghua Chen[b], Congjiao Sun[a, #] and Ning Yang[a, #]

[a]National Engineering Laboratory for Animal Breeding and MOA Key Laboratory of Animal Genetics and Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, 100193, China

[b]Wen's Nanfang Poultry Breeding Co. Ltd, Guangdong Province, Yunfu, 527400, China

[*]Correspondence: Ning Yang, Department of Animal Genetics and Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, 100193, China

E-mail: nyang@cau.edu.cn

or Congjiao Sun, Department of Animal Genetics and Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, 100193, China

E-mail: cjsun@cau.edu.cn

ABSTRACT Despite the convenience and noninvasiveness of fecal sampling, the fecal microbiota does not fully represent that of the gastrointestinal (GI) tract, and the efficacy of fecal sampling to accurately represent the gut microbiota in birds is poorly understood. Using chickens as a model, we collected 1,026 samples from 206 animals, including duodenum, jejunum, ileum, cecum and feces samples. Most taxa in the small intestine (94.10 – 94.82%) and ceca (99.57%) could be identified in feces. Microbial community membership was reflected with a gut anatomic feature, but community structure was not. Excluding shared microbes, the small intestine and ceca contributed 26.69 and 2.36% of the total fecal members, respectively. The composition of Firmicutes members in the small intestine and that of Actinobacteria, Bacteroidetes and Proteobacteria members in ceca could mirrored that observed in fecal samples well (ρ = 0.68 – 0.79 and 0.66 – 0.79, respectively, $P < 0.05$). Enterotype-like clustering was performed in GI tract and all sites were clustered into 2 or 3 enterotype-like clusters. Feces from different clusters reflected the GI microbiota with different efficacies, giving a new insight into observing efficacy of feces as a gut proxy. Our results provide evidences that the good potential of feces to identify most taxa in chicken guts, but microbial structure analyses using feces as a proxy for gut should be interpreted with caution.

IMPORTANCE Fecal sample is the important object used in gut microbial study, as the collection of feces is convenient and noninvasive. It is well known that the microbial community in fecal sample may not be fully representative of that in gastrointestinal tract (GI), but the extent to which the fecal sample reflect the microbiota in GI is not fully clear. Instead of focus on the variation of different sampling sites, this study demonstrated the spatial relationships of microbiota among duodenum, jejunum, ileum, cecum and feces and partitioned the efficacy of

feces as a gut proxy to quantitatively identify the extent to which fecal samples are appropriately used in gut study. Enterotype concept was used to confirm the efficacy of feces as a gut proxy. This work gives insights and provide future directions regarding the usage of fecal samples in studies of the gut microbiome.

KEYWORDS feces, proxy, gut, spatial relationships, enterotype, chicken

INTRODUCTION

Many studies have reported on the important roles of gut microbiota in affecting host metabolism and health in humans (1), other mammals (2) and in birds (3). Because of the convenience and noninvasiveness of fecal sampling, most studies use fecal samples as a proxy to study the gut microbiota, despite the increasing recognition that fecal microbial populations may not be fully representative of those in the gastrointestinal (GI) tract contents or mucosa (4, 5). Therefore, a comprehensive understanding of the efficacy of using fecal samples as a proxy to study the GI microbiota would help improve longitudinal analyses of microbiota.

Among birds, the chicken is frequently used as a research model in research, and its GI microbiota has been studied previously(6-10). In several studies, the microbiota present in different GI segments have been investigated using traditional sequencing methods (11) or high-throughput sequencing techniques (12, 13). However, these studies had small sample sizes (N = 3 – 8) and were primarily aimed at examining the spatial heterogeneity among different segments and did not focus on the spatial microbiota relationships between feces and the GI tract.

Compared with most mammals, the cecum has a specific structure in birds and has been reported to play important roles in bird metabolism, such as in the digestion of cellulose, starch and other resistant polysaccharides (14, 15) and in the absorption of nutrients (16) and water (17). Microbial compositions and functions in chicken ceca have been reported in many studies (18, 19). In addition, Stanley et al. (20) examined the microbial relationships between the ceca and feces and observed that 88.55% of all operational taxonomic units (OTUs) were shared. However, the microbial relationships between the ceca and small intestine (including the duodenum, jejunum and ileum) were not reported, data which could provide an integrated view of gut microbial

relationships.

The microbial communities in human feces were previously partitioned into 3 clusters, referred to as enterotypes (21), which varied from 2 to 3 in subsequent studies in humans (22-24). The concept was later extended to other mammalian hosts, such as chimpanzees (25) and mice (26). This concept is rarely used for birds, except in a study on fecal samples from 31 chickens (27). Thus, enterotype-like clustering was performed in the current study to identify the clusters present in different sampling sites and help improve our understanding of the microbial relationships among feces, ceca and small intestine.

Using chicken as a model, we performed large-scale sequencing surveys and focused on the efficacy of using feces to represent the GI microbiota in birds. The efficacy was partitioned into microbial community membership and structure to gain a comprehensive view concerning this issue. Using enterotype clustering methods (21), this study reports the enterotype-like clusters in the duodenum, jejunum, ileum and cecum of birds, improving our understanding of the efficacy of the use of feces as a proxy to study the gut microbiota and their spatial relationships in the gut.

RESULTS

Landscape and quantification of microbial relationships among feces, ceca and small intestine. To gain an overview of the microbial relationships among the chicken duodenum, jejunum, ileum, ceca and feces, unweighted UniFrac distances (community membership; presence/absence of taxa) and weighted UniFrac distances (community structure; taking the relative abundances of taxa into account) were used to perform principal co-ordinates analysis (PCoA; Fig. 1a, c). The variation in community memberships among different sites were primarily explained by the sites origin (Fig. 1a), but the community structures showed both the sites origin and inter-individual variation (Fig. 1c). In particular, the cecal microbial community exhibited a distant relationship with the small intestine community, and the microbial community in feces showed an intermediate relationship between those of the ceca and small intestines.

UniFrac distances between two samples from all assayed sites within each individual were calculated to quantify the spatial relationships of the gut microbiota. When the community membership was considered only, the UniFrac distance decreased along the gut anatomical locations from the farthest to the nearest sites between fecal and duodenal, jejunal, ileal or cecal samples (FD, FJ, FI or FC, respectively, in Fig. 1b), presenting clear anatomical differences. However, when taking the community structure into account, the UniFrac distance increased in FI and FC compared with that in FJ (Fig. 1d). This finding might be explained by the exchange of contents between the ileum and ceca, suggesting that the specific cecal microbial structure influences the microbial communities in the ileum and feces.

Among all pairs, the unweighted UniFrac distance between the cecal and duodenal samples was highest ($P < 0.05$), and that between duodenal and jejunal samples was lowest ($P < 0.05$; Fig. 1b

and Table S1). Regarding the weighted UniFrac distances, cecal samples had similar distances to the duodenal and jejunal samples and were higher than the other pairs ($P < 0.05$), whereas the lowest distance was observed between duodenal and jejunal samples ($P < 0.05$; Fig. 1d and Table S1). These results suggest that limited differences exist within small intestinal microbial communities, while the microbial structure in the ceca is quite distinct from those in the small intestine.

Analyses of shared and exclusive microbial members. Given that both the community membership and structure influence the microbial relationships among the feces, ceca and small intestine, we next evaluated the extent to which the spatial relationships were influenced by the above two factors. The shared and exclusive OTUs were calculated to assess the influence of the microbial community membership. We observed that 1604 OTUs, accounting for 50% of total OTUs, were shared across all sites (Fig. 2a), and these shared OTUs can be referred to as the "core" microbiota in the gut. These OTUs represented different proportions of sequences in different sites and were especially high in fecal samples (98.48%; Fig. 2b), indicating that the most abundant members detected in fecal samples belonged to these "core" microbiota. At the genus level, these core taxa were primarily classified as unclassified *Clostridiaceae*, *Lactobacillus*, *Bacteroides*, unclassified *Bacteroidales* and *Gallibacterium* (Fig. 2c). It is noteworthy that 5.66% of the "core" microbiota sequences were not assigned and that most of these sequences (71.11%) were detected in the cecal samples (small pie chart in Fig. 2c), suggesting that most of these unassigned taxa tended to be anaerobic microbes.

Most OTUs in the small intestine (94.1 – 94.82%) and cecal (99.57%) samples could be identified as fecal OTUs (Table 1), indicating that feces would be an excellent proxy for

identifying species in the gut microbiota. However, OTUs from the GI tract that were still present (5.18 – 5.90% in small intestinal and 0.43% in cecal samples) remained undetected in fecal samples (Table 1). Of these taxa, *Symbiobacterium*, members of the phylum Chloroflexi and α-, β- and γ-Proteobacteria appeared to be particularly undetected feces (Supplementary Table 2).

