

ORIGINAL ARTICLE

Analysis of the vaginal microbiome of giant pandas using metagenomics sequencing

Lan Zhang¹ | Caiwu Li² | Yaru Zhai¹ | Lan Feng¹ | Keke Bai¹ | Zhizhong Zhang² | Yan Huang² | Ti Li² | Desheng Li² | Hao Li¹ | Pengfei Cui¹ | Danyu Chen¹ | Hongning Wang¹ | Xin Yang¹ 

¹Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, Animal Disease Prevention and Food Safety Key Laboratory of Sichuan Province, College of Life Sciences, Sichuan University, Chengdu, PR China

²China Conservation and Research Center for the Giant Panda, Key Laboratory of State Forestry and Grassland Administration on Conservation Biology of Rare Animals in the Giant Panda National Park, Qionglai Mountains Conservation Biology of Endangered Wild Animals and Plants National Permanent Scientific Research Base, Dujiangyan, PR China

Correspondence

Xin Yang, Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, Animal Disease Prevention and Food Safety Key Laboratory of Sichuan Province, College of Life Sciences, Sichuan University, Chengdu, 610065, Sichuan, PR China.

Email: yangxin0822@163.com

Funding information

Panda International Cooperation Fund, Grant/Award Number: GH201714

Abstract

In this study, a total of 14 vaginal samples (GPV1-14) from giant pandas were analyzed. These vaginal samples were divided into two groups as per the region and age of giant pandas. All the vaginal samples were analyzed using metagenomic sequencing. As per the outcomes of metagenomic analysis, Proteobacteria (39.04%), Firmicutes (5.27%), Actinobacteria (2.94%), and Basidiomycota (2.77%) were found to be the dominant phyla in the microbiome of the vaginal samples. At the genus level, *Pseudomonas* (21.90%) was found to be the most dominant genus, followed by *Streptococcus* (3.47%), *Psychrobacter* (1.89%), and *Proteus* (1.38%). Metastats analysis of the microbial species in the vaginal samples of giant pandas from Wolong Nature Reserve, Dujiangyan and Ningbo Youngor Zoo, and Ya'an Bifengxia Nature Reserve was found to be significantly different ($p < 0.05$). Age groups, that is, AGE1 (5-10 years old) and AGE2 (11-16 years old), also demonstrated significantly different inter-group microbial species ($p < 0.05$). For the first time, *Chlamydia* and *Neisseria gonorrhoeae* were detected in giant pandas' reproductive tract. GPV3 vaginal sample (2.63%) showed highest *Chlamydia* content followed by GPV14 (0.91%), and GPV7 (0.62%). GPV5 vaginal sample (7.17%) showed the highest *Neisseria gonorrhoeae* content, followed by GPV14 (7.02%), and GPV8 (6.50%). Furthermore, we employed eggNOG, CAZy, KEGG, and NCBI databases to investigate the functional significance of giant panda's vaginal microbial community. The outcomes indicated that giant panda's vaginal microbes were involved in biological processes. The data from this study will help in improving the reproductive health of giant pandas.

KEYWORDS

giant panda, metagenomic sequencing, vaginal microbiome

Abbreviations: CAZy, Carbohydrate-Active enzymes Database; eggNOG, evolutionary genealogy of genes: Non-supervised Orthologous Groups; GPV, giant panda vagina; KEGG, kyoto encyclopedia of genes and genomes.

Lan Zhang and Caiwu Li contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *MicrobiologyOpen* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

The giant panda is a national treasure of China (Peng et al., (2001)). Giant pandas are endangered species with low fertility rates and on the brink of extinction. Their natural habitat occurs in small patches, which is spread across an area of 12,000 km² Hu, (1997); Peng et al., 2007). Increasing the reproductive rate of the giant panda is a crucial part of current conservation strategies (Hu, 1997).

Animals harbor diverse microorganisms in large numbers. Thus, the microbial investigation of the reproductive organs becomes crucial (Yan et al., 2019). As demonstrated in previous studies, microbiota affects the reproductive system either directly or indirectly (Bradford & Ravel, 2017; Ma et al., 2017; Moraes et al., 2014; Pal, 2015; Stumpf et al., 2013). Imbalanced vaginal microbiota and vaginal infection have been directly correlated to infertility as per the previous studies (Li et al., 2019). For instance, fungal infection of the reproductive system, such as human vulvovaginitis (Hedayati et al., 2015), cervicitis (Dascanio et al., 2010), and cow's clinical vaginitis (Al-Fatlawy, 2014), severely affects fertility. On par with the pigs (De Puyseleer et al., 2014; Schautteet & Vanrompay, 2011), cattle (Knudtson & Kirkbride, 1992), and horse's (Cafarchia et al., 2013) vaginal microbiome investigations, the vaginal microbiome of the giant panda was also investigated. These studies provided a comprehensive description of bacterial community composition in the vagina, uterus, and fungal community composition in the vagina of healthy giant pandas along with the diversity and abundance of these microbes (Chen et al., 2018; Yang et al., 2016, 2017).

The majority of the vaginal microbiome studies carried out so far have adapted either the conventional method of microbial isolation, culture, and identification or high-throughput sequencing. However, these methods could not elucidate the giant panda's complete vaginal microbiome. A majority of vaginal microorganisms have strict growth requirements, which makes their *in vitro* culture a challenging task. Vaginal microbes interact and restrict each other's growth, and thus protect reproductive efficiency in giant pandas. To date, limited studies have investigated the genitourinary tract of giant pandas. Due to its high efficiency and accuracy, in the current study, the metagenomic sequencing method was employed to investigate the giant panda's vaginal microbiome and functional characterization of the associated genes.

Compared with the 16S amplicon sequencing method, metagenomic sequencing is widely used to examine the microbiome, gene function, gene activity, and cooperative relationship between the microorganisms and the correlation between microorganisms and the environment. In this study, a comprehensive analysis of the reproductive tract microbiome from giant pandas of different ages residing in different locations was conducted using metagenomic sequencing (Doonan et al., 2016). In the giant pandas' vaginal samples, we examined microbial interaction and the effect of age and region on this interaction. Furthermore, in this study, the metagenomic sequencing method was employed to examine the vaginal microbiome of giant pandas for the first time. The data from this study can be utilized to improve the reproductive health of giant pandas in the future.

2 | MATERIALS AND METHODS

2.1 | Samples

A total of 14 vaginal samples from giant pandas (GPV 1-14) were collected for which 14 giant pandas were sampled from four different geographical locations in China, that is, Wolong Nature Reserve, Ya'an Bifengxia Nature Reserve, China Giant Panda Conservation Research Center Dujiangyan Base, and Ningbo Youngor Zoo. Long sterile swabs were used to collect the vaginal secretion of 5-23 years old giant panda *in estrous* during the period of June-August, 2018. Subsequently, each swab was immediately dipped in 2 ml phosphate-buffered saline (PBS) in sterile 5 ml screw-cap tubes (Yang et al., 2017). These samples were transported to the Animal Disease Prevention and Food Safety Key Laboratory of Sichuan Province and stored in freezers at -80°C until the DNA extraction procedure (Yang et al., 2016, 2017). The 14 vaginal samples taken from these giant pandas were divided into two groups based on age and region (Yang et al., 2016, 2017).

In the age group, 14 vaginal samples were categorized as AGE1, AGE2, AGE3. AGE1 contained seven vaginal samples from 5-10 years old giant pandas. These samples were marked as GPV1, GPV2, GPV3, GPV4, GPV5, GPV6, and GPV7. AGE2 contained five vaginal samples from 11 to 16 years old giant pandas, which were marked as GPV8, GPV9, GPV10, GPV11, and GPV12. AGE3

Basis of grouping	Group name	Sample serial number
5-10 years old	AGE1	GPV1-GPV7
11-16 years old	AGE2	GPV8-GPV12
17-23 years old	AGE3	GPV13-GPV14
Wolong Nature Reserve	WL	GPV1, GPV2, GPV8, GPV13, GPV14
Dujiangyan Base and Ningbo Youngor Zoo	DN	GPV3, GPV6, GPV7
Ya'an Bifengxia Nature Reserve	YA	GPV4, GPV5, GPV9, GPV10, GPV11, GPV12

TABLE 1 The sample group information.

contained two vaginal samples from 17-23 years old giant pandas, which were marked as GPV13 and GPV14.

In the region group, 14 vaginal samples were categorized as WL, YA, DN. WL contained five vaginal samples from Wolong region giant pandas, and these samples were marked as GPV1, GPV2, GPV8, GPV13, and GPV14. YA contained six vaginal samples from the Ya'an Bifengxia region's giant pandas, and these samples were marked as GPV4, GPV5, GPV9, GPV10, GPV11, and GPV12. DN contained three giant panda's vaginal samples from Duijiangyan Base and Ningbo Youngor Zoo and marked as GPV3, GPV6, and GPV7 (Table 1). As per the protection centers, none of the sampled giant pandas had any reproductive problems.

2.2 | DNA extraction and metagenomic sequencing

The whole-genome DNA was extracted from the vaginal samples using Universal Genomic DNA Kit (Kang for the century Biotechnology Co., Ltd.), as per the manufacturer's instruction. The microbial cells were lysed with 180 μ L of the enzymatic buffer. DNA purity and concentration were determined using 1% agarose gel electrophoresis before being sent to the Novogene Bioinformatics Technology Co. Ltd. The metagenomic sequencing of these DNA samples was performed by the quay Co. Ltd (Chen & Pachter, 2005; Yang et al., 2016, 2017) on the Illumina HiSeq sequencing platform and dual-end sequencing (paired-end).

2.3 | Analysis of samples

The metagenomic sequencing data analysis was primarily focused on gene prediction and abundance analysis, specimen annotation, dimensionality analysis of species abundance, LEfSe analysis of differential species between groups, cluster analysis of species abundance, and Metastats analysis of different species between groups (Avershina et al., 2013; Oh et al., 2014; Rivas et al., 2013; Segata et al., 2011; White et al., 2009).

2.4 | Metagenomic sequencing

Metagenomic analysis was used to unravel the microbial composition and microbial interaction between the 14 vaginal samples. Besides, the metagenomic sequencing data were analyzed to identify the metabolic pathways and gene function at the molecular level. The metagenomic sequencing was performed using the following steps:

2.4.1 | Library construction

1 μ g of DNA was taken per sample as input material. Sequencing libraries were generated using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB), as per the manufacturer's

recommendations, and index codes were added to attribute sequences to each sample. Briefly, DNA samples were fragmented by sonication to obtain DNA fragments of 350 bp size, and these DNA fragments were end-polished, A-tailed, and ligated with full-length adaptor for Illumina sequencing and PCR amplified. The PCR products were purified (AMPure XP system), and DNA libraries were analyzed for size distribution using the Agilent 2100 Bioanalyzer system and quantified using real-time PCR.

2.4.2 | Pretreatment of sequencing results

To obtain clean data for subsequent analysis, the raw data from the Illumina HiSeq sequencing platform were preprocessed using Readfq version 8 (<https://github.com/cjfields/readfq>) as follows: (a) reads with low-quality bases (default quality threshold value ≤ 38) above 40 bp length were removed; (b) reads containing a specific percentage of N base were removed (default length of 10 bp); (c) reads overlapping with the adapter (default length of 15 bp) were also removed.

2.4.3 | Metagenome assembly

The clean metagenomic data obtained from the non-complex environment of the giant panda's vaginal contents were assembled and analyzed (Luo et al., 2012) using SOAPdenovo software version 2.04 (<http://soap.genomics.org.cn/soapdenovo.html>; Brum et al., 2015; Feng et al., 2015; Qin et al., 2014; Scher et al., 2013). The parameters used for this analysis were as follows: -d 1, -M 3, -R, -u, -F, -K 55. The assembled scaffolds were interrupted from the N connection, and the scaffolds without N were removed (Mende et al., 2012; Nielsen et al., 2014; Qin et al., 2014) (<http://bowtiebio.sourceforge.net/bowtie2/index.shtml>).

2.4.4 | Gene prediction and abundance analysis

The scaffolds (≥ 500 bp) from each sample were assembled, and the ORFs were predicted using MetaGeneMark version 2.10 (<http://topaz.gatech.edu/GeneMark/>) software and reads shorter than 100 nt () were filtered from the final result using default parameters.

For the predicted ORF, CD-HIT (Fu et al., 2012; Li & Godzik, 2006) software version 4.5.8 (<http://www.bioinformatics.org/cd-hit>) was used to reduce redundancy and obtain the unique initial gene catalog (the genes here refers to the nucleotide sequences coded by unique and continuous genes), using the following parameters: -c 0.95, -G 0, -aS 0.9, -g 1, -d 0 (Sunagawa et al.,; Zeller et al., 2014).

The clean data of each sample were mapped to the initial gene catalog using Bowtie version 2.2.4 (<http://bowtiebio.sourceforge.net/bowtie2/index.shtml>) to obtain the number of reads mapped to the genes identified in each sample, with the following parameter setting:

-- end-to-end, --sensitive, -l 200, -X 400. The genes with ≤ 2 numbers of reads were filtered (Qin et al., 2012, 2014) for each sample to obtain the gene catalog (unigenes), which was used for subsequent analysis. The basic information: statistic, core-pan gene analysis, correlation analysis of samples, and Venn figure analysis of genes were based on the abundance of each gene in each sample in the gene catalog (Cotillard et al., 2013; Le Chatelier et al., 2013; Villar et al., 2013).

