



OPEN

Biological composition analysis of a natural medicine, Faeces Vespertilionis, with complex sources using DNA metabarcoding

Xiaoying Zhang, Wenxiu Wang, Xiaolei Yu, Yuxia Liu, Wenhui Li, Hongxia Yang, Ying Cui[✉] & Xiaoxuan Tian[✉]

Faeces *Vespertilionis* is a commonly used fecal traditional Chinese medicine. Traditionally, it is identified relying only on morphological characters. This poses a serious challenge to the composition analysis accuracy of this complex biological mixture. Thus, for quality control purposes, an accurate and effective method should be provided for taxonomic identification of Faeces *Vespertilionis*. In this study, 26 samples of Faeces *Vespertilionis* from ten provinces in China were tested using DNA metabarcoding. Seven operational taxonomic units (OTUs) were detected as belonging to bats. Among them, *Hipposideros armiger* (Hodgson, 1835) and *Rhinolophus ferrumequinum* (Schober and Grimberger, 1997) were the main host sources of Faeces *Vespertilionis* samples, with average relative abundances of 59.3% and 24.1%, respectively. Biodiversity analysis showed that Diptera and Lepidoptera were the most frequently consumed insects. At the species level, 19 taxa were clearly identified. Overall, our study used DNA metabarcoding to analyze the biological composition of Faeces *Vespertilionis*, which provides a new idea for the quality control of this special traditional Chinese medicine.

Traditional Chinese medicine is one of the oldest systems of traditional medicine, which originated more than 2500 years ago. Since the publication of the Yellow Emperor's Classic of Internal Medicine ("Huangdi Neijing" in Chinese) in the Han dynasty¹, many classical medical books and pharmacopeia in China have contained the prescriptions and clinical application of medicines with complex sources^{2,3}. Some medicines, such as Faeces *Vespertilionis*, Faeces *Troglodyterii*⁴ (the dry faeces of *Troglodytes xanthipes* Milne-Edwards, 1867), Faeces *Leporum*, Silkworm Faeces⁵ (the dry faeces of *Bombyx mori* Linnaeus, 1758), and *Cordyceps*⁶ (a complex mixture of *Cordyceps sinensis* (BerK.) Sacc. and larvae of the family Hepialidae), have been used currently. Faeces *Vespertilionis* is the dried excrement of bats, which can be collected in the wild all year round as a commercial medicinal material. After removing the sediment, removing impurities, and drying, Faeces *Vespertilionis* is usually used to improve eyesight, promote blood circulation, promote digestion and alleviate food retention in clinical practice^{3,7-9}.

Although some authors have argued that the diversity of hosts and prey might affect Faeces *Vespertilionis* quality and efficacy^{2,10,11}, the species involved in this medicine have not been precisely restricted by pharmacopeia or other practical criteria thus far. In appearance, Faeces *Vespertilionis* has been described as oblong, slightly pointed at both ends, easy to break, and able to be crumbled into small particles or powder². All of these characteristics are shared by most bat faeces. Although previous studies have shown that approximately 100 species, 24 genera and 6 families of Chiroptera distributed in China could be taken as the origin of Faeces *Vespertilionis*, there has been no survey of the actual sources of this fecal traditional Chinese medicine¹². It is worth noting that bats are hosts to a broad range of different viruses, therein also comprising a reservoir of potential zoonotic pathogens. In particular, the spread of coronavirus disease 2019 (COVID-19) has aroused people's attention to the hosts of Faeces *Vespertilionis*¹³. On the other hand, the diet of bats has been regarded as the direct source of medicinal effects of Faeces *Vespertilionis* in previous reports^{2,10,11}. These bats mainly feed on nocturnal flying insects² belonging to Lepidoptera, Diptera, Hymenoptera and Coleoptera. The variety of insects makes it hard to guarantee the uniform quality of this traditional Chinese medicine.

State Key Laboratory of Component-Based Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China. ✉email: CQL8179270@126.com; tian_xiaoxuan@tjutcm.edu.cn

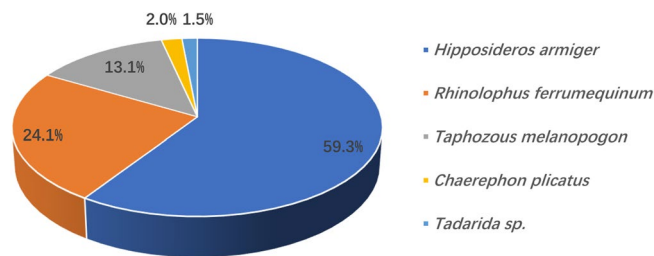


Figure 1. Taxonomic information and average relative abundance of bat species in the samples.