Microbial communities in the small intestine and ceca did not contribute equally to the fecal microbial members, as 27.01% of fecal OTUs were not identified in cecal samples, most of which (26.69%) could be identified in small intestinal niches (Table 1 and Fig. 2d). These OTUs were primarily from the orders *Alteromonadales*, *Bacillales*, *Lactobacillales*, *Actinomycetales*, *Sphingomonadales* and *Enterobacteriales* (Fig. 2e) and were considered exclusive contributors of the small intestinal microbiota to fecal microbial members. The ceca exclusively contributed 2.36% of OTUs to the observed fecal members, representing 0.11% of the fecal sample sequences and consisting of taxa primarily from the orders *Cenarchaeales*, *Bacteroidales*, *SB-45*, *Chlamydiales* and *Thermococcales* (Fig. 2f).

Correlation analyses of microbial abundances. Because community structure also affects the spatial relationships of gut microbiota, we next performed Spearman correlation analyses between the mean fecal and segmental genera abundance to evaluate the effects of community structure and assess the extent to which the microbial community in GI tract was reflected in the fecal samples (Fig. 3). If a high correlation was observed between two sites, the differences in abundance between sites were considered highly consistent, so that the abundance at one site had the potential to be a good proxy for the abundance at another. The microbial composition of feces was moderately correlated with those in the small intestine (Spearman: $\rho = 0.43$; $P < 0.05$) and in the combination of small intestine and ceca ($\rho = 0.50$; $P < 0.05$; Fig. 3a). We then performed similar

analyses to identify the correlation bias in predominant phyla (Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria; Fig. S2). Genera of the Firmicutes and Proteobacteria phyla in fecal samples showed moderate to high correlations with those in all of four GI sites (ρ = 0.43 – 0.79, P < 0.05). In particular, fecal samples were highly representative of Firmicutes members in the small intestine (ρ = 0.68 – 0.79, P < 0.05) and of Actinobacteria, Bacteroidetes and Proteobacteria members in ceca (ρ = 0.66 – 0.79, P < 0.05). However, Actinobacteria members in the small intestine might not be well represented in fecal samples (ρ = 0.15 – 0.25, P > 0.05; Fig. S2).

A follow-up question concerned the extent to which each microbe correlated between two sites. To address this issue, Spearman correlation tests were performed for each genus between two sites. The genera with abundances of > 0.1% at either compared site with a significant correlation (P < 0.05) are summarized in Fig. S3. Between the fecal and each of the 4 gut segmental samples, a limited number of significant correlations (P < 0.05) were observed, and these correlations were not high (ρ = -0.2 – 0.4, P < 0.05) for each genus. Most genera with significant correlations belonged to the phyla Firmicutes and Proteobacteria. However, more significant and moderate correlations were observed between two of the small intestinal segments, and most of the genera with significant correlations were also from the phyla Firmicutes and Proteobacteria (Table S3). The results suggest that the gut microbiota structures could be moderately reflected by fecal samples when taking all genera into consideration simultaneously, but analyses of fluctuations in abundance for a specific genus should be interpreted with caution.

Although microbes at one site were weakly correlated with the corresponding microbes at another site, certain patterns were observed in some cases, as exemplified by the genus

Campylobacter (Table S3). The abundance of this genus in ceca exhibited consistent correlations with that observed in the jejunum (ρ = 0.21, P < 0.05) and ileum (ρ = 0.36, P < 0.05). In ileal samples, this genus was correlated with that measured in fecal samples (ρ = 0.19, P < 0.05), while no correlation was observed between cecal and fecal samples. This finding indicates that *Campylobacter* has great colonization ability in the distal gut of chickens, especially in ceca, and most *Campylobacter* contributions to the fecal composition are probably from the ileum, but not ceca.

Enterotype clustering analyses. The above analyses showed the spatial relationships of the gut microbiota among chicken feces, ceca and small intestine by regarding each site as an entirety, and the "entirety" could be clustered into different enterotype-like groups (21), which might improve our observations. To address this issue, enterotype analyses of different sites were implemented using the methods reported by Arumugam et al. (21).

The microbial communities in the duodenum, jejunum and ceca were clustered into two clusters, while those in ileum and feces were clustered into three (Fig. 4a). Each of the clusters was characterized by signature taxa that were overrepresented in one cluster compared to their representation in the others within each site. The signature taxa of clusters in different sites presented different but anatomy-linked features (Fig. 4b). In the small intestine, *Ochrobactrum* and *Rhodococcus* were overrepresented in duodenal cluster 1, jejunal cluster 2 and ileal cluster 2. An unassigned taxon was overrepresented in duodenal cluster 2, jejunal cluster 2 and ileal cluster 2, and this taxon was identified as New.ReferenceOTU2622 based on OTU level analyses. These signature taxa showed conserved overlaps among clusters within the small intestine. Similarly, the signature taxa in fecal clusters presented linkages with those in GI clusters. Fecal cluster 1 was

overrepresented by unclassified *Clostridiaceae*, which also occurred in jejunal cluster 1 and ileal cluster 1. *Bacteroides* was overrepresented in fecal cluster 2, as well as in ileal cluster 4 and cecal cluster 1. However, *Gallibacterium* and *Lactobacillus* were overrepresented in fecal cluster 3, which did not occur any of the GI clusters (Fig. 4b).

To understand whether enterotype-like clustering would affect the spatial relationships of gut microbiota, individual animals were divided into 3 groups according to the fecal enterotype-like clusters (Enterotype-like clusters 1, 2 and 3, F1, F2 and F3). Within each group, the UniFrac distances between feces and each of four GI segments were calculated, and each distance was compared among three groups. The results showed that both the community membership (unweighted UniFrac distance) and structure (weighted UniFrac distance) were associated with the enterotype-like clustering (Fig. 5). Although membership was not significantly affected by the clustering between F1 and F3, fecal samples in F2 exhibited a higher membership similarity with cecal samples but a lower similarity with small intestine samples than the other two cluster groups (Fig. 5a). Regarding community structure, fecal samples in F2 also exhibited a higher similarity with cecal samples than the other two groups. Nevertheless, similarities between fecal samples and each of the GI segments were lower in F1 than in F2 and F3 (Fig. 5b). The results suggested that the fecal microbial community in F2 might be influenced by cecal microbiota and be more representative of the cecal microbial community. Furthermore, the fecal samples in F1 had reduced abilities to reflect the microbial community structure of the GI tract.

Spearman correlation analyses at the genus level were performed to assess the extent to which GI communities were reflected by fecal samples in different fecal clusters. All three enterotype groups presented moderate correlations between fecal and GI samples (Fig. 3b-d), and the

correlations were consistent with the results in comparisons of the weighted UniFrac distance, as shown in Fig. 5b, confirming the microbial relationships between feces and GI tract among three enterotype groups. To assess the correlation bias in the abundant phyla (Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria), we next performed similar correlation analyses between feces and each of the GI segments within each fecal cluster and specific phylum, and the distribution of correlations is shown in Fig. S4. Fecal samples in all three groups showed correlations with corresponding cecal samples (ρ = 0.36 – 0.84, P < 0.05) in all four phyla. In particular, F2 showed high correlations (ρ = 0.69 – 0.84, P < 0.05) between fecal and cecal samples, confirming the close relationships between these two sites in F2. In the phylum Actinobacteria, most correlations between fecal and small intestine samples in both F1 and F3 were low or not significant (ρ = 0.34, P < 0.05). In the phylum Firmicutes, fecal samples in F2 and F3 were highly correlated with small intestine samples (ρ = 0.66 – 0.83, P < 0.05). The results provide additional evidence that the enterotype-like clustering influences the efficacy of fecal sampling for studying gut microbial communities.

DISCUSSION

The efficacy of using fecal samples to reflect the gastrointestinal microbiota. To our knowledge, this is the first large-scale sequencing assessment on the efficacy of using fecal samples as a proxy for the gut microbiota in birds. In this study, we comprehensively examined the membership, structure and enterotype-like clusters of the chicken gut microbiome at five different biogeographic sites within 206 individual animals. We showed that fecal samples were good proxies for detecting the presence/absence of GI microbial members because most GI tract members could be detected within anatomic features in fecal samples (microbial communities in feces showed increasing similarities to those in the GI tract along the duodenum-jejunum-ileum-ceca axis). However, phyla bias and inter-individual and enterotype-like clustering effects were observed to affect the efficacy of using fecal samples to study GI microbial abundance.