2.4.5 | Taxonomy prediction

The DIAMOND (Buchfink et al., 2015) software version 0.9.9 (<https://github.com/bbuchfink/diamond/>) was used to blast the unigenes to the sequences of bacteria, fungi, archaea, and viruses fetched from the NR database version 2018-01-02 (<https://www.ncbi.nlm.nih.gov/>) of NCBI with the parameter setting: `blastp -e 1e-5`. Since each sequence may have multiple aligned results, for the final aligned results of each sequence, the cutoff was set to e-value \leq the smallest e-value $\times 10$ to take the LCA algorithm, which was applied to the systematic classification of MEGAN (Huson et al., 2011) software to ensure the species annotation information of sequences (Oh et al., 2014).

To obtain the table containing the gene number and abundance of each sample at each taxonomical level, taxonomical hierarchy (kingdom, phylum, class, order, family, genus, and species) for each sample was deduced based on the LCA annotation result and the gene abundance table. The species abundance in one sample equals the sum of the gene abundance annotated for the species. The gene number of a species in a sample equals the number of genes whose abundance is nonzero.

For Krona analysis, the exhibition of generation situation of relative abundance, the exhibition of abundance cluster heat map, PCA (Avershina et al., 2013; R `ade4` package version 2.15.3), and NMDS (Rivas et al., 2013; R `vegan` package version 2.15.3) decrease-dimension analysis were based on the abundance table of each taxonomical hierarchy. The difference between groups was tested using ANOSIM (R `vegan` package version 2.15.3). Different species between groups were searched using Metastats and LEfSe analyses. P-value was determined for each taxon using the permutation test between groups in the Metastats analysis. Furthermore, Benjamini and Hochberg False Discovery Rate were used to correct the p-values and acquire the q-values (White et al., 2009). LEfSe analysis was conducted using the LEfSe software (the default LDA score was 3; Segata et al., 2011). Lastly, random forest (RandoForest; R `pROC` and `randomForest` packages (Breiman, 2001) version 2.15.3) software was used to construct a random forest model. Significant species were screened using `MeanDecreaseAccuracy` and `MeanDecreaseGini`, and each model was cross-validated (default 10 times), and the ROC curves were plotted.

2.4.6 | Common functional database annotation

The DIAMOND software version 0.9.9 was employed to blast the unigenes to the functional database with the parameter setting of

`blastp, -e 1e-5` (Feng et al., 2015; Qin et al., 2014). The functional databases used in this analysis were as follows: KEGG (Kanehisa et al., 2006, 2014) version 2018-01-01 (<http://www.kegg.jp/kegg/>), eggNOG (Powell et al.) version 4.5 (<http://eggnogdb.embl.de/#/app/home>), and CAZy (Cantarel et al., 2009) version 201801 (<http://www.cazy.org/>). For each sequence's blast result, the best blast hit was used for subsequent analysis (Bäckhed et al., 2015; Feng et al., 2015; Qin et al., 2014).

For the statistic of the relative abundance of different functional hierarchy, the relative abundance of each functional hierarchy equals the sum of relative abundance annotated to that functional level. As per the outcome of the functional annotation and gene abundance table, the gene number table of each sample in each taxonomic hierarchy was obtained. The gene number of a function in a sample equals the gene number that was annotated to this function, and the abundance is nonzero.

Based on the abundance table of each taxonomy hierarchy, the counting of annotated gene numbers, the exhibition of the general relative abundance situation, the exhibition of abundance cluster heat map, and the decrease-dimension analysis of PCA and NMDS were conducted. Similarly, ANOSIM test of the difference between groups (inside) based on the functional abundance, comparative analysis of metabolic pathways, the Metastats and LEfSe analyses of the functional difference between groups were also performed.

3 | RESULTS

3.1 | Sequencing data

The sequencing data of giant panda's vaginal samples in this study were acquired through the Illumina HiSeq sequencing platform (Odintsova et al., 2017). 94,662.28 Mbp of raw reads was obtained from the metagenomic sequencing, and the average sequencing data volume was 6,761.59 Mbp. The total and average data after quality control were 94,455.81 Mbp and 6,746.84 Mbp, respectively. The effective data rate of quality control was 99.78% (Table A1: <https://doi.org/10.5281/zenodo.4067664>). A total of 537,000,891 bp scaffolds were assembled with an average length of 1,727.76 bp, the maximum length of 898,339 bp, N50 of 4,454.20 bp, and N90 of 697,60 bp. The scaffolds were interrupted from N to generate scaftigs. It yielded a total of 477,921,089 bp scaftigs. The average length of scaftigs was 1,610 bp. The N50 was 3,830 bp, and the N90 was 701 bp (Table A2: <https://doi.org/10.5281/zenodo.4067664>).

3.2 | Gene prediction and abundance analysis

A total of 74,805 genes with an average length of 566.80 bp and 54.96% GC content, accounting for 18.14% of all non-redundant genes, were identified. The three vaginal samples, GPV1, GPV5, and

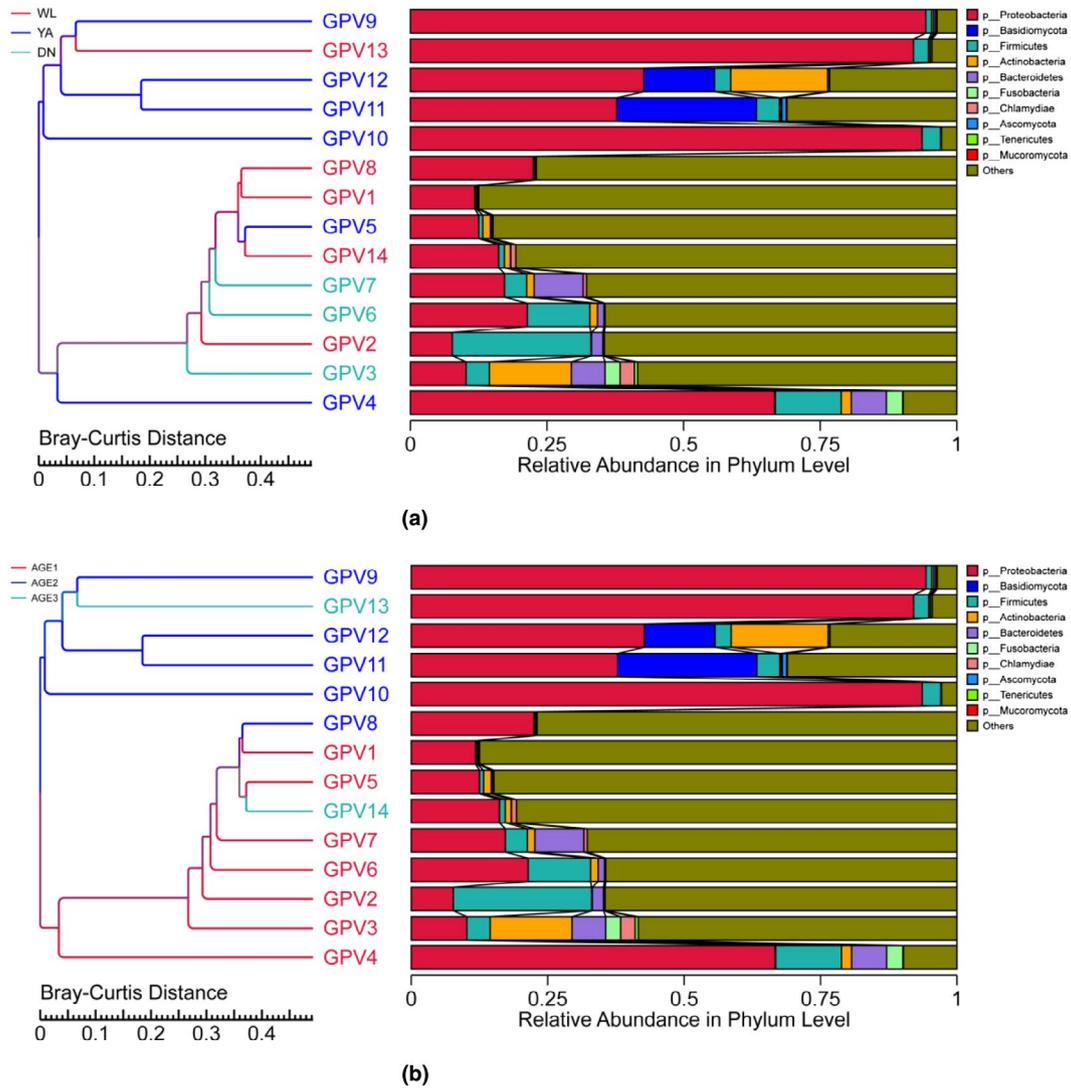


FIGURE 1 Clustering tree based on Bray–Curtis distance. (a) Clustering tree based on Bray–Curtis distance in the regional groups. (b) Clustering tree based on Bray–Curtis distance in the age groups.

GPV8, did not show any unique genes. However, GPV12 showed the highest number of unique genes. Out of the three regional groups of giant pandas, the highest number of unique genes appeared in the vaginal samples from the YA group. Out of the three age groups of giant pandas, the highest number of unique genes appeared in the vaginal samples from the AGE2 group.

3.3 | Species annotation

3.3.1 | Specimen annotation overview

A total of 412,312 predicted genes were obtained after the original de-redundancy of the 14 vaginal samples. 293,671 (71.23%) ORFs were annotated using the NR database. A total of 33 phyla, 78 classes, 184 orders, 383 families, and 1000 genera were identified from the all giant pandas' vaginal samples. The top 10 species with the highest relative abundance in each sample (group) were

selected from the relative abundance table of different classification levels. The remaining species were designated as "Others." Figure 1 demonstrates the relative abundance of the top 10 microorganisms at the phylum level. Proteobacteria, Firmicutes, Actinobacteria, and Basidiomycota, which accounted for 39.04%, 5.77%, 2.94%, and 2.77%, respectively, were identified as the dominant phyla (Figure 1). However, as demonstrated in Figure 1, different groups showed different dominant phylum. Phyla with more than 1% relative abundance were defined as the dominant phyla.

Proteobacteria and Firmicutes were identified as the dominant phyla in the WL group, whereas Proteobacteria, Basidiomycetes, Firmicutes, Actinomycetes, and Bacteroides were identified as the dominant phyla in the YA group. The dominant phyla in the DN group were Proteobacteria, Firmicutes, Actinomycetes, Bacteroides, and Chlamydiae. In the AGE1 group, Proteobacteria, Firmicutes, Actinomycetes, and Bacteroidetes were identified as the dominant phylum. However, the AGE2 group showed Proteobacteria, Basidiomycota, Firmicutes, and Actinomycetes as the dominant

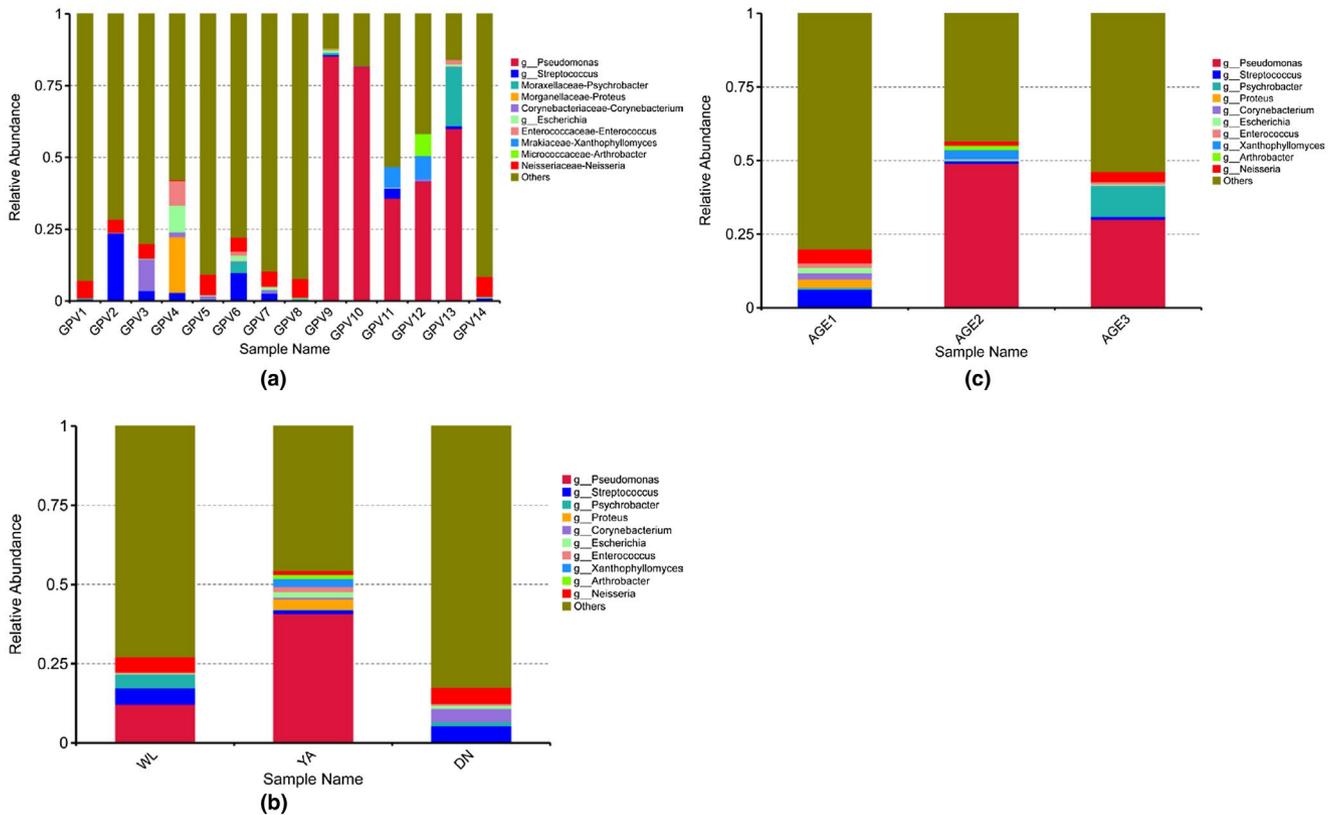


FIGURE 2 The relative abundance of features based on metagenome sequencing reads on the genus level. (a) Genus level relative abundance in all the samples. (b) Genus level relative abundance in regional groups. (c) Genus level relative abundance in age groups.

phyla, and AGE3 showed Proteobacteria and Firmicutes as dominant phyla. Basidiomycota phylum was identified as a dominant phylum only in the YA and AGE2 groups, but it was not identified in any of the other groups. Interestingly, the relative abundance of *Chlamydia* was higher in the DN group only.