Unfortunately, traditional means cannot easily provide species-specific identification for this kind of complicated natural medicine¹². Broken debris of insects, such as appendages, eyes, setae and thorns¹¹, could be observed by microscope. Nevertheless, identifying hosts or other biological composition at the species level is impracticable. Moreover, no species-specific compounds of Faeces Vespertilionis have yet been reported. In the natural medicine identification field, DNA barcoding has emerged as a new molecular tool for species identification¹⁴. Furthermore, DNA barcoding technology is able to identify traditional medicines with complex species sources¹⁵. Therefore, there is an urgent need to develop novel diagnostic tools to solve this problem.

DNA metabarcoding has revolutionized species identification methods¹⁶. This method is based on the high-throughput sequencing (HTS) of DNA barcode regions, amplified using universal polymerase chain reaction (PCR) primers¹⁷. Currently, HTS technologies can generate millions of sequences concurrently. Therefore, DNA metabarcoding enables us to quickly characterize a very large number of species present in an environmental sample in a single experiment^{16,18,19}. Dietary analyses have been facilitated by the advent of metabarcoding and its application to the analysis of faeces or stomach contents²⁰, especially for the analysis of the diet of insectivorous bats^{21,22}. It is feasible that the biological composition of Faeces Vespertilionis could be discovered by metabarcoding, regardless of whether diversity is from the diet or species introduced due to collection, storage or transportation.

To our knowledge, this report is the first application of metabarcoding to depict the biological composition of a natural medicine with complex sources. Twenty-six Faeces Vespertilionis samples from ten provinces in China were selected, and the taxa involved were explored. We aimed to identify the bat hosts and other biological compositions of Faeces Vespertilionis, which should be an innovation for the quality control of this kind of fecal medicines.

Results

Sequencing results and data filtering. The Illumina sequencing of 26 PCR products generated a total of 17,794,946 paired-end reads. A total of 12,175,994 reads were obtained after filtering out low-quality reads, trimming tags and primers, and merging the paired-end reads. After retaining 232 bp sequences, dereplicating reads and removing chimeras, unique sequences were clustered at a 98% similarity threshold. Then, operational taxonomic units (OTUs) with a relative abundance greater than 0.01% of total reads of the 26 samples were retained. Finally, 243 OTUs and 4,758,016 reads remained. Rarefaction curves analysis indicated that the number of sequences and sequencing depth were sufficient for this study (Fig. S1).

Taxonomic identification of bats. We identified five bat species comprising 1,037,606 reads (Table S1). Among them, the average relative abundance (i.e., the total relative abundance of each taxon across all samples divided by 26) of *H. armiger* accounted for 59.3% of the 26 Faeces Vespertilionis samples, followed by *R. ferrumequinum* (24.1%) (Fig. 1). Two or more bat species could be identified in each sample, with *H. armiger* and *R. ferrumequinum* being the main host types within most collected samples in this study (Fig. 2).

Biodiversity analysis. Principal component analysis (PCA)²³ and non-metric multidimensional scaling (NMDS)²⁴ were utilized to calculate beta diversity and to analyze the differences between samples (Fig. 3 and Fig. S2). The PCA results showed that the samples were clustered into three groups. The first was composed of yms1 from HUBEI Province and yms14 from SICHUAN Province. The second group comprised yms7 from JIANGXI Province, yms16 from SICHUAN Province, yms17 from SHANGDONG Province, yms21 from ZHEJIANG Province, and yms25 from ZHEJIANG Province. The remaining samples formed the third group. Both the PCA and NMDS results implied that the origin was not the main explanation for the biodiversity variation between samples.

The alpha diversity²⁵ of each sample, calculated with observed OTUs, and the Shannon entropy index were used to evaluate the biodiversity within samples. The observed OTUs constituted the number of different OTUs observed within an individual sample. The Shannon entropy index indicated the richness and evenness of the species present. As shown in Fig. 4, approximately 150–230 OTUs belonging to invertebrates could be observed in every 10 g of sample. Although the number of OTUs observed varied from sample to sample, the distribution of Shannon entropy was mainly between 2 and 3 when considering OTU evenness. It is worth noting that although samples yms1 and yms14 had relatively low diversity, the former had fewer OTUs, and the latter presented a more unbalanced species distribution. Both indices were useful to assess the complexity of Faeces Vespertilionis.