Similar to the current study, a high proportion of shared OTUs has been previously observed between fecal and cecal samples in chickens (20). Similarly, a study in house mice observed that 93.3% of OTUs were shared between fecal and lower GI samples (28). Another chicken study indicated that the GI origin is a primary determinant for the chicken fecal microbiota composition (29), supporting the high proportion of shared OTUs between feces and the four gut segments observed in the current study. These results indicate that fecal samples have good potential for identifying microbial members derived from the GI tract. However, another chicken study (12) observed low percentages of shared OTUs between segments. A major reason for the differences among studies might be the small sample size in Choi's study, which would increase the sensitivity of the results with respect to individual variation. Moreover, the presence/absence of

microbial members in the GI tract was observed to be reflected by fecal samples in a given anatomical feature, i.e., fecal samples had more similarities in community membership to those in ileal and cecal samples than in duodenal and jejunal samples, consistent with previous reports in birds (13) and mammals (28, 30).

As for microbial community structure, the efficacy of using fecal samples to represent the gut microbiota structure did not work as well as for community membership. First, the weighted UniFrac distances between feces and each of intestinal segments was significantly higher than the corresponding unweighted UniFrac distances (Fig. S5), suggesting that taking the abundance into account significantly increased the dissimilarity between feces and each of the GI segments. Second, the abundances of most taxa were significantly different between fecal and GI samples (Table S4), consistent with previous studies(11, 13, 31). Third, the correlations between the mean fecal and segmental genera abundances were moderate, similar to the results obtained for rhesus macaques (32). However, these correlations display bias among different phyla, i.e., different phyla in the GI tract are differentially mirrored by fecal samples. Fourth, significant correlations ($P < 0.05$) of each microbe between fecal and segmental samples were low and rare, suggesting that the efficacy of using fecal samples to represent microbial abundance was affected by the inter-individual effect. A similar effect has also been observed in humans (4).

In addition, the efficacy of using fecal samples to represent the gut microbiota was affected by enterotype-like clustering. In this study, we observed that fecal samples in cluster F2 were more representative of the cecal microbiota than in F1 and F3. We infer that this phenomenon might be explained by the greater susceptibility of the fecal microbial community in cluster F2 group to the cecal microbiota, making the compositions of the microbial community in feces and ceca more

identical, e.g., the dominant genus in the cecum and the overrepresented genus in the feces in cluster F2 was *Bacteroides*. The *Bacteroides* enterotype is broadly observed in fecal samples in chickens (27), humans (21, 24) and other mammals (25, 26), but the relationships between fecal and GI samples within this enterotype have been rarely reported. Therefore, more studies are required to elucidate the enterotype-like clustering effects on the efficacy of using fecal samples to reflect gut microbial profiles.

Previous studies in humans (4, 33) and other mammals (32, 34) have also addressed the issue of whether fecal samples are good representatives for GI microbial analyses. Although the conclusions may not be fully consistent, nearly all studies reached a consensus that microbial communities in fecal samples do not represent the whole GI microbiota. Studies in humans suggest that microbial communities in the duodenum and colon are not represented by those in feces because of the large differences in microbial profiles (33), and these studies emphasized the need to examine tissue biopsies in addition to fecal samples (5), proposing that standard forceps mucosal biopsy samples can represent bacterial populations (4). Compared with human studies, studies in other mammals are more comprehensive because a larger number of gut segments can be involved in the analyses. Several studies in mice (28, 34) support the utility of fecal samples for studying the gut microbiota, because microbial communities in fecal samples were observed to be similar to those in the lower GI tract, which is supported by studies conducted in rhesus macaques (32), pigs (35) and equines (36).

Compared with previous studies, the strength of the current study lies in the following: 1) it involved the use of gut segments from the upper GI tract to the lower GI tract and feces, providing a relatively comprehensive view of the spatial relationships of the gut microbiota; 2) the microbial

relationships were partitioned into two parts, i.e., microbial community membership and structure, providing multi-angle observations to identify microbial relationships between feces and the GI tract; 3) a massive number of individuals was sampled, which is significant for investigations of gut spatial relationships, as the sizes of most of the above studies did not exceed twenty; and 4) the use of enterotype-like clustering analyses provided new insights into exploring the utility of fecal samples in studies of the gut microbiota.

Specific roles of the cecal microbial community in the gut. Because of the specific and significant roles in nutrition and health (15, 37), ceca have been widely investigated in birds (38, 39), especially chickens (18, 19, 40). *Bacteroides* was observed as the dominant taxa in our sturdy (Fig. S6) and in most other studies (41, 42), although some reports observed a predominance of *Clostridiales* members in ceca (12, 43). Although the cecal microbial community may sometimes be linked to diet (38), the nearly consistent results across studies suggests the cecal microbial community is stable. This finding might be due to ceca having a special blind-ended structure and are located in the lower GI tract, providing a stable and anaerobic environment for microbes and longer storage periods of the contents, in contrast to the rapid transit environment in the small intestine (44). In addition to the microbial composition, Stanley et al. (20) also compared microbial differences and similarities between ceca and feces in chicken. They observed that 88.55% of all OTUs, containing 99.25% of all sequences, were shared by the ceca and feces, similar to the observations in the current study. These results indicate that except for some rare microbial members, most microbes in the ceca can be detected in fecal samples.

The microbial relationships between the ceca and small intestine have been rarely reported in birds. Choi et al. (12) compared the percentage of shared OTUs among ceca and three small

intestinal sections but observed low percentages between the segments (ranging from 1.2 to 2.9%, representing from 38.7 to 65.5% of sequences). The percentages reported in another study (60.2% for the duodenum, 50.5% for the jejunum and 43.5% for the ileum, which were calculated from Fig. 3 in their article) were higher than those in Choi's study. In contrast, the results of Xiao's study presented an opposite trend from our findings, i.e., the percentages of shared OTUs in Xiao's study decreased from the duodenum to the jejunum and ileum, demonstrating a reversed-anatomical feature compared with the current study. These inconsistent results might be attributable to differences among species, diets or other environmental factors, but the small sample size in Xiao's study may be an important reason for these inconsistencies.

Enterotype-like clustering in chicken gut microbiota. Enterotype-like clustering is of increasing concern and has recently led to heated discussions (45, 46). To the best of our knowledge, the current study describes the first attempt of enterotype-like clustering for the GI tract of birds. Three enterotype-like clusters (*Ochrobactrum*, *Rhodococcus* and New.ReferenceOTU2622) appeared to be conserved in the small intestine, demonstrating a close microbial relationship within the small intestine. Additionally, the enterotype-like cluster *Bacteroides* and unclassified *Clostridiaceae* appeared in the jejunum and ileum with the oxygen concentration, pH and nutrient changes along the GI tract. Both New.ReferenceOTU2622 clusters in the duodenum and ileum had a lower diversity than the corresponding clusters (Fig. S7), suggesting a potential depression role of this taxon compared to many other microbes. Cecal samples were clustered into two enterotype-like clusters, but the silhouette coefficient was low (0.08, Fig. S1f). This issue might be attributable to similar dominant genera (*Bacteroides*, unclassified *Bacteroidales* and unclassified *Clostridiales*) in both enterotype-like clusters, such

that the similarity between the two clusters was high (ADONIS: $R^2 = 0.021$, $P < 0.05$; Fig. S6). Similarly, the clustering in ceca did not significantly lead to a difference in microbial diversity (Fig. S7). These results also provided evidence that the microbial community in ceca tended to be more stable than other GI segments.

The chicken fecal samples were previously clustered into four potential enterotypes: enterotype 1, dominated by Firmicutes; enterotype 2, dominated by Firmicutes and Proteobacteria; enterotype 3, dominated by Firmicutes and Actinobacteria; and enterotype 4, dominated by Firmicutes and Bacteroidetes (27). This result is not fully consistent with the enterotype-like clusters observed in our study. Regarding cluster numbers, Kaakoush et al. (27) observed four potential enterotypes, while we observed three. At the phylum level, the clusters in our study were overrepresented by Firmicutes, Bacteroidetes and Proteobacteria, but not Actinobacteria. At the genus level, the *Lactobacillus* and *Bacteroides* clusters were observed in both Kaakoush's and our studies, despite differences in the species, environment and sample size. This finding indicates that *Lactobacillus* and *Bacteroides* clusters have the potential to be conserved clusters in chickens. The *Lactobacillus* cluster would account for the largest microbial proportion (53.0% in our study) in fecal samples, because *Lactobacillus* has been widely reported as the dominant genus in the feces of domestic poultry (29, 47) and pet birds (48). Because microbial diversity is linked to the stability of the microbial community (49, 50), the community in the *Bacteroides* cluster has the potential to be more stable with a higher diversity than in other fecal clusters ($P < 0.001$; Fig. S7). However, the results are not fully consistent across studies in chickens (27), humans (23, 45) and mammals (26, 46), suggesting that the relationship between enterotype-like clustering and microbial diversity might be influenced by differences in species, cohorts within species and the environment.