Out of the total 1,000 genera identified in the giant panda's vaginal samples, the top 10 genera with the highest relative abundance are depicted in Figure 2. In all the fourteen samples, *Pseudomonas* was found to be the most abundant genus (21.90%), followed by *Streptococcus* (3.47%), *Psychrobacter* (1.89%), and *Proteus* (1.38%; Figure 2a). Dominant genera did not differ between the region and age group. However, out of the top ten dominant microorganisms with relatively highest abundance, *Neisseria* was found to be one of the most dominant genera (Figure 2b, c).

The top 35 genera were selected from the relative abundance table of different classification levels, and a heat map was constructed based on the abundance of each sample (Figure 3). The three groups, WL, DN, and YA, showed significant differences at the genus level (Figure 3a). The *Mycoplasma* in the Tenericutes phylum was mainly from the DN group. The abundant genus identified in the YA group belonged to phyla Proteobacteria, Basidiomycetes, and Actinomycetes. However, the proportion of *Fusobacterium* in phylum Fusobacteria was identical in the YA and DN groups. The abundant genera identified in the WL group were primarily from phylum Firmicutes. In the AGE group, we observed significant

differences at the genus level between each group (Figure 3b). The *Mycoplasma* in the Tenericutes phylum was primarily contributed by the AGE1 group, whereas the abundant genera in the AGE3 group belonged to the Proteobacteria phylum. Six of the thirteen abundant genera in the AGE2 group belonged to the Basidiomycota phylum.

The outcomes of the analysis demonstrated that the average relative abundance of *Chlamydia*, which is the most common cause of vaginal infection, was less than 0.44% in giant panda's vaginal samples. The highest content of *Chlamydia* was found in GPV3 (2.63%), followed by GPV14 (0.90%) and GPV7 (0.62%).

3.3.2 | Dimensionality analysis of species abundance

The WL and DN groups showed identical microbial composition within the group and also between the groups irrespective of the PCA or NMDS analysis (Figure 4a, b). The YA group showed relatively more diverse microbial composition within the group. The microbial composition of the AGE group is depicted in Figure 4c and 4d. As per the PCA or NMDS analysis, the majority of the samples in AGE1 and AGE3 groups showed a highly similar microbial composition. However, each sample in the AGE2 group showed significantly different microbial composition.

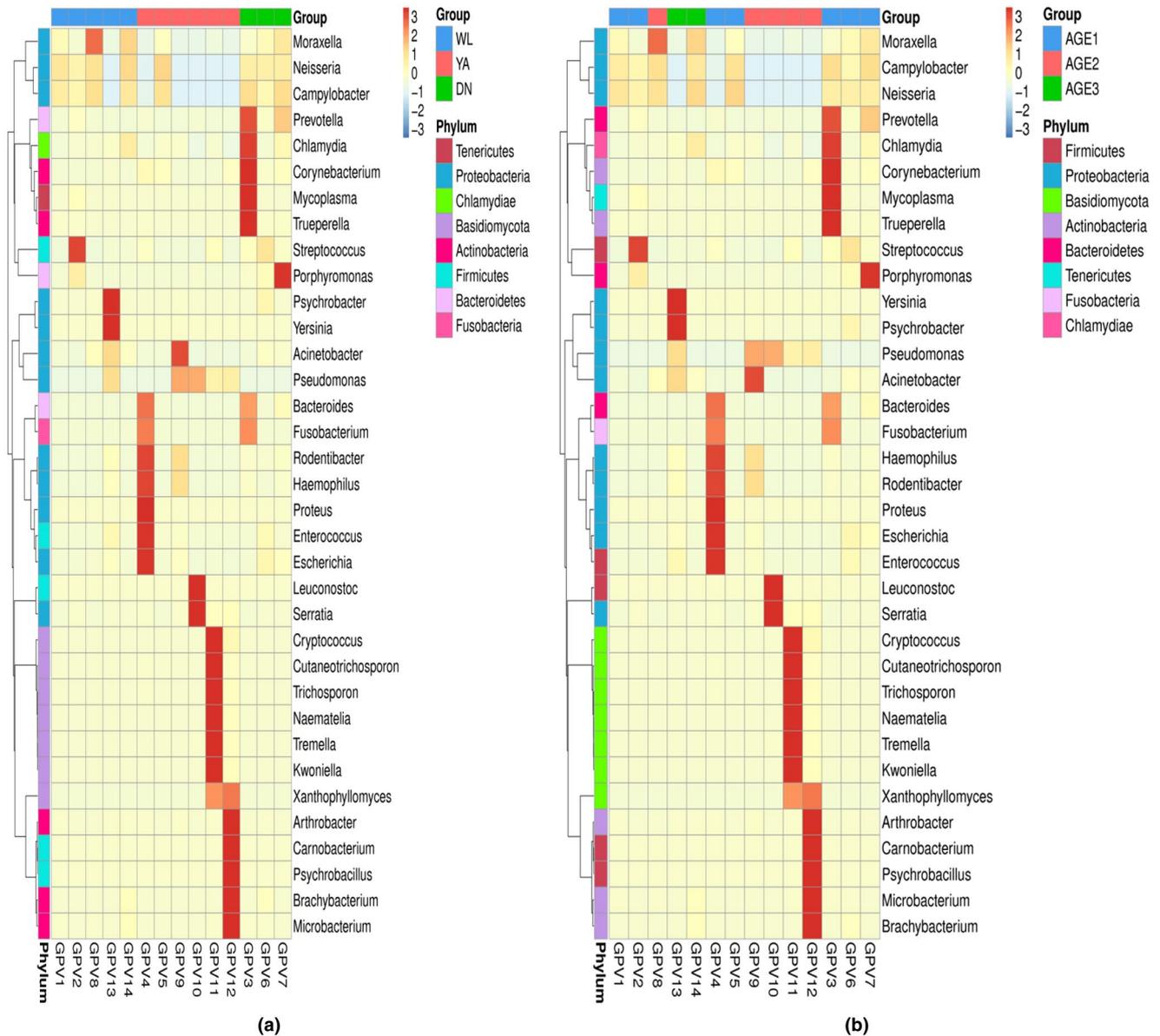


FIGURE 3 Clustering heat map based on the relative abundance of features based on metagenome sequencing at the genus level. (a) Genus level relative abundance clustering heat map in regional groups. (b) Genus level relative abundance clustering heat map in age groups.

3.3.3 | Cluster analysis of species abundance and Metastats analysis of differential species between groups

The three regional groups (WL, DN, and YA; Figure 1a) and three age groups (AGE1, AGE2, and AGE3; Figure 1b) were clustered separately. The giant pandas sampled in this study were all fed at the Giant Panda Research and Protection Center in an identical environment. Region and age significantly affect the components of the microbiota in the reproductive tract of the giant panda. Hypothesis testing of species abundance data between the groups was performed using Metastats analysis. The species with significant differences ($p < 0.05$, $q < 0.05$) were screened as per the p - and q -values. The phyla with significant differences between the three groups (WL, YA, and DN) were as follows: Proteobacteria, Cyanobacteria, Chloroflexi, Euryarchaeota,

Planctomycetes, Acidobacteria, and Deinococcus-Thermus (Table A3: <https://doi.org/10.5281/zenodo.4067664>). Moreover, the phyla, which differed significantly between the AGE1 and AGE2 groups, were as follows: Proteobacteria, Firmicutes, Bacteroidetes, Cyanobacteria, Planctomycetes, Acidobacteria, and Deinococcus-Thermus (Table A4: <https://doi.org/10.5281/zenodo.4067664>).

3.3.4 | LefSe analysis of different species between groups

LDA score (LDA Score > 4) was used to indicate the significant differences between the species in different groups, as depicted in Figure A1. The LefSe test identified the differences in microbial communities between the three groups (WL, YA, and DN). Microbiota from all the samples was

classified at phylum, class, order, family, genus, and species levels, and a total of 43 differences were detected at each of these levels. Out of these 43 differences, 37 were contributed by the YA group, 5 by the DN group, and 1 by the WL group. The majority of the microbes enriched in the YA group belonged to Saccharomycetes, Tricholomataceae, Agaricomycetes, Microbotryaceae, Microbotryales, Cryptococcaceae, Naemateliaceae, and Tremellaceae. Additionally, the majority of the microbes enriched

in the DN group belonged to Actinomycetaceae, *Neisseria subflava*, *Ornithobacterium*, and *Helicobacter macacae*. Furthermore, the primary microbial population enriched in the WL group was of *Moraxella osloensis*.

A total of 87 differences were detected between the AGE2 and AGE3 groups (Figure A1b). Out of these 87 differences, 82 differences were contributed by the AGE2 group, and 5 differences by the AGE3 group. The majority of the microbial populations enriched

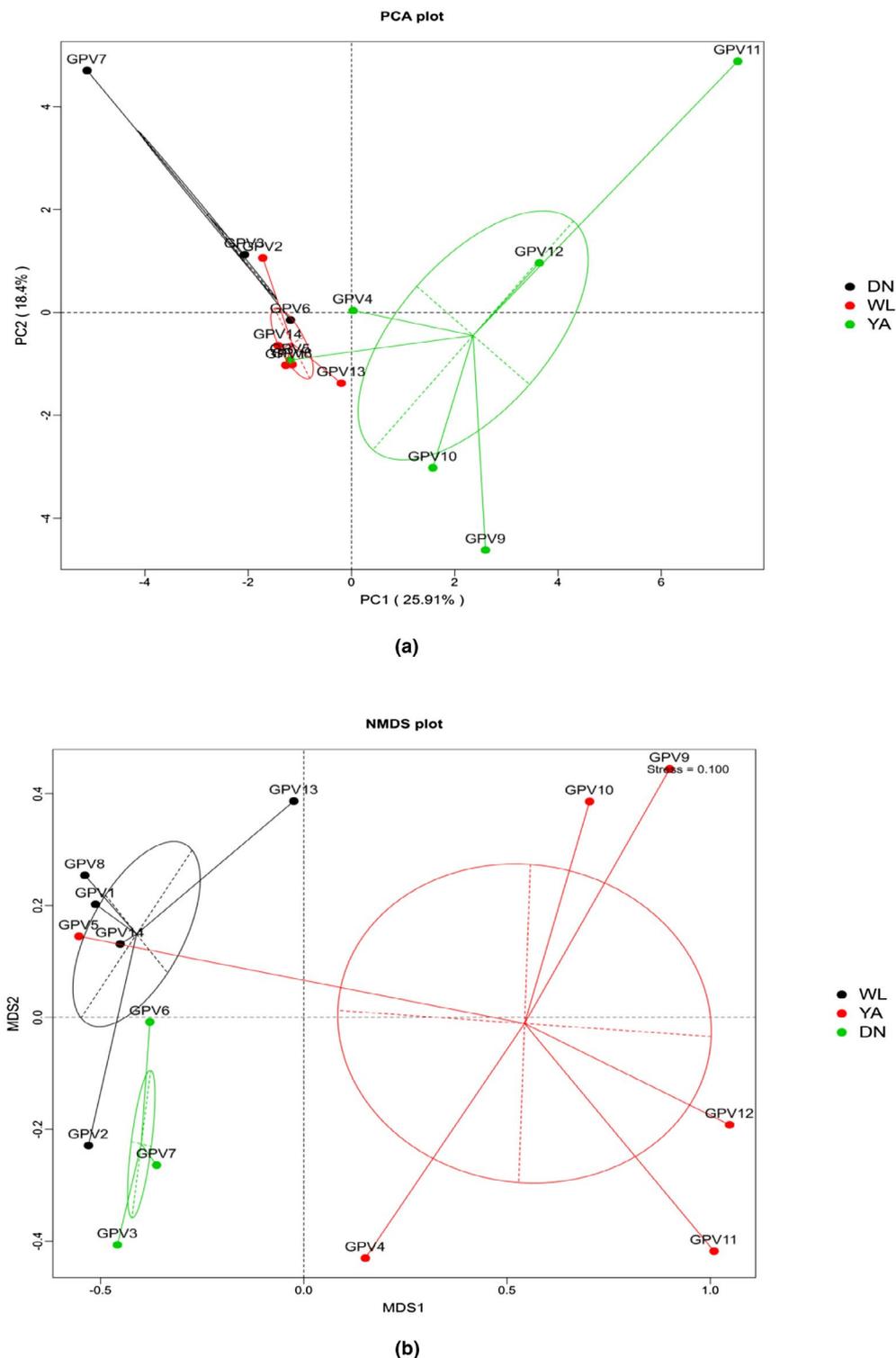


FIGURE 4 (Continued)

