



OPEN

A database of animal metagenomes

DATA DESCRIPTOR

Ruirui Hu^{1,2,5}, Rui Yao^{2,5}, Lei Li^{2,5}, Yueren Xu², Bingbing Lei², Guohao Tang³, Haowei Liang³, Yunjiao Lei², Cunyuan Li^{2,4}, Xiaoyue Li², Kaiping Liu², Limin Wang¹, Yunfeng Zhang¹, Yue Wang², Yuying Cui², Jihong Dai², Wei Ni², Ping Zhou¹✉, Baohua Yu³✉ & Shengwei Hu^{1,2}✉

With the rapid development of high-throughput sequencing technology, the amount of metagenomic data (including both 16S and whole-genome sequencing data) in public repositories is increasing exponentially. However, owing to the large and decentralized nature of the data, it is still difficult for users to mine, compare, and analyze the data. The animal metagenome database (AnimalMetagenome DB) integrates metagenomic sequencing data with host information, making it easier for users to find data of interest. The AnimalMetagenome DB is designed to contain all public metagenomic data from animals, and the data are divided into domestic and wild animal categories. Users can browse, search, and download animal metagenomic data of interest based on different attributes of the metadata such as animal species, sample site, study purpose, and DNA extraction method. The AnimalMetagenome DB version 1.0 includes metadata for 82,097 metagenomes from 4 domestic animals (pigs, bovines, horses, and sheep) and 540 wild animals. These metagenomes cover 15 years of experiments, 73 countries, 1,044 studies, 63,214 amplicon sequencing data, and 10,672 whole genome sequencing data. All data in the database are hosted and available in figshare <https://doi.org/10.6084/m9.figshare.19728619>.

Background & Summary

Microorganisms play essential roles in specialized niches, including internal host biology as well as the external environment. Microbes are found in diverse habitats, including deep seas, saline marshes, and glaciers¹. The roles of microorganisms in biodiversity have become a focus of interest for researchers due to their omnipresence². This diversity represents a vast genetic resource that could be exploited for the discovery of novel genes, biomolecules for metabolic pathways, and potentially valuable end-products³. The structure and function of the microbial community has received significant attention for decades, notably in association with research concerning human microbiota⁴. However, veterinarians, animal nutritionists, and microbiologists have begun to focus on studying the microbes of domestic (horses, pigs, and ruminants) and wild animals⁵. For domestic animals, a better understanding of disease-causing microbes of livestock can contribute to achieving the goals of better foods and a cleaner environment⁶. This is not only conducive to the healthy development of domestic animal and poultry breeding industries but also reduces the risk of food-borne diseases being transmitted to humans, which is conducive to public health security⁷. Regarding wild animals, wildlife microbiota are also natural hosts for animal and human pathogens; mapping their distributions can shed light on the timing and pathways of their transmission to humans, as is the case in the current COVID-19 pandemic⁸.

With the development of ultra-high throughput metagenomic sequencing technologies, including 16S rRNA gene sequencing and whole-genome sequencing, the number and scope of metagenomic sequencing projects have increased rapidly⁹. This has led to an exponential growth of metagenomics data under different experimental conditions. Therefore, metagenomic data contain an overwhelming volume of complex information, posing challenges not only for data storage but also for metadata annotation and management. Several pioneering studies have been designed to construct resources for storing raw sequencing data, including the National Center for Biotechnology Information (NCBI)¹⁰, the Sequence Read Archive (SRA)¹¹, and the European Nucleotide Archive (ENA)¹². These public resources contain human and animal metagenomic information that will serve as an important reference for current studies¹³. For instance, a previous study¹⁴ meta-analyzed 20 publicly available datasets from 16S rRNA gene-sequencing studies of the swine gut microbiota and demonstrated that GI tract

¹State Key Laboratory of Sheep Genetic Improvement and Healthy Production, Xinjiang Academy of Agricultural and Reclamation Science, Shihezi, Xinjiang, China. ²College of Life Sciences, Shihezi University, Shihezi, Xinjiang, 832003, China. ³College of Information Science and Technology, Shihezi University, Shihezi, Xinjiang, 832003, China. ⁴College of Animal Science and Technology, Shihezi University, Shihezi, Xinjiang, 832003, China. ⁵These authors contributed equally: Ruirui Hu, Rui Yao, Lei Li. ✉e-mail: zhpxqf@163.com; 1498322833@qq.com; hushengwei@163.com