In humans, an enterotype classifier with a flow that includes two routes to obtain enterotype assignments has been developed (45). We applied the classifier to our dataset, and none of the duodenum, jejunum, ileum and feces samples were similar to stool samples from large-scale projects, such as MetaHIT and HMP. In ceca, 71 samples matched the classifier sample criteria, and most were clustered as enterotype *Bacteroides*, consistent with the current study (data not shown). Differences in the physiological structure of the GI tract (e.g., the short colon with numerous flat villi and relatively few goblet cells) (51), diet and the environment between humans and birds make this enterotype classifier not fully suitable for data collected from birds, making it necessary to develop an enterotype classifier for birds in the future that can be used to gain a better understanding of their gut microbiota and to detect disorders.

In conclusion, we assessed the efficacy of using fecal samples to represent GI microbiota in birds and analyzed potential factors affecting this efficacy. With highly shared microbial members, fecal samples could be used to detect most microbial species in the small intestine and ceca with gut anatomical features. However, analyses of microbial structures using fecal samples as the proxy for the gut in longitudinal microbial studies should be interpreted with caution. Moreover, we described the first attempt to perform enterotype-like clustering in GI segments and observed that the clustering affected the efficacy of using fecal samples to represent the GI microbiota. This study is one of the first attempts to identify the microbial relationships between feces and the intestine in birds, which will help extend our understanding of the bird gut microbiota and provide future directions regarding the usage of fecal samples in studies of the gut microbiome.

MATERIALS AND METHODS

Animal model. The complete procedure was performed according to the guidelines established by the Animal Care and Use Committee of China Agricultural University (permit number: SYXK 2013-0013).

The slow-growing yellow broiler was used as the animal model in this study, and the birds were obtained from Wen's Nanfang Poultry Breeding Co., Ltd. in Guangdong Province of China. Two hundred and six birds with similar body weights were selected and raised on the ground with *ad libitum* feeding and nipple drinkers. The birds were fed a common maize-soybean-based diet throughout the duration of the experiment. No antibiotics were applied during the thirty-five days before sample collection. Because chickens are the largest population of birds on earth, the chicken was selected as a bird model for this investigation. The slow-growing yellow broiler has not been highly selected for production, making this breed of chicken closer to the ancestral birds.

Sample collection. Fresh fecal samples were collected from each bird as soon as excreta was discharged through the cloaca at 77 days of age. Next, all the birds were humanely euthanized and dissected. The contents and mucosal surfaces of the duodenum, jejunum, ileum and cecum were collected immediately after dissection. To ensure the consistency of samples among individuals, a 10-cm-long fixed section of the duodenum and jejunum, and the whole ileum and a pair of ceca were selected for sampling from each bird. The contents and mucosa were mixed uniformly before collection. All samples were immediately placed in liquid nitrogen and then stored at -80°C. Both the intestinal contents and mucosa were sampled based on the consideration that the microbes from both sources may contribute to host interactions with respect to nutrient metabolism and immunity (52).

DNA extraction and 16S rRNA gene sequencing. DNA was extracted from intestinal and fecal samples using a QIAamp DNA stool mini kit (QIAGEN, cat#51504) (53) following the manufacturer's instructions. PCR amplification of the V4–V5 region of the bacterial 16S rRNA gene was performed using the forward primer 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and the reverse primer 907R (5'-CCGTCAATTCMTTTRAGTTT-3'). Sample-specific 7-bp barcodes were incorporated into the primers for multiplex sequencing. The PCR reactions contained 5 μl of Q5 reaction buffer (5×), 5 μl of Q5 High-Fidelity GC buffer (5×), 0.25 μl of Q5 High-Fidelity DNA Polymerase (5 U/μl), 2 μl of dNTPs (2.5 mM), 1 μl (10 μM) of each forward and reverse primer, 2 μl of DNA template, and 8.75 μl of ddH_2O. Thermal cycling consisted of initial denaturation at 98°C for 2 min, followed by 25 cycles of denaturation at 98°C for 15 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s, with a final extension at 72°C for 5 min. PCR amplicons were purified using Agencourt AMPure Beads (Beckman Coulter, Indianapolis, IN) and quantified using a PicoGreen dsDNA Assay kit (Invitrogen, Carlsbad, CA, USA). After the quantification step, amplicons were pooled in equal amounts, and 2 × 300 bp paired-end sequencing was performed using an Illumina MiSeq platform with the MiSeq Reagent kit v3 at Shanghai Personal Biotechnology Co., Ltd. (Shanghai, China). The raw data on which the conclusions of the manuscript rely has been deposited in the National Center for Biotechnology Information (NCBI) database (accession number SRP139192, SRP139193 and SRP139195).

Analysis of sequencing data. Data analysis was performed using the Quantitative Insights Into Microbial Ecology (QIIME, v1.8.0) pipeline (54). Briefly, raw sequencing reads with exact matches to the barcodes were assigned to respective samples and identified as valid sequences. The low-quality sequences were filtered based on the following criteria (55, 56): length < 150 bp,

average Phred score < 20, ambiguous bases, and mononucleotide repeats > 8 bp. Paired-end reads were assembled using FLASH (57), and chimera detection was performed with QIIME. After quality control, four fecal samples were excluded due to low sequence quality that was potentially caused by a technical artifact. The remaining high-quality sequences were clustered into operational taxonomic units (OTUs) at a 97% sequence identity using an open-reference OTU picking protocol against the Greengenes database (58).

We focused on open-reference OTU picking for these analyses because this method yields substantially more taxonomic identifications with sequences that failed to hit the reference database than do closed-reference methods. The open-reference method can provide more information for enterotype-like clustering and comparisons among intestinal segments or feces. The singleton OTUs were discarded because such OTUs can occur due to sequencing errors. Only OTUs representing more than 0.001% of the total filtered OTUs were retained to improve the efficiency of the analysis. Because the sequencing and sampling quantity varied among individuals, we rarefied the data to the lowest sequences per sample to control for sampling effort in diversity analyses. Alpha and beta diversity of individual OTUs were calculated with post-rarefaction data and the phylogenetic tree. Principal coordinate analysis (PCoA) was performed using the unweighted or weighted UniFrac distance (59) for different intestinal segments and feces.

Enterotype-like clustering was performed as previously described (21). In brief, the Calinski–Harabasz (CH) index was calculated with PAM clustering to determine the optimal number of clusters (Fig. S1a-e). The silhouette scores were calculated for each cluster to assess the robustness of the clustering (Fig. S1f). Principal coordinate analysis (PCoA) was implemented with the

dudi.pco function using the *ade4* package, and the cluster plots were performed using the *rgl* packages in **R**. The correlations between the mean fecal and segmental genera abundance were calculated using the method described in a study of rhesus macaques (32).

Statistical analysis. We used Mann-Whitney tests to identify overrepresented genera in each cluster within the same intestinal segment or fecal sample. Because abundant unassigned taxa were detected that were significantly different in the duodenal, jejunal and ileal clusters, identification of overrepresented taxa was also implemented at the OTU level in these samples. ADONIS analyses were performed with 999 permutations for analysis of similarities and dissimilarities using the *vegan* package in **R**. Venn plots were generated for intestinal segment or feces samples at the OTU level using the VennDiagram package in **R**. The package *psych* in **R** was used to calculated Spearman correlations.

Acknowledgements

We thank Prof. Guiyun Xu for the anatomizing demonstration; Guangqi Li, Zhongyi Duan, Shanshan Xie, Jingwei Yuan, Dehe Wang, Zebin Zhang, Xingzheng Li, Yajie Li, Chunning Mai and Zhenfei Jiang for assistance with the sample collection; and Dr. Zhengsheng Xue for advice on the study. The current study was funded in part by Programs for Changjiang Scholars and Innovative Research in University (IRT_15R62) and China Agriculture Research System (CARS-40).

WY, JXZ and NY designed the study. WY, JXZ, CLW, CLJ, DXZ, YHC and CJS collected the samples. WY analyzed the data and wrote the manuscript. CLW assisted in construction of the figures. CJS and NY assisted in data analyzing and contributed to the revisions. All authors read and approved the final manuscript.

Conflict of Interest

The authors declare no conflicts of interest.