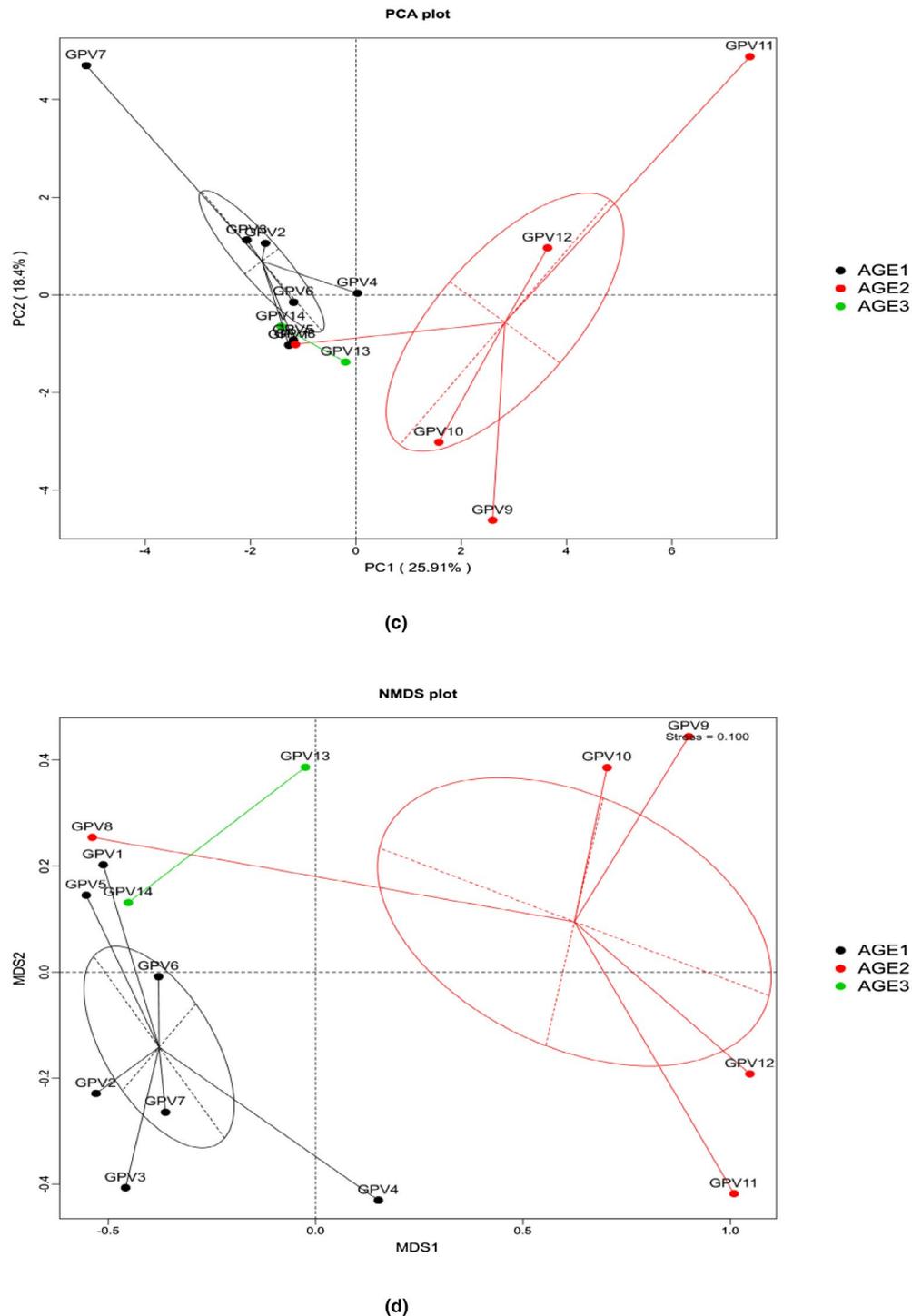


FIGURE 4 PCA and NMDS analysis of microorganisms composition at the phylum level. (a) PCA analysis of microorganisms composition in regional groups. (b) NMDS analysis of microorganisms composition in regional groups. (c) PCA analysis of microorganisms composition in age groups. (d) NMDS analysis of microorganisms composition in age groups.

in the AGE1 group belonged to Epsilonproteobacteria, *Bacteroides*, and *Corynebacterium*, and in the AGE2 group were Burkholderiales, *Burkholderiaceae*, Gammaproteobacteria, Pseudomonadales, *Pseudomonadaceae*, Saccharomycetes, *Metschnikowiaceae*, Agaricomycetes, Polyporales, *Tricholomataceae*, *Microbotryaceae*, Cryptococcaceae, *Cuniculitremaceae*, *Naemateliaceae*, Tremellales, *Tremellaceae*, Tremellomycetes, *Trichosporonaceae*.

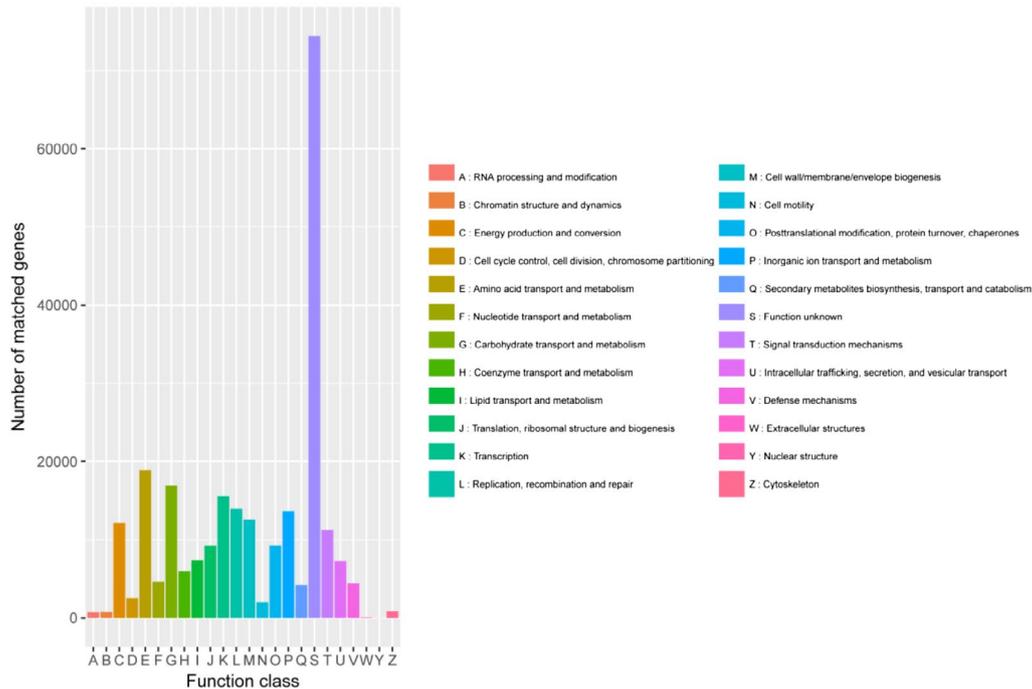
3.4 | Functional annotation and analysis

412,312 genes identified from all the 14 samples were analyzed for functional characterization. Out of the 412,312 genes, 240,050 (58.22%) genes were analyzed using eggNOG database, 9,978 (2.42%) in CAZy database, 259,596 (62.96%) in KEGG database, of which 143,757 (34.87%) genes were queried against 7,791 KEGG ortholog group (KO) in the KEGG database.

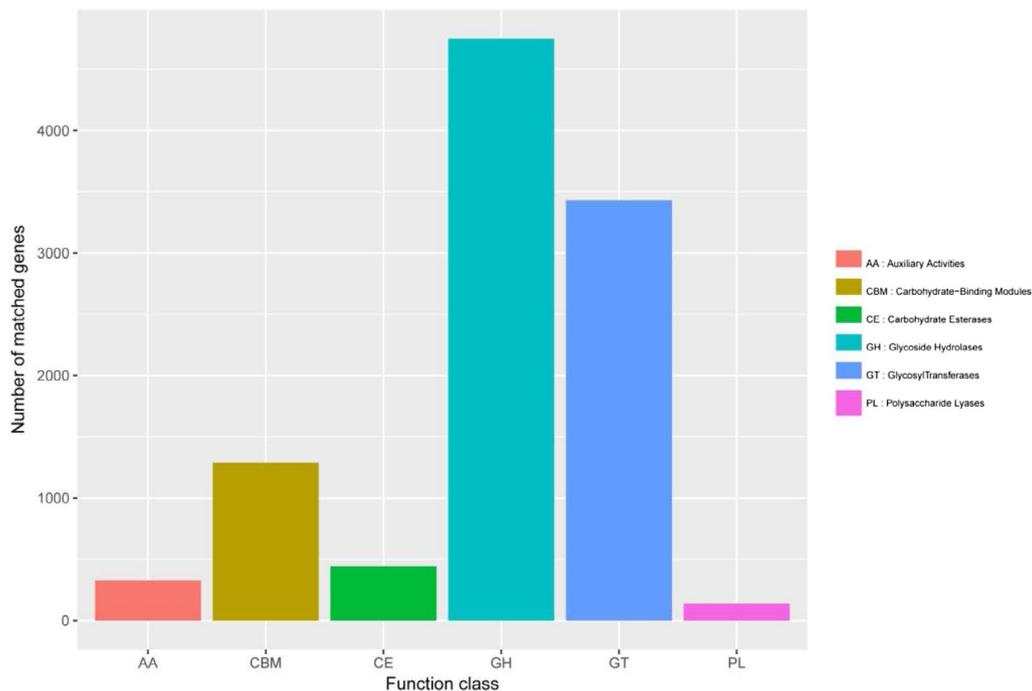
As per the functional analysis of genes using eggNOG, except for the unknown functions category, 24 functional categories were enriched, and a majority of the genes were enriched in functional categories, such as amino acid transport and metabolism, carbohydrate transport and metabolism, transcription, and replication, recombination, and repair (Figure 5a). Few genes were enriched in the nuclear structure and extracellular structure (each $\leq 0.01\%$) categories. The

functional characterization of the metagenomic sequencing data revealed that amino acid transport and metabolism, carbohydrate transport and metabolism were the dominant functions of the genes identified from the giant pandas' vaginal samples. These functions are also involved in microbial growth and development (Feng et al., 2018).

Besides, we examined the changes associated with the altered carbohydrate-active enzyme genes abundance. In the CAZY



(a)



(b)

FIGURE 5 (Continued)

KEGG pathway annotation

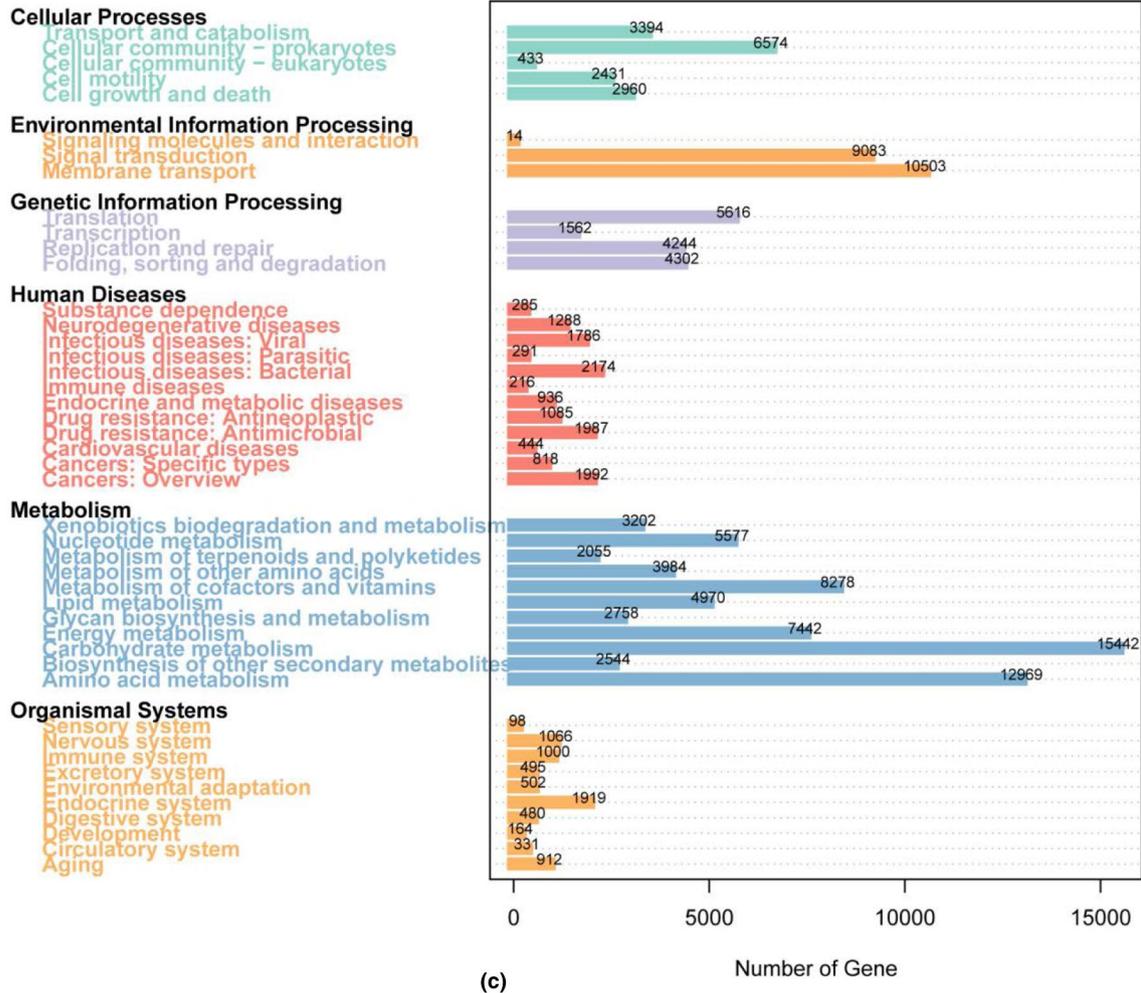


FIGURE 5 Annotated gene number statistics graphs and cluster trees based on Bray-Curtis distance for each database. (a) EggNOG's Unigenes annotation number statistical graph (level 1). (b) CAZy Unigenes annotation number statistical graph (level 1). (c) The number of functional annotations of genes at KEGG (level 1).

analysis, the abundant enzymes in descending order of abundance were glycoside hydrolases (GHs), glycosyltransferases (GTs), carbohydrate-binding modules (CBMs), carbohydrate esterases (CEs), auxiliary activities (AAs), and polysaccharide lyases (PLs; Figure 5b).

At level two of the CAZy family, 252 carbohydrate-active enzymes were identified using metagenomic analysis (Table A5: <https://doi.org/10.5281/zenodo.4067664>). GT2, which constitutes cellulose synthase (EC 2.4.1.12); chitin synthase (EC 2.4.1.16); dolichyl-phosphate β -D-mannosyltransferase (EC 2.4.1.83), and so on, was identified as the most abundant carbohydrate-active enzyme in all the 14 vaginal samples. Thus, we speculated that glycoside hydrolase is involved in the metabolic function of the giant panda's vagina.

Apart from eggNOG and CAZy functional analysis, we also performed a KEGG functional analysis of the 14 samples. Overall, the enriched functional profiles were divided into seven categories, and the most highly enriched category was metabolism, followed by environmental information processing, genetic information processing, cellular processes, human diseases, and organismal systems

(Figure 5c, Table A6: <https://doi.org/10.5281/zenodo.4067664>). In the level two of KEGG analysis, carbohydrate metabolism was the most highly enriched category followed by membrane transport, amino acid metabolism, metabolism of cofactors and vitamins, signal transduction, energy metabolism, cellular community-prokaryotes, nucleotide metabolism, translation, and lipid metabolism (Table A7: <https://doi.org/10.5281/zenodo.4067664>). A total of 392 functional groups were identified through the "KEGG Orthology" analysis (Table A8: <https://doi.org/10.5281/zenodo.4067664>). The signaling pathway enrichment analysis of genes from vaginal samples indicated that these pathways were involved in metabolism, cellular processes, environmental, and genetic information processing. The 392 pathways were divided into 612 functional modules, and out of these modules, M00178 (ribosome, bacteria) and M00179 (ribosome, archaea) modules, part of the map03010 pathways, were most highly enriched (Table A9: <https://doi.org/10.5281/zenodo.4067664>). It indicated that the dominant microbes present in the vagina of giant pandas were mainly bacteria and archaea.

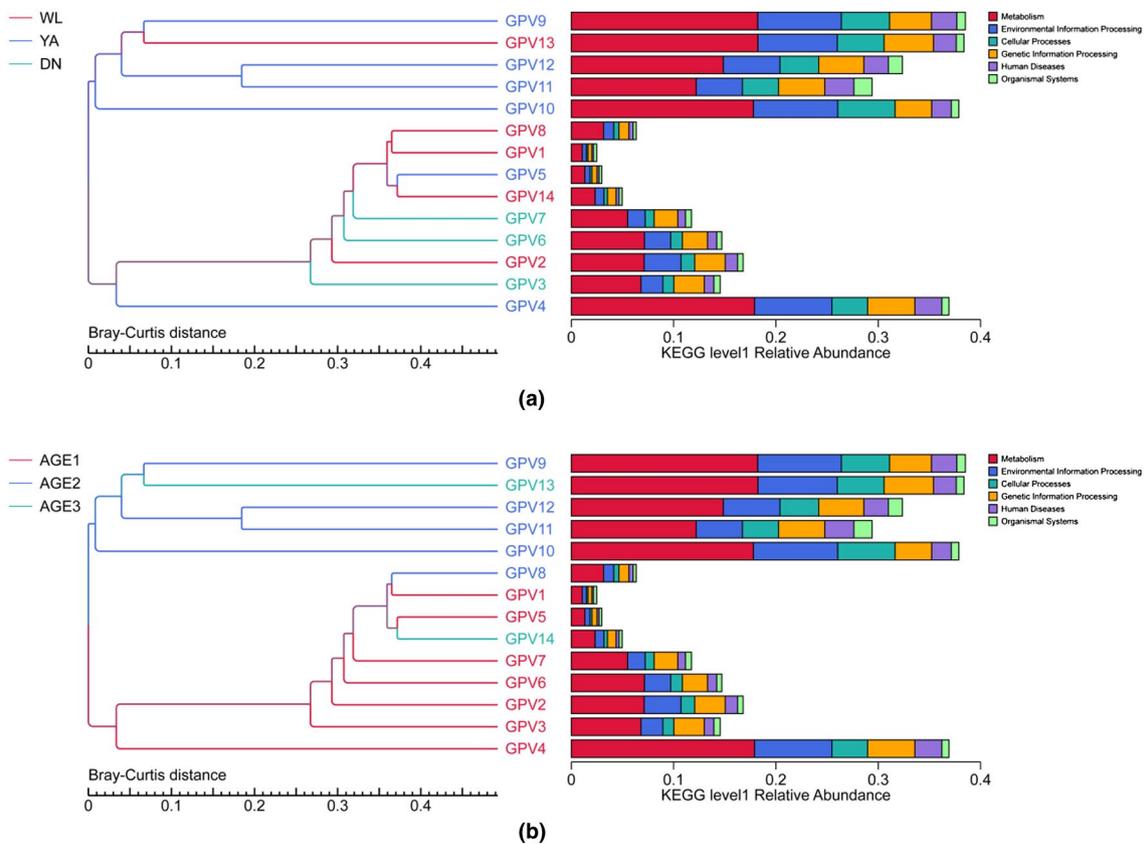


FIGURE 6 Cluster the trees according to the Bray-Curtis distance of the KEGG database. (a) Clustering tree based on Bray-Curtis distance in the regional groups. (b) Clustering tree based on Bray-Curtis distance in the age groups.

As per the KEGG database enrichment analysis, the ABC transporters, two-component system, purine metabolism, quorum sensing, ribosome, pyrimidine metabolism, bacterial secretion system, and phosphotransferase system (PTS) were the most highly enriched functional categories (Table A8: <https://doi.org/10.5281/zenodo.4067664>). Figure A2 depicts the top 10 enriched signaling pathways identified through the KEGG pathway analysis. ABC transporters (PATH: ko02010), involved in membrane transport, were enriched in most of the vaginal samples.

AGE and YA groups were identified as the most functionally active groups. Although the relative abundances of the pathways identified in all the vaginal samples did not differ significantly, the relative abundance of the phosphotransferase system (PTS) in AGE2 was significantly lower than AGE1 and AGE3 groups. Besides, the abundance of *Pseudomonas aeruginosa*, a biofilm-forming strain in the YA group, was significantly lower than WL and DN groups. These outcomes indicated that differences in age and region impact the function of microbes present in the giant panda's vagina.

4 | DISCUSSION

In the current study, we investigated the vaginal microbiome and associated functional profiles in the vaginal samples collected from the giant pandas from different geographical locations in China and different age groups using metagenomic sequencing. Currently,

metagenomic sequencing technology is being widely employed in sequencing studies. Although the metagenomic sequencing method cannot distinguish DNA from dead and living cells, it can substantially eliminate the limitations posed by conventional methods, such as microbial isolation and culture. Apart from being convenient and highly precise technology, metagenomic sequencing can expand the utilization space of microbial resources.

As implicated in the previous studies, apart from region and age, other factors also affect the microbiota of the vagina of giant panda significantly. In humans and other animals, feeding conditions severely impact the microbial makeup. Feeding style acts as a crucial driving factor for the colonization of intestinal microbes and influences the structure and function of intestinal microbiota in newborns (Azad et al., 2013) as a majority of the microbes in mammals are contributed through the lactation process in infants. All the giant pandas sampled in this study were fed by the Panda Research and Protection Center in an identical feeding environment. Besides, all the giant panda vaginal samples were collected and processed by the trained staff using a standard protocol. Thus, we decided that the samples can be analyzed to study the effect of region and age.

Microorganisms exist in different environmental matrices, such as water, air, and soil, where they are involved in a myriad of ecological functions (Abia et al., 2019). However, age and region of the host organism may also affect the microbial composition in a different environment. In 2017, we employed 16S rRNA high-throughput sequencing to analyze the reproductive tract microbiota of giant panda, which revealed

Proteobacteria (59.2%), Firmicutes (34.4%), Actinobacteria (5.2%) as the most abundant phyla and *Pseudomonas* (8.0%), *Streptococcus* (6.3%) as most abundant genera in these samples (Yang et al., 2017). Interestingly, in our previous study on fungal microbiome using HiSeq sequencing, we observed similar dominant microbiota in the giant panda's vagina, as observed in the current study (Chen et al., 2018). Although the proportion of microorganisms in the giant panda's vagina varied as per the analyses in this study, the outcome was in line with the previous study. For instance, *Pseudomonas* was the most abundant genus in vaginal samples from both the studies.

In this study, we determined the effect of age and region on vaginal samples using metagenomic sequencing. The species annotation of metagenomic data from vaginal samples in the age group (AGE1, AGE2, AGE3) showed that the AGE2 group had higher microbial diversity than the AGE1 and AGE3 group. A similar analysis of the vaginal samples in the region group (YA, WL, DN) showed that the YA group had higher microbial diversity as compared to the samples in the WL and DN groups (Tables A10 and A11: <https://doi.org/10.5281/zenodo.4067664>). The outcomes of PCA, heat map, and species clustering abundance analysis further validated these findings. Thus, we found that age and region affect the microbiome of the giant panda's vagina.

Chlamydia, a strict intracellular bacterial parasite (Sachse et al., 2015), causes uterine infection, premature birth, or miscarriage in humans (Moragianni et al., 2019), and in pigs, it causes respiratory diseases, diarrhea, conjunctivitis, and reproductive problems (Li et al., 2011). In females, *Chlamydia* infection is associated with sexual dysfunction and infertility (Pfennig, 2019). In the current study, *Chlamydia* was detected in all 14 samples but in different proportions. It is noteworthy that the DN group showed *Chlamydia* as the dominant bacteria. A majority of the vaginal samples in the DN group were from younger giant pandas. Out of all the 14 vaginal samples, GPV14 showed the highest *Chlamydia* content (9%).

Neisseria gonorrhoeae is a major cause of infertility in female (Costa-Lourenço et al., 2017). In the current study, *Neisseria* was detected in all the 14 vaginal samples. Krona was employed to examine the result of species annotations (Ondov et al., 2011). The outcome of this analysis showed that except for GPV1, GPV2, GPV4, and GPV12 samples, *Neisseria gonorrhoeae* was identified in all the other vaginal samples. GPV11 showed the highest abundance of (4%) of *Neisseria gonorrhoeae*. Humans are the only natural host of *Neisseria gonorrhoeae* known so far (Chan et al., 2016). In general, an abnormal microbiome may be one of the crucial causes of pregnancy failure in animals (García-Velasco et al., 2017). We speculated that *Chlamydia* and *Neisseria gonorrhoeae* infections could have an adverse effect on the reproductive ability of pandas as well.

The metagenomic sequencing technology has revolutionized the research and extended the understanding of microbial communities in various environmental matrices, such as water (Abia et al., 2018), soil (Su et al., 2014), and even air (Be et al., 2015). Also, functional screening can allow us to deduce the role of these microbial communities. Besides, it can unravel the novel genes that were not reported in previous sequencing analysis datasets

(Abia et al., 2019; Al-Hebshi et al., 2018; Bao et al., 2020; Shobo et al., 2020). In the current study, functional databases of eggNOG, CAZy, and KEGG were employed to analyze the metagenomic sequencing data, and it was found that the giant panda's vaginal microbiome was actively involved in metabolic functions. As per the eggNOG database annotation of the genes, a majority of the genes belonged to the amino acid transport and metabolism category. As per the CAZy analysis, 252 different carbohydrate-active enzymes were identified from the metagenomic sequencing data of the vaginal samples. The eggNOG and CAZy analysis revealed that in the AGE group, the AGE2 showed significantly higher gene numbers than AGE1 and AGE3 groups. Similarly, in the region group of vaginal samples, the number of genes annotated in the YA group was significantly higher than the WL and the DN groups. A high number of annotated genes in the AGE2 and YA groups indicate that age and region affect the giant panda's vaginal microbial function.