REFERENCES

1. Rosenbaum M, Knight R, and Leibel RL. 2015. The gut microbiota in human energy homeostasis and obesity. Trends in Endocrinology & Metabolism Tem 26:493.
2. Shin NR, Lee JC, Lee HY, Kim MS, Whon TW, Lee MS, and Bae JW. 2014. An increase in the Akkermansia spp. population induced by metformin treatment improves glucose homeostasis in diet-induced obese mice. Gut. 63:727-735.
3. Brisbin JT, Gong J, and Sharif S. 2008. Interactions between commensal bacteria and the gut-associated immune system of the chicken. Anim Health Res Rev 9:101-110.
4. Lavelle A, Lennon G, O'Sullivan O, Docherty N, Balfe A, Maguire A, Mulcahy HE, Doherty G, O'Donoghue D, Hyland J, Ross RP, Coffey JC, Sheahan K, Cotter PD, Shanahan F, Winter DC, and O'Connell PR. 2015. Spatial variation of the colonic microbiota in patients with ulcerative colitis and control volunteers. Gut 64:1553-61.
5. Gevers D, Kugathasan S, Denson LA, Vázquezbaeza Y, Van TW, Ren B, Schwager E, Knights D, Song SJ, and Yassour M. 2014. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe 15:382-392.
6. Dumonceaux TJ, Hill JE, Hemmingsen SM, and Van Kessel AG. 2006. Characterization of intestinal microbiota and response to dietary virginiamycin supplementation in the broiler chicken. Appl Environ Microbiol 72:2815-2823.
7. Yeoman CJ, Chia N, Jeraldo P, Sipos M, Goldenfeld ND, and White BA. 2012. The microbiome of the chicken gastrointestinal tract. Anim Health Res Rev 13:89-99.
8. Stanley D, Hughes RJ, and Moore RJ. 2014. Microbiota of the chicken gastrointestinal tract: influence on health, productivity and disease. Appl Microbiol Biotechnol 98:4301-4310.

9. Shaufi MAM, Sieo CC, Chong CW, Gan HM, and Ho YW. 2015. Deciphering chicken gut microbial dynamics based on high-throughput 16S rRNA metagenomics analyses. Gut Pathog 7:4.

10. Clavijo V, and Mjv F. 2018. The gastrointestinal microbiome and its association with the control of pathogens in broiler chicken production: A review. Poultry Sci 97:1006-1021.

11. Gong J, Si W, Forster R, Huang R, Yu H, Yin Y, and Yang CY. 2007. 16S rRNA gene-based analysis of mucosa-associated bacterial community and phylogeny in the chicken gastrointestinal tracts: from crops to ceca. Fems Microbiol Ecol 59:147.

12. Choi JH, Kim GB, and Cha CJ. 2014. Spatial heterogeneity and stability of bacterial community in the gastrointestinal tracts of broiler chickens. Poult Sci 93:1942-1950.

13. Xiao Y, Xiang Y, Zhou W, Chen J, Li K, and Yang H. 2016. Microbial community mapping in intestinal tract of broiler chicken. Poult Sci 96:1387-1393.

14. Mead GC. 1989. Microbes of the avian cecum: types present and substrates utilized. Journal of Experimental Zoology Part A Ecological Genetics & Physiology 252:48-54.

15. Clench MH, and Mathias JR. 1995. The Avian Cecum: A Review. Wilson Bulletin 107:93-121.

16. Obst BS, and Diamond JM. 1989. Interspecific variation in sugar and amino acid transport by the avian cecum. Journal of Experimental Zoology Part A Ecological Genetics & Physiology 252:117-126.

17. Gasaway WC, White RG, and Dan FH. 1976. Digestion of Dry Matter and Absorption of Water in the Intestine and Cecum of Rock Ptarmigan. Condor 78:77-84.

18. Stanley D, Geier MS, Hughes RJ, Denman SE, and Moore RJ. 2013. Highly variable microbiota development in the chicken gastrointestinal tract. Plos One 8:e84290.

19. Sergeant MJ, Constantinidou C, Cogan TA, Bedford MR, Penn CW, and Pallen MJ. 2014.

Extensive Microbial and Functional Diversity within the Chicken Cecal Microbiome. Plos One 9:e91941.

20. Stanley D, Geier MS, Chen H, Hughes RJ, and Moore RJ. 2015. Comparison of fecal and cecal microbiotas reveals qualitative similarities but quantitative differences. Bmc Microbiol 15:1-11.

21. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J, MetaHIT C, Weissenbach J, Ehrlich SD, and Bork P. 2011. Enterotypes of the human gut microbiome. Nature 473:174-180.

22. Mi YL, Rho M, Song YM, Lee K, Sung J, and Ko GP. 2014. Stability of Gut Enterotypes in Korean Monozygotic Twins and Their Association with Biomarkers and Diet. Sci Rep-Uk 4:7348.

23. Liang C, Tseng HC, Chen HM, Wang WC, Chiu CM, Chang JY, Lu KY, Weng SL, Chang TH, and Chang CH. 2017. Diversity and enterotype in gut bacterial community of adults in Taiwan. Bmc Genomics 18:932.

24. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, and Lewis JD. 2011. Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. Science (New York, N.y.) 334:105-108.

25. Moeller AH, Degnan PH, Pusey AE, Wilson ML, Hahn BH, and Ochman H. 2012. Chimpanzees and humans harbour compositionally similar gut enterotypes. Nat Commun 3:1179.

26. Wang J, Linnenbrink M, Künzel S, Fernandes R, Nadeau M, Rosenstiel P, and Baines JF. 2014.

Dietary history contributes to enterotype-like clustering and functional metagenomic content in the intestinal microbiome of wild mice. Proceedings of the National Academy of Sciences 111:E2703.

27. Kaakoush NO, Sodhi N, Chenu JW, Cox JM, Riordan SM, and Mitchell HM. 2014. The interplay between Campylobacter and Helicobacter species and other gastrointestinal microbiota of commercial broiler chickens. Gut Pathog 6:18.

28. Suzuki TA, and Nachman MW. 2016. Spatial Heterogeneity of Gut Microbial Composition along the Gastrointestinal Tract in Natural Populations of House Mice. Plos One 11:e0163720.

29. Sekelja M, Rud I, Knutsen SH, Denstadli V, Westereng B, Næs T, and Rudi K. 2012. Abrupt temporal fluctuations in the chicken fecal microbiota are explained by its gastrointestinal origin. Applied & Environmental Microbiology 78:2941-2948.

30. Gu S, Chen D, Zhang JN, Lv X, Wang K, Duan LP, Nie Y, and Wu XL. 2013. Bacterial Community Mapping of the Mouse Gastrointestinal Tract. Plos One 8:e74957.

31. Yan W, Sun C, Yuan J, and Yang N. 2017. Gut metagenomic analysis reveals prominent roles of Lactobacillus and cecal microbiota in chicken feed efficiency. Sci Rep-Uk 7:45308.

32. Yasuda, Koji, Keunyoung, Ren, Boyu, Nbsp T, and Franzosa. 2015. Biogeography of the Intestinal Mucosal and Lumenal Microbiome in the Rhesus Macaque. Cell Host Microbe 17:385.

33. Stearns JC, Lynch MDJ, Senadheera DB, Tenenbaum HC, Goldberg MB, Cvitkovitch DG, Croitoru K, Morenohagelsieb G, and Neufeld JD. 2011. Bacterial biogeography of the human digestive tract. Sci Rep-Uk 1:170.

34. Li D, Chen H, Mao B, Yang Q, Zhao J, Gu Z, Zhang H, Chen YQ, and Chen W. 2017. Microbial Biogeography and Core Microbiota of the Rat Digestive Tract. Sci Rep-Uk 8:45840.

35. Zhao W, Wang Y, Liu S, Huang J, Zhai Z, He C, Ding J, Wang J, Wang H, Fan W, Zhao JG, and

Meng H. 2015. The dynamic distribution of porcine microbiota across different ages and gastrointestinal tract segments. Plos One 10:e0117441.

36. Ericsson AC, Johnson PJ, Lopes MA, Perry SC, and Lanter HR. 2016. A Microbiological Map of the Healthy Equine Gastrointestinal Tract. Plos One 11:e0166523.

37. Waite DW, and Taylor MW. 2014. Characterizing the avian gut microbiota: membership, driving influences, and potential function. Front Microbiol 5:223.

38. Wienemann T, Schmitt-Wagner D, Meuser K, Segelbacher G, Schink B, Brune A, and Berthold P. 2011. The bacterial microbiota in the ceca of Capercaillie (Tetrao urogallus) differs between wild and captive birds. Systematic & Applied Microbiology 34:542-551.

39. Matsui H, Kato Y, Chikaraishi T, Moritani M, Bantokuda T, and Wakita M. 2010. Microbial diversity in ostrich ceca as revealed by 16S ribosomal RNA gene clone library and detection of novel Fibrobacter species. Anaerobe 16:83.

40. Oakley BB, Lillehoj HS, Kogut MH, Kim WK, Maurer JJ, Pedroso A, Lee MD, Collett SR, Johnson TJ, and Cox NA. 2014. The chicken gastrointestinal microbiome. Fems Microbiol Lett 360:100-112.

41. Wei S, Morrison M, and Yu Z. 2013. Bacterial census of poultry intestinal microbiome. Poultry Sci 92:671-683.

42. Tillman GE, Haas GJ, Wise MG, Oakley B, Smith MA, and Siragusa GR. 2011. Chicken intestine microbiota following the administration of lupulone, a hop-based antimicrobial. Fems Microbiol Ecol 77:395-403.