Meanwhile, metabolic pathway analysis using the KEGG dataset demonstrated enrichment of the membrane transport and nucleotide metabolic pathways. The ABC transporters (PATH: ko02010), a crucial pathway in microbes (Smart & Fleming, 1996), facilitates the lipid mobilization across the lipid bilayer. Also, it plays a crucial role in the construction and maintenance of the lipid bilayer. The enrichment of the ko02010 pathway may be due to membrane transport function associated with the vagina, such as the secretion of fluids. ABC transporters were one of the largest family of membrane transporters (Veen & Konings, 1998). Recent studies have shown that the ABC transporter is detected in all organisms, and it is involved in the active transport of various substrates across the lipid bilayer (Sasser et al., 2012). The loss-of-function mutation of the ABC transporter culminates in genetic disorders (Paumi et al., 2009). In the current study, as per the KEGG analysis, ABC transporters may also affect the reproduction rate of giant pandas. The cluster analysis of age and regional group vaginal samples presented an apparent difference between these two groups. (Figure 6, Figure A3).

In the current metagenomics analysis study of the giant pandas' vaginal samples from different age groups and regions, we identified a total of 33 phyla and 1000 genera. Taxonomic profiling showed that the microbiome present in the vagina of giant pandas is affected by age and region. The functional pathways of vaginal microbes were correlated to the composition of the microbial community. A majority of these functions were induced by age and region. Therefore, compared to the region, age had a clear impact on the species and functional abundance of the microbiome in giant pandas' reproductive tract.

5 | CONCLUSIONS

In this study, we analyzed the vaginal microbiome of giant pandas using metagenomic sequencing. As per the outcomes of our analyses, Proteobacteria, Basidiomycota, Firmicutes, and Actinobacteria were identified as the dominant phyla in vaginal samples from giant

pandas. Besides, we found that the geographical location and age affect the giant panda's vaginal microbiome and associated functions involving signaling pathways, proteins, and CAZy. Overall, *Chlamydia* and *Neisseria gonorrhoeae* in the reproductive tract of giant pandas may adversely affect the reproductive tract of giant pandas.

ACKNOWLEDGMENTS

This research was supported by the Giant Panda International Cooperation Fund from State Forestry Administration of China: an investigation on the microbiome of the reproductive tract of the giant panda (Grant number: GH201714).

CONFLICTS OF INTERESTS

None declared.

AUTHOR CONTRIBUTION

Lan Zhang: Data curation (equal); Formal analysis (equal); Writing-original draft (lead). **Caiwu Li:** Funding acquisition (equal); Investigation (equal). **Yaru Zhai:** Formal analysis (equal). **Lan Feng:** Data curation (equal). **Keke Bai:** Formal analysis (equal). **Zhizhong Zhang:** Funding acquisition (equal); Investigation (equal). **Yan Huang:** Funding acquisition (equal); Investigation (equal). **Ti Li:** Funding acquisition (equal); Investigation (equal). **Desheng Li:** Funding acquisition (equal); Investigation (equal). **Hao Li:** Data curation (equal); Formal analysis (equal). **Pengfei Cui:** Data curation (equal); Formal analysis (equal). **Danyu Chen:** Data curation (equal). **Hong-Ning Wang:** Supervision (equal). **Xin Yang:** Writing-original draft (equal); Writing-review & editing (equal).

ETHICS STATEMENT

The animal study was reviewed and approved by the Sichuan University Animal Ethics Committee.

DATA AVAILABILITY STATEMENT

Sequence data generated during the current study are available in the NCBI repository at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA659379>. Supplementary tables are available in Zenodo at <https://doi.org/10.5281/zenodo.4067664>: Table A1. Sample sequencing data processing; Table A2. Sample assembly - Scaffolds basic information statistics (≥ 500 bp); Table A3. Species with significant differences at the phylum level; Table A4. Species with significant differences at the phylum level; Table A5. Functional category prediction based on genes extracted from giant pandas' vagina samples from the CAZy database (level 2); Table A6. Functional category prediction based on genes extracted from giant pandas' vagina samples from the KEGG database (level 1); Table A7. Functional category prediction based on genes extracted from giant pandas' vagina samples from the KEGG database (level 2); Table A8. Functional category prediction based on genes extracted from giant pandas' vagina samples from the KEGG database (level 3); Table A9. Relative abundances of the KEGG modules for the giant pandas' vaginal samples (level 4); Table A10. Table of relative abundance of each sample annotated at the phylum level; Table A11. Table of relative abundance of each sample annotated at the genus level.

ORCID

Xin Yang  <https://orcid.org/0000-0002-7940-714X>

REFERENCES

- Abia, A. L. K., Alisoltani, A., Keshri, J., & Ubomba-Jaswa, E. (2018). Metagenomic analysis of the bacterial communities and their functional profiles in water and sediments of the Apies River, South Africa, as a function of land use. *Science of the Total Environment*, 616–617, 326–334. <https://doi.org/10.1016/j.scitotenv.2017.10.322>
- Abia, A. L. K., Alisoltani, A., Ubomba-Jaswa, E., & Dippenaar, M. A. (2019). Microbial life beyond the grave: 16S rRNA gene-based metagenomic analysis of bacteria diversity and their functional profiles in cemetery environments. *Science of the Total Environment*, 655, 831–841. <https://doi.org/10.1016/j.scitotenv.2018.11.302>
- Al-Fatlawy, H. J. (2014). Isolation and identification of aerobic vaginal bacteria and fungi from adult cows in AL-Kufa district/AL-Najaf province-Iraq. *Mrsva*, 3, 19–23.
- Al-Hebshi, N. N., Baraniya, D., Chen, T., Hill, T., Puri, S., Tellez, M., Hasan, N. A., Colwell, R. B., & Ismail, A. (2018). Metagenome sequencing-based strain-level and functional characterization of supragingival microbiome associated with dental caries in children. *Journal of Oral Microbiology*, 11(1), 1557986. <https://doi.org/10.1080/20002297.2018.1557986>
- Avershina, E., Frisli, T., & Rudi, K. (2013). De novo semi-alignment of 16S rRNA gene sequences for deep phylogenetic characterization of next generation sequencing data. *Microbes and Environments*, 28(2), 211–216. <https://doi.org/10.1264/jsm2.ME12157>
- Azad, M. B., Konya, T., Maughan, H., Guttman, D. S., Field, C. J., Chari, R. S., Sears, M. R., Backer, A. B., Scott, J. A., & Kozyrskyj, A. L. (2013). Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *Canadian Medical Association Journal*, 185(5), 385–394. <https://doi.org/10.1503/cmaj.121189>
- Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., Khan, M. T., Zhang, J., Li, J., Xiao, L., Al-Aama, J., Zhang, D., Lee, Y. S., Kotowska, D., Colding, C., ... Wang, J. (2015). Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host & Microbe*, 17(6), 690–703. <https://doi.org/10.1016/j.chom.2015.04.004>
- Bao, T., Deng, S., Yu, K., Li, W., & Dong, A. (2020). Metagenomic insights into seasonal variations in the soil microbial community and function in a Larix gmelinii forest of Mohe, China. *Journal of Forestry Research*, <https://doi.org/10.1007/s11676-019-01090-w>
- Be, N. A., Thissen, J. B., Fofanov, V. Y., Allen, J. E., Rojas, M., Golovko, G., Fofanov, Y., Koshinsky, H., & Jaing, C. J. (2015). Metagenomic analysis of the airborne environment in urban spaces. *Microbial Ecology*, 69(2), 346–355. <https://doi.org/10.1007/s00248-014-0517-z>
- Bradford, L. L., & Ravel, J. (2017). The vaginal mycobiome: A contemporary perspective on fungi in women's health and diseases. *Virulence*, 8(3), 342–351. <https://doi.org/10.1080/21505594.2016.1237332>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Brum, J. R., Ignacio-Espinoza, J. C., Roux, S., Doulcier, G., Acinas, S. G., Alberti, A., Chaffron, S., Cruaud, C., de Vargas, C., Gasol, J. M., Gorsky, G., Gregory, A. C., Guidi, L., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Poulos, B. T., ... Sullivan, M. B. (2015). Patterns and ecological drivers of ocean viral communities. *Science*, 348(6237), 1261498. <https://doi.org/10.1126/science.1261498>
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1), 59–60. <https://doi.org/10.1038/nmeth.3176>
- Cafarchia, C., Figueredo, L. A., & Otranto, D. (2013). Fungal diseases of horses. *Veterinary Microbiology*, 167(7), 215–234. <https://doi.org/10.1016/j.vetmic.2013.01.015>

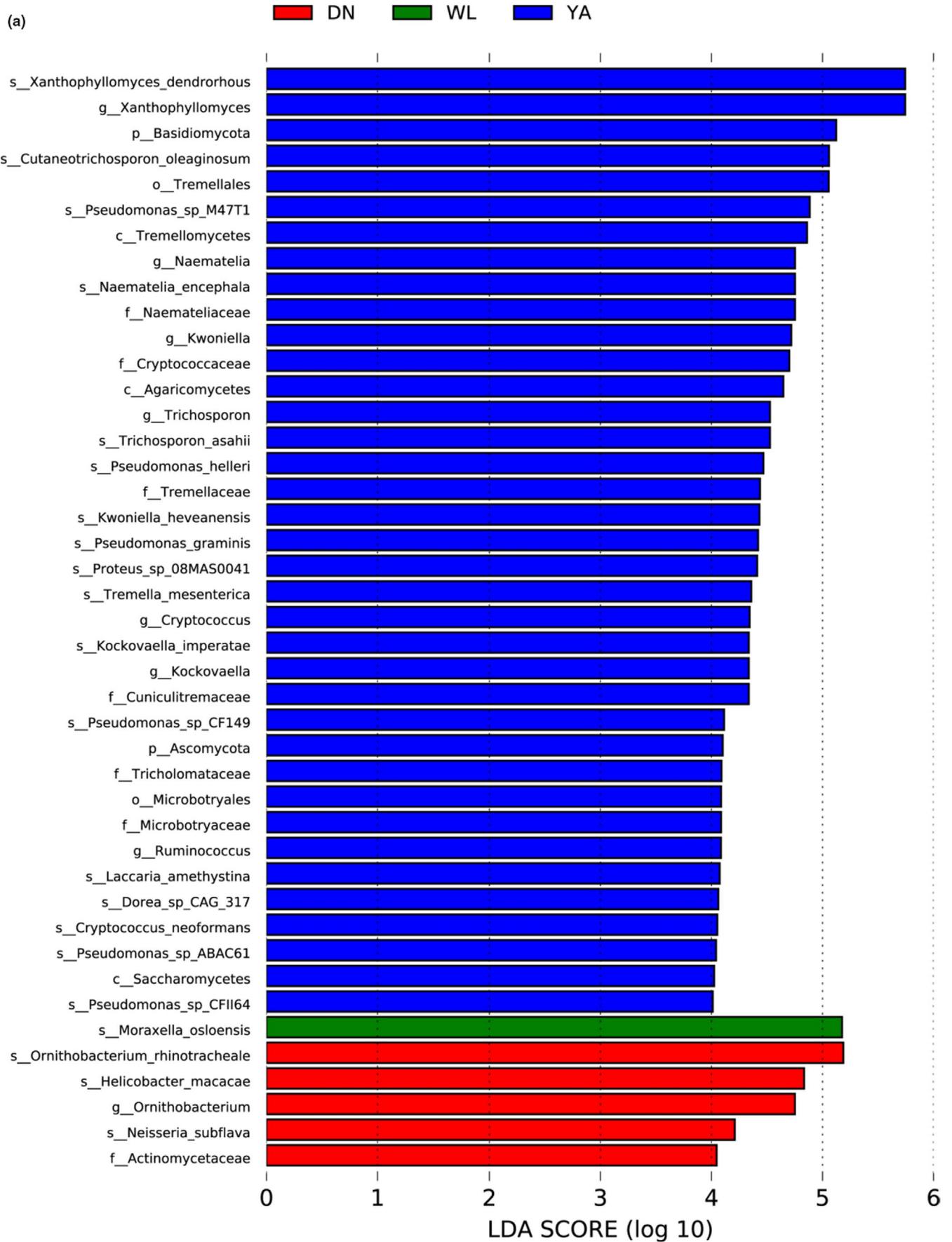
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, *37(Database, issue):* D233–238. <https://doi.org/10.1093/nar/gkn663>
- Chan, P. A., Robinette, A., Montgomery, M., Almonte, A., Cu-Uvin, S., Lonks, J. R., & Hardy, E. J. (2016). Extragenital infections caused by *Chlamydia trachomatis* and *Neisseria gonorrhoeae*: a review of the literature. *Infectious Diseases in Obstetrics and Gynecology*, *2016*, 5758387.
- Chen, D., Li, C., Feng, L., Zhang, Z., Zhang, H., Cheng, G., Li, D., Zhang, G., Wang, H., Chen, Y., Feng, M., Wang, C., Wu, H., Deng, L., Ming, H., & Yang, X. (2018). Analysis of the influence of living environment and age on vaginal fungal microbiome in giant pandas (*Ailuropoda melanoleuca*) by high throughput sequencing. *Microbial Pathogenesis*, *115*, 280–286. <https://doi.org/10.1016/j.micpath.2017.12.067>
- Chen, K., & Pachter, L. (2005). Bioinformatics for whole-genome shotgun sequencing of microbial communities. *Plos Computational Biology*, *1(2)*, e24. <https://doi.org/10.1371/journal.pcbi.0010024>
- Costa-Lourenço, A. P. R., Barros dos Santos, K. T., Moreira, B. M., Fracalanza, S. E. L., & Bonelli, R. R. (2017). Antimicrobial resistance in *Neisseria gonorrhoeae*: history, molecular mechanisms and epidemiological aspects of an emerging global threat. *Brazilian Journal of Microbiology*, *48(4)*, 617–628.
- Cotillard, A., Kennedy, S. P., Kong, L. C., Prifti, E., Pons, N., Le Chatelier, E., Almeida, M., Quinquis, B., Levenez, F., Galleron, N., Gougis, S., Rizkalla, S., Batto, J. M., Renault, P., Doré, J., Zucker, J.-D., Clément, K., & Ehrlich, S. D. (2013). Dietary intervention impact on gut microbial gene richness. *Nature*, *500(7464)*, 585–588. <https://doi.org/10.1038/nature12480>
- Dascanio, J. J., Schweizer, C., & Ley, W. B. (2010). Equine fungal endometritis. *Equine Veterinary Education*, *13*, 324–329. <https://doi.org/10.1111/j.2042-3292.2001.tb00122.x>
- De Puyssseleyr, K., De Puyssseleyr, L., Geldhof, J., Cox, E., & Vanrompay, D. (2014). Development and Validation of a Real-Time PCR for *Chlamydia suis* diagnosis in Swine and Humans. *PLoS One*, *9(5)*, e96704. <https://doi.org/10.1371/journal.pone.0096704>
- Doonan, J., Denman, S., McDonald, J. E., & Golyshin, P. N. (2016). Shotgun metagenomic sequencing analysis of soft-rot enterobacteriaceae in polymicrobial communities. *Methods in Molecular Biology*, *1539*, 85–97. https://doi.org/10.1007/978-1-4939-6691-2_7
- Feng, G., Xie, T., Wang, X., Bai, J., Tang, L., Zhao, H., Wei, W., Wang, M., & Zhao, Y. (2018). Metagenomic analysis of microbial community and function involved in cd-contaminated soil. *BMC Microbiology*, *18(1)*, 11. <https://doi.org/10.1186/s12866-018-1152-5>
- Feng, Q., Liang, S., Jia, H., Stadlmayr, A., Tang, L., Lan, Z., Zahng, D., Xia, H., Xu, X., Jie, Z., Su, L., Li, X., Li, J., Xiao, L., Huber-Schönauer, U., Niederseer, D., Xu, X., Alaama J. Y., ... Wang J. (2015). Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nature. Communication*, *6*, 6528. <https://doi.org/10.1055/s-0035-1551729>
- Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28(23)*, 3150–3152.
- García-Velasco, J. A., Menabrito, M., & Catalán, I. B. (2017). What fertility specialists should know about the vaginal microbiome: a review. *Reproductive Biomedicine Online*, *35(1)*, 103–112. <https://doi.org/10.1016/j.rbmo.2017.04.005>
- Hedayati, M. T., Taheri, Z., Galinimoghadam, T., Aghili, S. R., Yazdani Cherati, J., & Mosayebi, E. (2015). Isolation of different species of candida in patients with vulvovaginal candidiasis from sari. *Iran. Jundishapur Journal of Microbiology*, *8(4)*, e15992. [https://doi.org/10.5812/jjm.8\(4\)2015.15992](https://doi.org/10.5812/jjm.8(4)2015.15992)
- Hu, J. C. (1997). Existing circumstances and prospect of the giant panda. *Journal of Sichuan Teachers College (Natural Science)*, *18*, 129–133.
- Huson, D. H., Mitra, S., Ruscheweyh, H. J., Weber, N., & Schuster, S. C. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Research*, *21(9)*, 1552–1560. <https://doi.org/10.1101/gr.120618.111>
- Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., & Hirakawa, M. (2006). From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research*, *34(90001)*, D354–D357. <https://doi.org/10.1093/nar/gkj102>
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Research*, *42(D1)*, D199–D205. <https://doi.org/10.1093/nar/gkt1076>
- Karlsson, F. H., Fåk, F., Nookaew, I., Tremaroli, V., Fagerberg, B., Petranovic, D., Bäckhed, F., & Nielsen, J. (2012). Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nature Communication*, *3*, 1245. <https://doi.org/10.1038/ncomms2266>
- Karlsson, F. H., Tremaroli, V., Nookaew, I., Bergstrom, G., Behre, C. J., Fagerberg, B., Nielsen, J., & Bäckhed, F. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature*, *498(7452)*, 99–103. <https://doi.org/10.1038/nature12198>
- Knudtson, W. U., & Kirkbride, C. A. (1992). Fungi Associated with Bovine Abortion in the Northern Plains States (USA). *Journal of Veterinary Diagnostic Investigation*, *4(2)*, 181–185. <https://doi.org/10.1177/104063879200400211>
- Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J. M., Kennedy, S., Leonard, P., Li, J., Burgdorf, K., Grarup, N., Jørgensen, T., Brandslund, I., Nielsen, H. B., Juncker, A. S., Bertalan, M., ... Pedersen, O. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature*, *500(7464)*, 541–546.
- Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Arumugam, M., Kultima, J. R., Prifti, E., Nielsen, T., Juncker, A. S., Manichanh, C., Chen, B., Zhang, W., Levenez, F., Wang, J., Xu, X., Xiao, L., Liang, S., ... Wang, J. (2014). An integrated catalog of reference genes in the human gut microbiome. *Nature Biotechnology*, *32(8)*, 834–841. <https://doi.org/10.1038/nbt.2942>
- Li, T., Liu, Z. H., Li, K., & Bai, H. H. (2019). Evaluation of the vaginal microbiome in clinical diagnosis and management of vaginal infectious diseases. *Chinese Medical Journal*, *132(9)*, 1100–1103. <https://doi.org/10.1097/cm9.0000000000000211>
- Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22(13)*, 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>
- Li, Y., Wang, Y., Nie, F., Xiao, J., Wang, G., Yuan, L., & Li, Z. (2011). A nested multiplex polymerase chain reaction assay for the differential identification of three zoonothropotic chlamydial strains in porcine swab samples. *Journal of Veterinary Diagnostic Investigation*, *23(4)*, 673–681. <https://doi.org/10.1177/1040638711406972>
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., ... Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*, *1(1):*18. <https://doi.org/10.14322/publons.r3715>
- Ma, X., Li, C., Hou, J., & Gu, Y. (2017). Isolation and identification of culturable fungi from the genitals and semen of healthy giant pandas (*Ailuropoda melanoleuca*). *BMC Veterinary Research*, *13(1)*, 344. <https://doi.org/10.1186/s12917-017-1231-0>
- Mende, D. R., Waller, A. S., Sunagawa, S., Aino, J., Michelle, C., Manimozhiyan, A., Chaffron, S., Cruaud, C., de Vargas, C., Gasol, J. M., Gorsky, G., Gregory, A. C., Guidi, L., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Poulos, B. T., ... Sullivan, M. B. (2012). Assessment of metagenomic assembly using simulated next

- generation sequencing data. *PLoS One*, 7(6237), e31386. <https://doi.org/10.1371/journal.pone.0114063>
- Moraes, I. A., Stussi, J. S., Lilenbaum, W., Pissinatti, A., Luz, F. P., & Ferreira, A. M. R. (2014). Isolation and identification of fungi from vaginal flora in three species of captive *Leontopithecus*. *American Journal of Primatology*, 64(3), 337–343. <https://doi.org/10.1002/ajp.20082>
- Moragianni, D., Dryllis, G., Andromidas, P., Kapeta-Korkouli, R., Kouskouni, E., Pessach, I., Papalexis, P., Kodonaki, A., Athanasiou, N., Poulidakis, A., & Baka, S. (2019). Genital tract infection and associated factors affect the reproductive outcome in fertile females and females undergoing in vitro fertilization. *Biomedical Reports*, 10(4), 231–237. <https://doi.org/10.3892/br.2019.1194>
- Nielsen, H. B., Almeida, M., Juncker, A. S., Rasmussen, S., Li, J., Sunagawa, S., Plichta, D. R., Gautier, L., Pedersen, A. G., Le Chatelier, E., Pelletier, E., Bonde, I., Nielsen, T., Manichanh, C., Arumugam, M., Batto, J.-M., Quintanilha dos Santos, M. B., Blom, N., Borruel, N., ... Ehrlich, S. D. (2014). Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nature Biotechnology*, 32(8), 822–828. <https://doi.org/10.1038/nbt.2939>
- Odintsova, V., Tyakht, A., & Alexeev, D. (2017). Guidelines to statistical analysis of microbial composition data inferred from metagenomic sequencing. *Current Issues in Molecular Biology*, 24, 17–36. <https://doi.org/10.21775/cimb.024.017>
- Oh, J., Byrd, A. L., Deming, C., Conlan, S., Kong, H. H., & Segre, J. A. (2014). Biogeography and individuality shape function in the human skin metagenome. *Nature*, 514(7520), 59–64. <https://doi.org/10.1038/nature13786>
- Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, 12, 385. <https://doi.org/10.1186/1471-2105-12-385>
- Pal, M. (2015). Growing role of fungi in mycotic abortion of domestic animal. *Journal of Bacteriology and Mycolgy*, 2, 1009.
- Paumi, C. M., Chuk, M., Snider, J., Stagljar, I., & Michaelis, S. (2009). ABC transporters in *Saccharomyces cerevisiae* and their interactors: New technology advances the biology of the ABCC (MRP) subfamily. *Microbiology and Molecular Biology Reviews*, 73(4), 577–593. <https://doi.org/10.1128/MMBR.00020-09>
- Peng, J. J., Jiang, Z. G., & Hu, J. C. (2001). Status and conservation of giant panda (*Ailuropoda melanoleuca*): A review. *Folia Zoologica Praha*, 50(2), 81–88.
- Peng, R., Zeng, B., Meng, X., Yue, B., Zhang, Z., & Zou, F. (2007). The complete mitochondrial genome and phylogenetic analysis of the giant panda (*Ailuropoda melanoleuca*). *Gene*, 397(1–2), 76–83. <https://doi.org/10.1016/j.gene.2007.04.009>
- Pfennig, C. L. (2019). Sexually transmitted diseases in the emergency department. *Emergency Medicine Clinics of North America*, 37(2), 165–192.
- Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., Gabaldón, T., Rattei, T., Creevey, C., Kuhn, M., Jensen, L. J., von Mering, C., & Bork, P. (2014). eggNOG v4.0: Nested orthology inference across 3686 organisms. *Nucleic Acids Research*, 42(D1), D231–D239. <https://doi.org/10.1093/nar/gkt1253>
- Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Guan, Y., Shen, D., Peng, Y., Zhang, D., Jie, Z., Wu, W., Qin, Y., Xue, W., Li, J., Han, L., Lu, D., Wu, P., Dai, Y., ... Wang, J. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*, 490(7418), 55–60. <https://doi.org/10.3410/f.717957500.793464381>
- Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., Guo, J., Chatelier, E. L., Yao, J., Wu, L., Zhou, J., Ni, S., Liu, L., Pons, N., Batto, J. M., Kennedy, S. P., Leonard, P., Yuan, C., Ding, W., ... Li, L. (2014). Alterations of the human gut microbiome in liver cirrhosis. *Nature*, 513(7516), 59–64. <https://doi.org/10.1038/nature13568>
- Rivas, M. N., Burton, O. T., Wise, P., Zhang, Y., Hobson, S. A., Lloret, M. G., Chehoud, C., Kuczynski, J., DeSantis, T., Warrington, J., Hyde, E. R., Petrosino, J. F., Gerber, G. K., Bry, L., Oettgen, H. C., Mazmanian, S. K., & Chatila, T. A. (2013). A microbiota signature associated with experimental food allergy promotes allergic sensitization and anaphylaxis. *Journal of Allergy and Clinical Immunology*, 131(1), 201–212. <https://doi.org/10.1016/j.jaci.2012.10.026>
- Sachse, K., Bavoil, P. M., Kaltenboeck, B., Stephens, R. S., Kuo, C. C., Rosselló-Móra, R., & Horn, M. (2015). Emendation of the family Chlamydiaceae: Proposal of a single genus, *Chlamydia*, to include all currently recognized species. *Systematic and Applied Microbiology*, 38(2), 99–103. <https://doi.org/10.1016/j.syapm.2014.12.004>
- Sasser, T. L., Padolina, M., & Fratti, R. A. (2012). The yeast vacuolar ABC transporter Ybt1p regulates membrane fusion through Ca²⁺ transport modulation. *Biochemical Journal*, 448(3), 365–372. <https://doi.org/10.1042/BJ20120847>
- Schautteet, K., & Vanrompay, D. (2011). Chlamydiaceae infections in pig. *Veterinary Research*, 42(1), 29. <https://doi.org/10.1186/1297-9716-42-29>
- Scher, J. U., Sczesnak, A., Longman, R. S., Segata, N., Ubeda, C., Bielski, C., Rostron, T., Cerundolo, V., Pamer, E. G., Abramson, S. B., Huttenhower, C., & Littman, D. R. (2013). Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *Elife*, 2, e01202. <https://doi.org/10.7554/eLife.01202>
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, 12(6), 1–18. <https://doi.org/10.1186/gb-2011-12-6-r60>
- Shobo, C. O., Alisoltani, A., Abia, A. L. K., Mtshali, P. S., Lsmail, A., Zishiri, O., Horn, J. D., Brysiewicz, P., Essack, S. Y., & Bester, L. A. (2020). Bacterial diversity and functional profile of microbial populations on surfaces in public hospital environments in South Africa: A high throughput metagenomic analysis. *Science of the Total Environment*, 719, 137360. <https://doi.org/10.1016/j.scitotenv.2020.137360>
- Smart, C. C., & Fleming, A. J. (1996). Hormonal and environmental regulation of a plant PDR5-like ABC transporter. *Journal of Biological Chemistry*, 271(32), 19351–19357.
- Stumpf, R. M., Wilson, B. A., Rivera, A., Yildirim, S., Yeoman, C. J., Polk, J. D., White, B. A., & Leigh, S. R. (2013). The primate vaginal microbiome: Comparative context and implications for human health and disease. *American Journal of Physical Anthropology*, 152, 119–134. <https://doi.org/10.1002/ajpa.22395>
- Su, J. Q., Wei, B., Xu, C. Y., Qiao, M., & Zhu, Y. G. (2014). Functional metagenomic characterization of antibiotic resistance genes in agricultural soils from China. *Environment International*, 65, 9–15. <https://doi.org/10.1016/j.envint.2013.12.010>
- Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., ... Velayoudon, D. (2015). Structure and function of the global ocean microbiome. *Science*, 348(6237), 1261359. <https://doi.org/10.1126/science.1261359>
- van Veen, H. W., & Konings, W. N. (1998). The ABC family of multidrug transporters in microorganisms. *Biochimica Et Biophysica Acta*, 1365(1–2), 31–36.
- Villar, E., Farrant, G. K., Follows, M., Garczarek, L., Speich, S., Audic, S., Bittner, L., Blanke, B., Brum, J. R., Brunet, C., Casotti, R., Chase, A., Dolan, J. R., d'Ortenzio, F., Gattuso, J.-P., Grima, N., Guidi, L., Hill, C. N., Jahn, O., ... Iudicone, D. (2015). Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science*, 348(6237), 1261447. <https://doi.org/10.1126/science.1261447>
- White, J. R., Nagarajan, N., & Pop, M. (2009). Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Computational Biology*, 5(4), e1000352. <https://doi.org/10.1371/journal.pcbi.1000352>