43. Cressman MD, Yu Z, Nelson MC, Moeller SJ, Lilburn MS, and Zerby HN. 2010. Interrelations between the Microbiotas in the Litter and in the Intestines of Commercial Broiler Chickens.

Applied & Environmental Microbiology 76:6572-6582.

44. Clench MH, and Mathias JR. 1992. A complex avian intestinal motility response to fasting. American Journal of Physiology 262:498-504.

45. Costea PI, Hildebrand F, Manimozhiyan A, Bäckhed F, Blaser MJ, Bushman FD, De Vos WM, Ehrlich SD, Fraser CM, Hattori M, Huttenhower C, Jeffery IB, Knights D, Lewis JD, Ley RE, Ochman H, O'Toole PW, Quince C, Relman DA, Shanahan F, Sunagawa S, Wang J, Weinstock GM, Wu GD, Zeller G, Zhao L, Raes J, Knight R, and Bork P. 2018. Enterotypes in the landscape of gut microbial community composition. Nat Microbiol 3:8.

46. Ramayo-Caldas Y, Mach N, Lepage P, Levenez F, Denis C, Lemonnier G, Leplat JJ, Billon Y, Berri M, Doré J, Rogel-Gaillard C, and Estelle J. 2016. Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. Isme J 10:2973.

47. Pauwels J, Taminiau B, Janssens GPJ, Beenhouwer MD, Delhalle L, Daube G, and Coopman F. 2015. Cecal drop reflects the chickens' cecal microbiome, fecal drop does not. J Microbiol Meth 117:164.

48. Garcia-Mazcorro JF, Castillo-Carranza SA, Guard B, Gomez-Vazquez JP, Dowd SE, and Brigthsmith DJ. 2016. Comprehensive Molecular Characterization of Bacterial Communities in Feces of Pet Birds Using 16S Marker Sequencing. Microb Ecol:1-12.

49. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S, Harris HM, Coakley M, Lakshminarayanan B, and O'Sullivan O. 2012. Gut microbiota composition correlates with diet and health in the elderly. Nature 488:178-184.

50. Clarke SF, Murphy EF, O'Sullivan O, Lucey AJ, Humphreys M, Hogan A, Hayes P, O'Reilly M, Jeffery IB, and Woodmartin R. 2014. Exercise and associated dietary extremes impact on gut

microbial diversity. Gut 63:1913.

51. Clauss W, Dantzer V, and Skadhauge E. 1991. Aldosterone modulates electrogenic Cl secretion in the colon of the hen (Gallus domesticus). Am J Physiol 261:1533-41.

52. Smith CC, Snowberg LK, Gregory CJ, Knight R, and Bolnick DI. 2015. Dietary input of microbes and host genetic variation shape among-population differences in stickleback gut microbiota. Isme J 9:2515.

53. Zhao LL, Wang G, Siegel P, He C, Wang HZ, Zhao WJ, Zhai ZX, Tian FW, Zhao JX, Zhang H, Sun Z, Chen W, Zhang Y, and Meng H. 2013. Quantitative Genetic Background of the Host Influences Gut Microbiomes in Chickens. Sci Rep-Uk 3:1163.

54. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, and Gordon JI. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335.

55. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraserliggett CM, and Nelson KE. 2006. Metagenomic Analysis of the Human Distal Gut Microbiome. Science 312:1355-1359.

56. Chen H, and Jiang W. 2014. Application of high-throughput sequencing in understanding human oral microbiome related with health and disease. Front Microbiol 5:508.

57. Magoč T, and Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957-2963.

58. Desantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, and Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied & Environmental Microbiology 72:5069-72.

59. Lozupone C, and Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Applied & Environmental Microbiology 71:8228.

Table 1 Shared and exclusive OTUs between two of sampling sites

Site1	Site2	Shared OTUs		Exclusive OTUs	
		In site1, %	In site2, %	In site1, %	In site2, %
D[1]	J	95.03[2]	96.00	4.97	4.00
		99.95[3]	99.96	0.05	0.04
D	I	98.57	89.77	1.43	10.23
		99.90	99.80	0.10	0.20
J	I	98.78	89.05	1.22	10.95
		99.93	99.81	0.07	0.19
C	D	82.71	65.79	17.29	34.21
		97.16	43.13	2.84	56.87
C	J	81.38	65.40	18.62	34.60
		96.27	48.30	3.73	51.70
C	I	95.31	69.05	4.69	30.95
		99.58	59.14	0.42	40.86
F	D	86.81	94.20	13.19	5.80
		99.51	98.78	0.49	1.22
F	J	85.83	94.10	14.17	5.90
		99.49	99.05	0.51	0.95

		95.94	94.82	4.06	5.18
F	I				
		99.65	98.84	0.35	1.16
		72.99	99.57	27.01	0.43
F	C				
		99.12	>99.99	0.88	<0.01

[1]D, J, I, C and F denoted the microbial community of duodenum, jejunum, ileum, cecum and feces, respectively. [2]The percentage of shared or exclusive OTUs; [3]The percentage of sequence reads shared or exclusive OTUs represented.

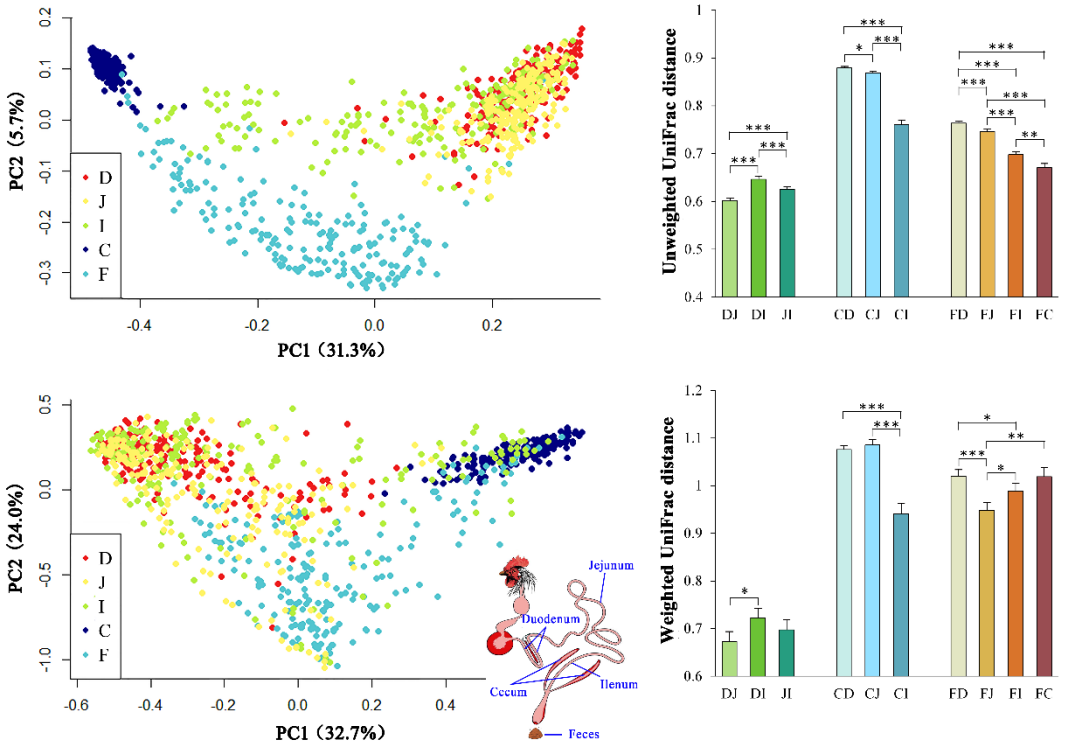

FIG 1 Site origin and inter-individual effects on the shape of microbial community membership and structure. **(a)** Principal coordinates analysis (PCoA) with unweighted UniFrac distance. Each dot represents a sample from a duodenum (D), jejunum (J), ileum (I), cecum (C) or feces (F). PC1 and PC2 represent the top two principal coordinates that captured the most variation, with the fraction of variation captured by that coordinate shown as a percent. **(b)** Unweighted UniFrac distance (mean ± SEM) between two sampling sites. DJ represents the UniFrac distance between the duodenal and jejunal microbial community, and it was the same as DI, JI, CD, CJ, CI, FD, FJ, FI and FC. Asterisks indicate the significance of the paired *t*-test: $^{***}P < 0.001$, $^{**}P < 0.01$, and $^{*}P < 0.05$. **(c)** PCoA plot with weighted UniFrac distance, similar to **(a)**. **(d)** Weighted UniFrac distance between two sampling sites, similar to (**b**).