- Yan, X., Zhang, H., Li, D., Wu, D., Zhou, S., Sun, M., Hu, H., Liu, X., Mou, S., He, S., Owen, M. A., & Huang, Y. (2019). Acoustic recordings provide detailed information regarding the behavior of cryptic wildlife to support conservation translocations. *Scientific Reports*, 9(1), 5172. <https://doi.org/10.1038/s41598-019-41455-z>
- Yang, X., Cheng, G., Li, C., Yang, J., Li, J., Chen, D., Zou, W., Jin, S., Zhang, H., Li, D., He, Y., Wang, C., Wang, M., & Wang, H. (2017). The normal vaginal and uterine bacterial microbiome in giant pandas (*Ailuropoda melanoleu ca*). *Microbiology Research*, 199, 1–9. <https://doi.org/10.1016/j.micres.2017.01.003>
- Yang, X., Yang, J., Wang, H., Li, C., He, Y., Jin, S., Zahng, H., Li, D., Wang, P., Xu, Y., Xu, C., Fan, C., Xu, L., Huang, S., Qu, C., & Li, G. (2016). Normal vaginal bacterial flora of giant pandas (*ailuropoda melanoleuca*) and the antimicrobial susceptibility patterns of the isolates. *Journal of Zoo and Wildlife Medicine*, 47(1), 671–675. <https://doi.org/10.1638/2015-0203.1>
- Zeller, G., Tap, J., Voigt, A. Y., Sunagawa, S., Kultima, J. R., Costea, P. L., Amiot, A., Böhm, J., Brunetti, F., Habermann, N., Hercog, R., Koch, M., Luciani, A., Mende, D. R., Schneider, M. A., Schrotz-King, P., Tournigand, C., Nhieu, J. T. V., Yamada, T., ... Bork, P. (2014). Potential of fecal microbiota for early stage detection of colorectal cancer. *Molecular Systems Biology*, 10(11), 766. <https://doi.org/10.15252/msb.20145645>

How to cite this article: Zhang L, Li C, Zhai Y, et al. Analysis of the vaginal microbiome of giant pandas using metagenomics sequencing. *MicrobiologyOpen*. 2020;9:e1131. <https://doi.org/10.1002/mbo3.1131>

APPENDIX A1



(b) AGE1 AGE2

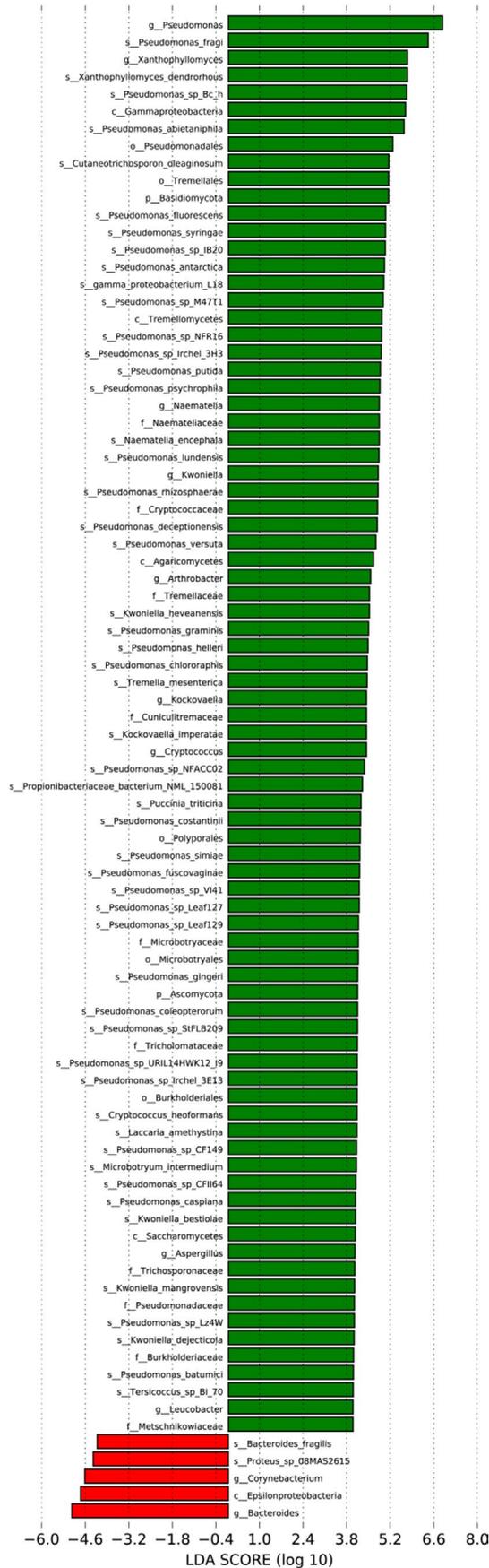


Figure A1 LDA value distribution map and evolutionary branch diagram of different microorganisms composition in regional and age groups. (a) LDA value distribution map of different microorganisms composition in regional groups. (b) LDA value distribution map of different microorganisms composition in age groups.

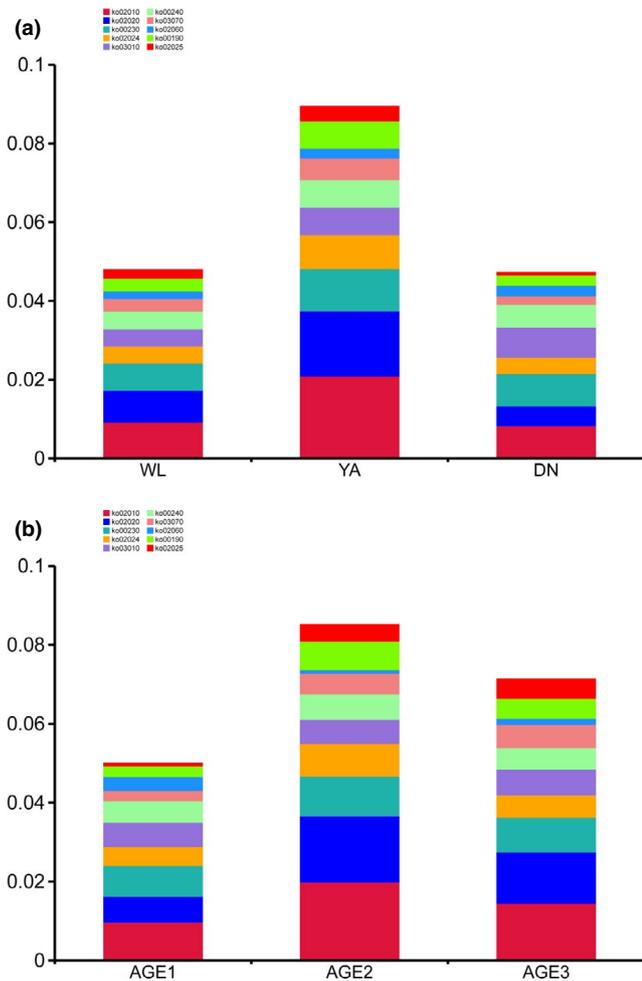


Figure A2 Relative abundances of the top 10 KEGG pathways in the region and age groups. (a) Relative abundances of the top 10 KEGG pathways in the region group (level 3). (b) Relative abundances of the top 10 KEGG pathways in the age group (level 3).

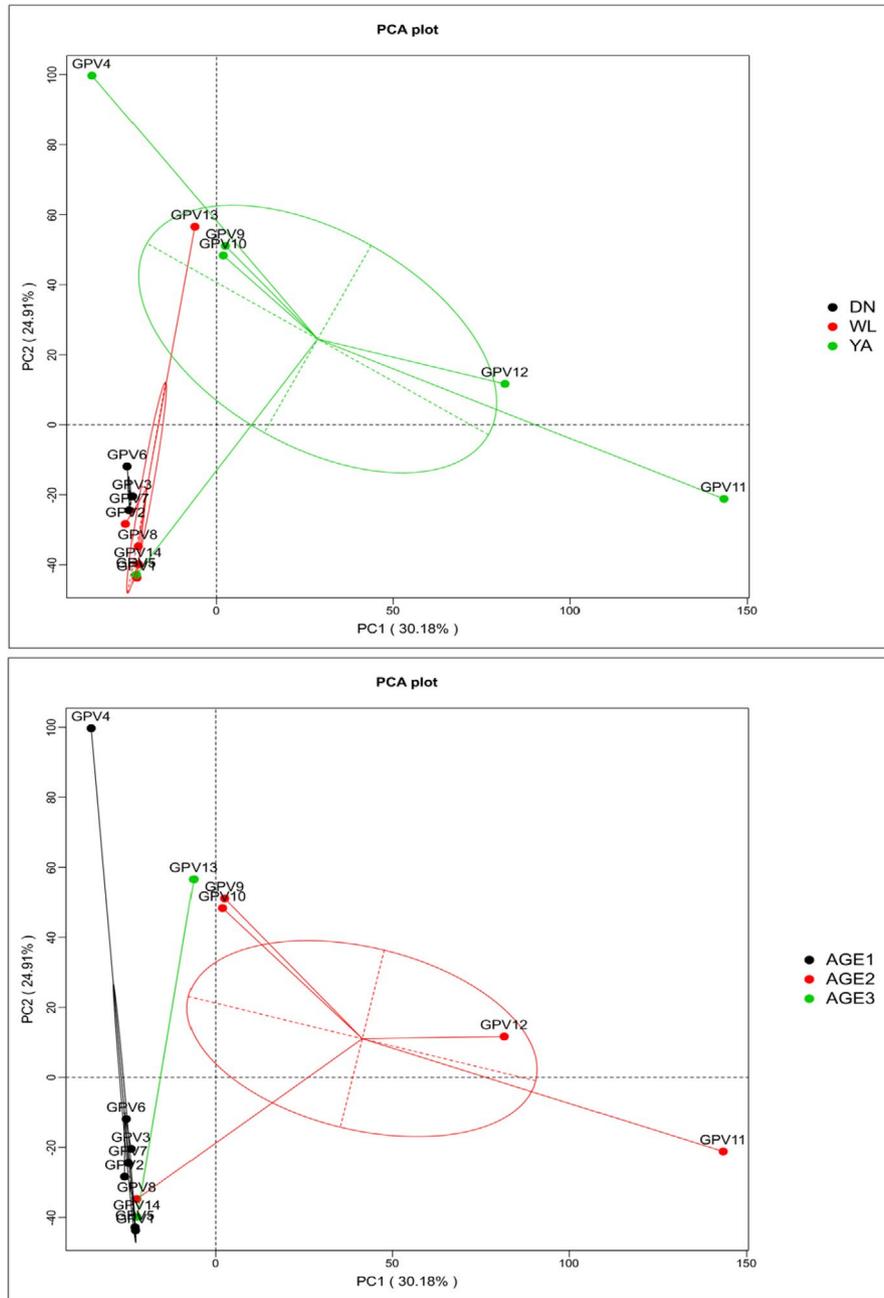


Figure A3 PCA analysis diagram based on functional abundance. PCA results show that the abscissa represents the first principal component and the percentage represents the contribution of the first principal component to the sample difference. The vertical axis represents the second principal component, and the percentage represents the contribution of the second principal component to the sample difference. Each dot in the diagram represents a sample, and the samples in the same group are shown in the same color.