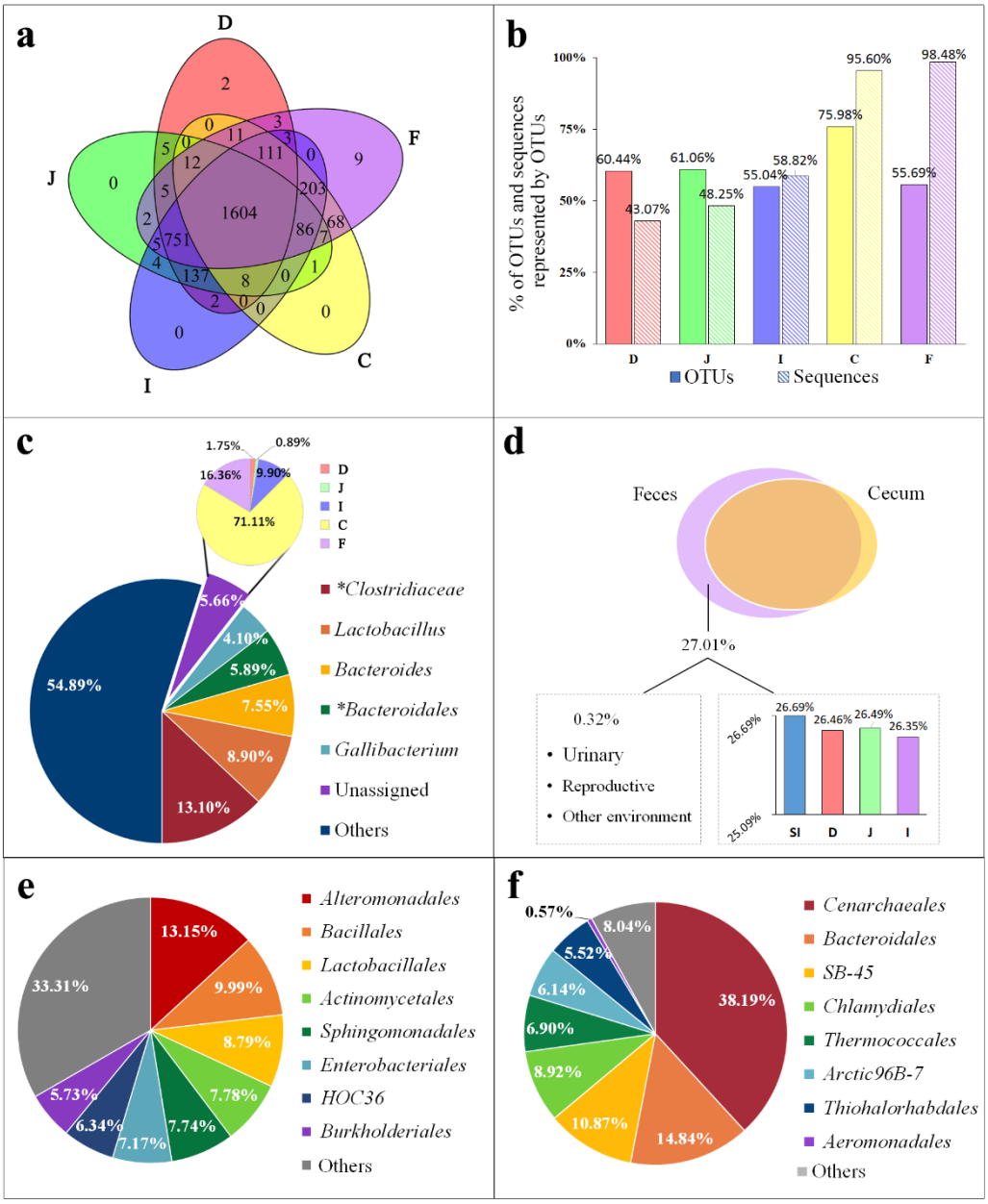

FIG 2 OTUs shared across different sampling sites. **(a)** Venn diagram demonstrating that the taxa overlap among different sampling sites. D, J, I, C and F denote the microbial communities of the duodenum, jejunum, ileum, cecum and feces, respectively. **(b)** The percentage of core OTUs and sequences represented by these OTUs in the duodenal, jejunal, ileal, cecal and fecal samples. **(c)**, related to **(b)**. The dominant taxa of the core microbiota. The sequences of unassigned taxa are partitioned to different sampling sites (duodenum, jejunum, ileum, ceca and feces) shown in the

small pie chart. **(d)** The percentage of OTUs in feces exclusively contributed by the small intestine, and the percentage of OTUs in feces was below the limit of detection in the gastrointestinal tract. SI = the microbial community in small intestine. Taxa exclusively shared between feces and small intestine **(e)** or ceca **(f)** are shown at the order level. Each section in the pies **(c, e** and **f)** indicate the percentage of sequences represented by the corresponding taxon in the core taxa **(c)** or exclusively shared taxa **(e** and **f)**.

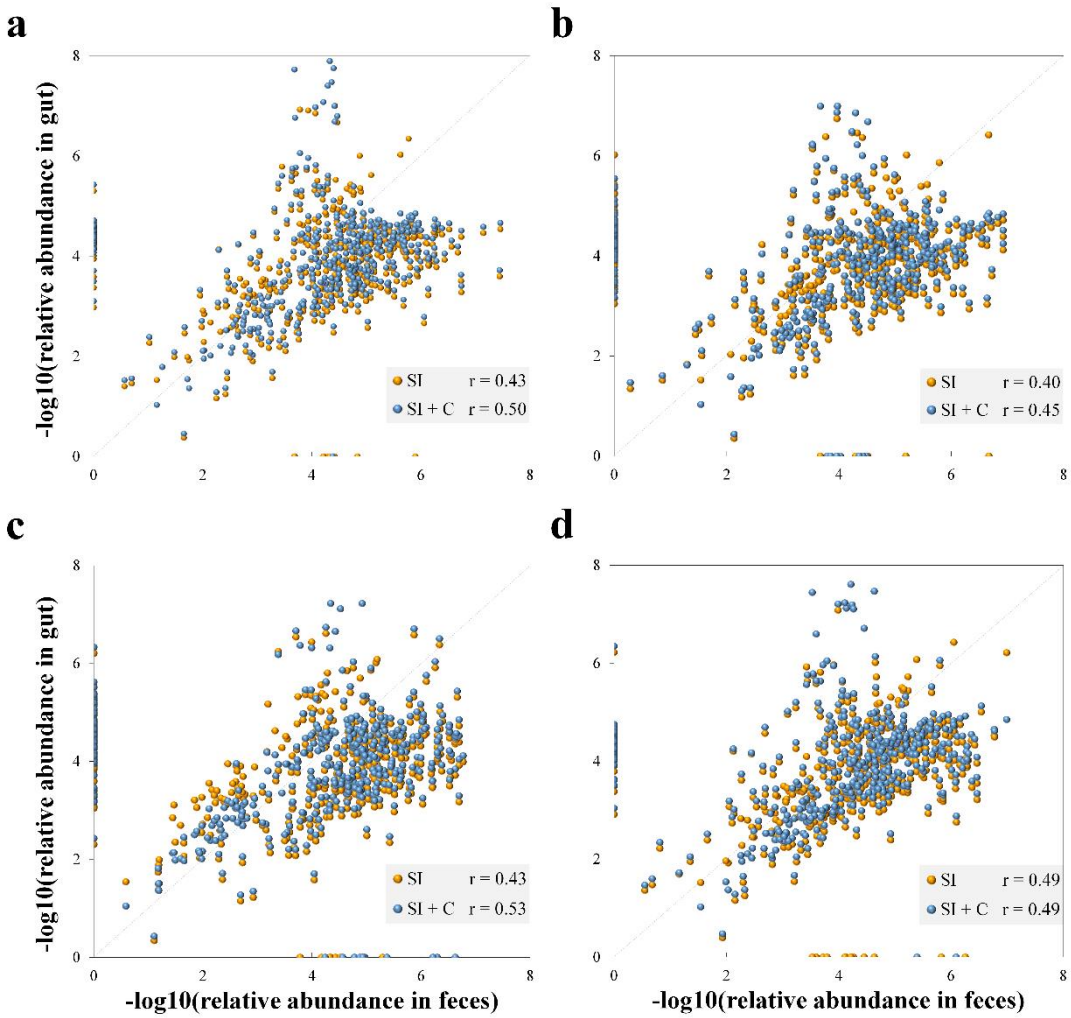

FIG 3 Microbial compositions in feces mirror those in the gastrointestinal tract. Each dot represents the average relative abundance of a genus across all individuals **(a)**, individuals in fecal enterotype-like cluster 1 **(b)**, individuals in fecal enterotype-like cluster 2 **(c)** and individuals in fecal enterotype-like cluster 3 **(d)** for each region (feces at the x-axis; SI: small intestine or SI + C: intestine including small intestine and ceca at the y-axis). Spearman's rho was calculated between fecal and SI (or SI + C) negative logarithm-transferred relative abundances. All Spearman correlations shown are significant ($P < 0.05$).

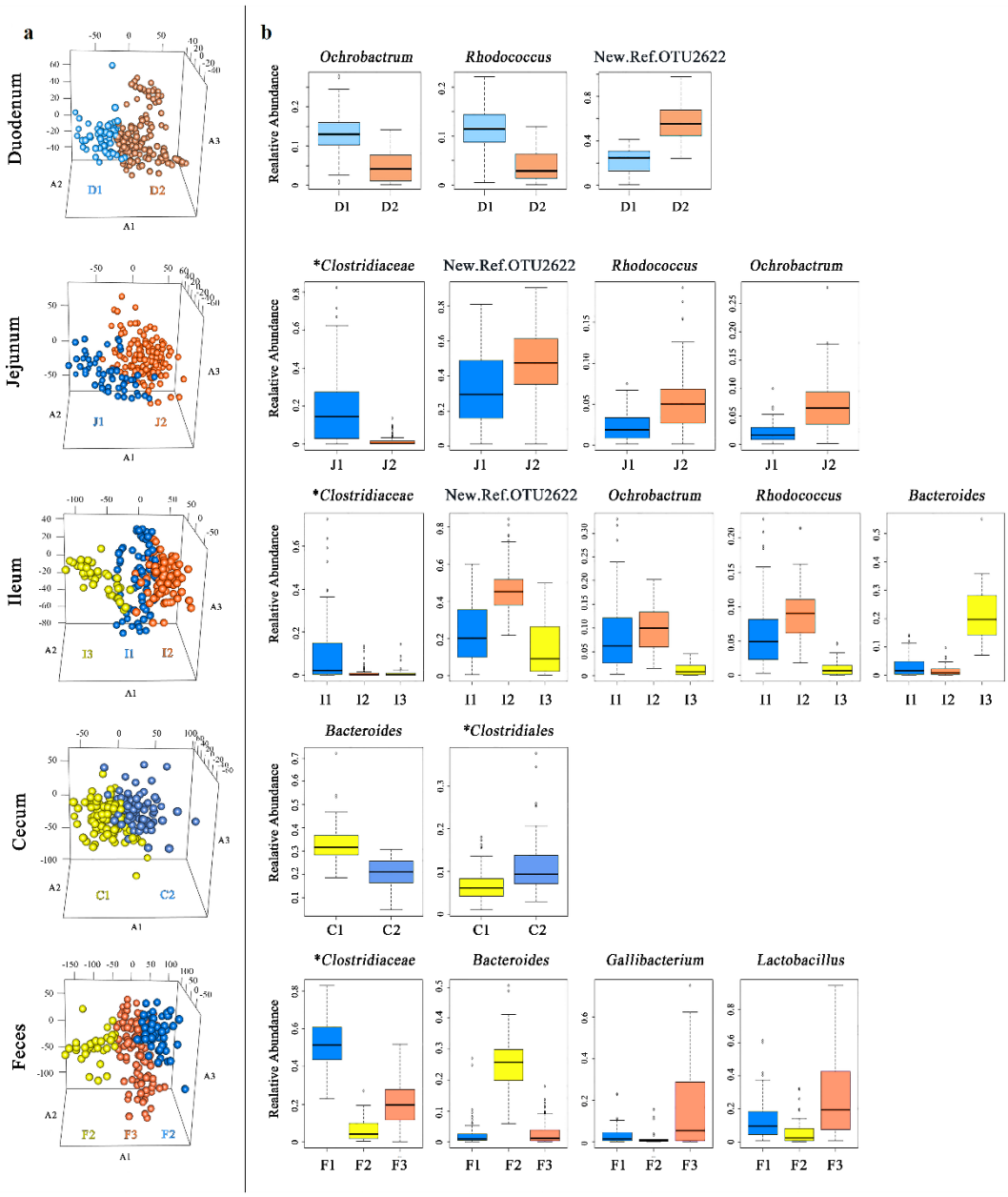

FIG 4 Enterotype-like clusters in different sites. **(a)** Analyses of enterotype-like clustering with the partitioning around medoid (PAM) clustering algorithm, which visualizes results from PCoA and clustering at the OTU level using a 97% similarity threshold. **(b)** Relative abundances of overrepresented taxa of each enterotype. D, J, I, C and F denote the microbial communities of the duodenum, jejunum, ileum, cecum and feces, respectively. The numbers following D, J, I, C and F represent enterotype-like cluster 1, 2 or 3. Unclassified genera under a higher rank are marked by asterisks.

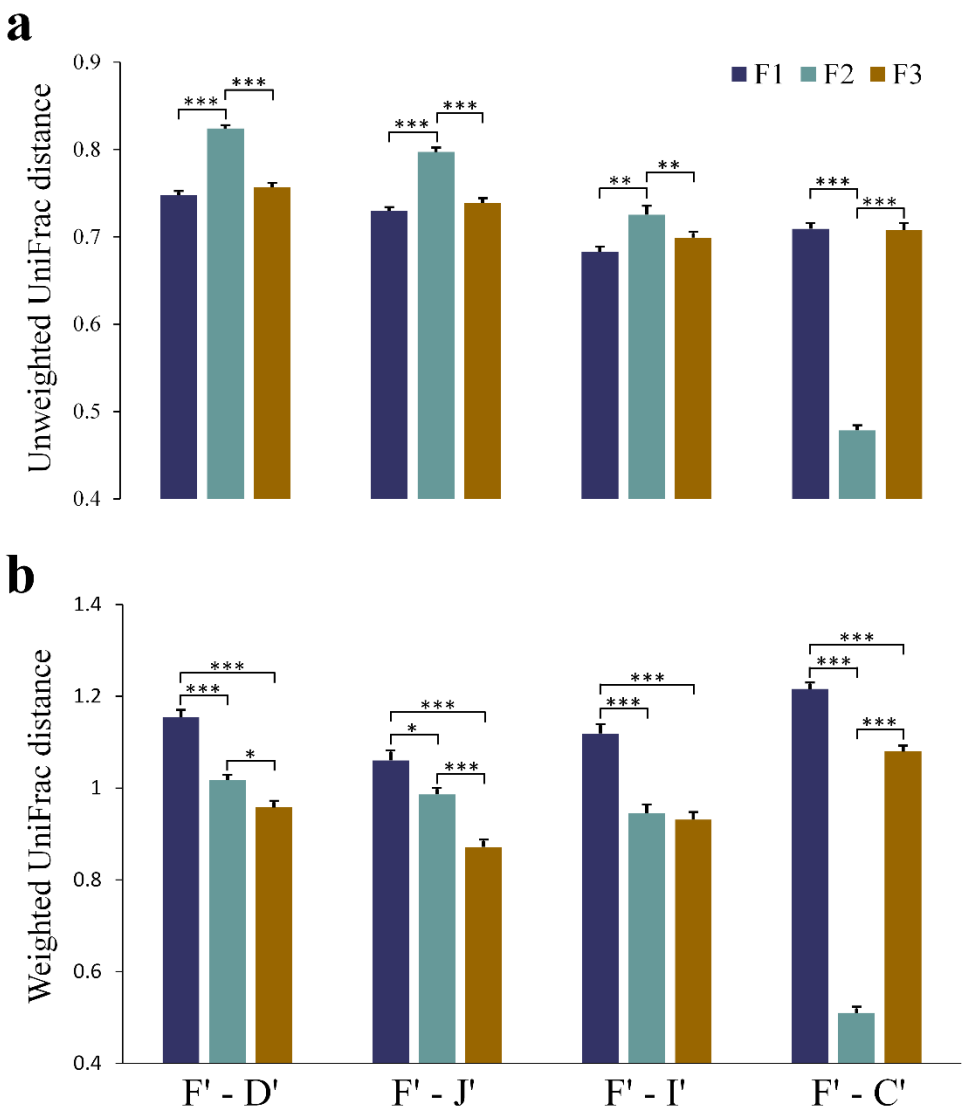

FIG 5 Enterotype-like clustering affects microbial relationships between fecal and gastrointestinal samples. Unweighted **(a)** and weighted **(b)** UniFrac distance (mean ± SEM) between fecal and each of the four segmental samples in different fecal enterotype-like clusters. F1, F2 and F3 denote fecal enterotype-like clusters 1, 2 and 3. D', J', I', C' and F' represent the duodenal, jejunal, ileal, cecal and fecal samples, respectively, in corresponding fecal enterotype-like clusters. Asterisks indicate the significance of the paired *t*-test: $^{***}P < 0.001$, $^{**}P < 0.01$, and $^{*}P < 0.05$.