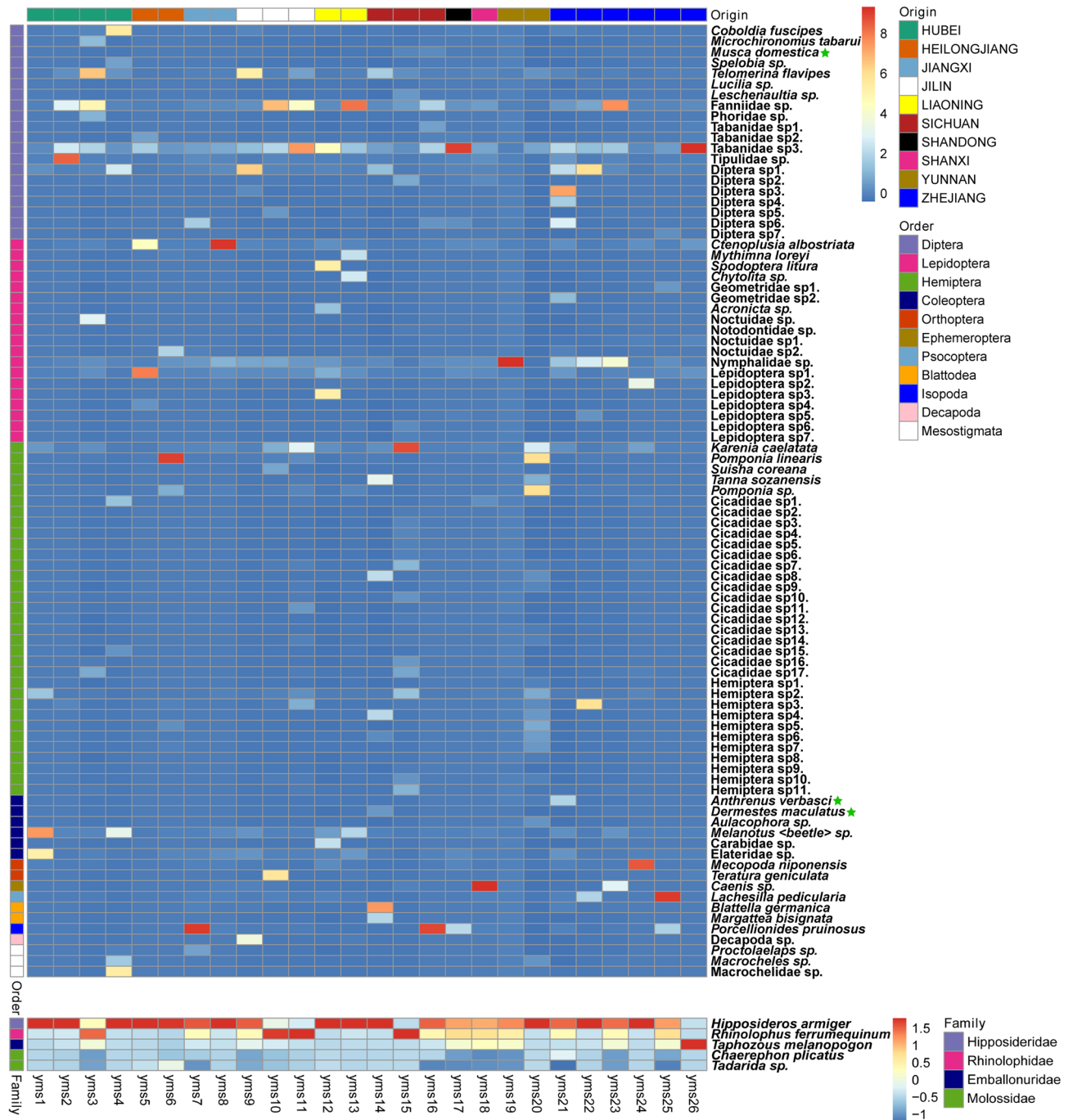


Figure 2. Detection of species in each Faeces Vespertilionis sample. Heatmap was plotted using the package pheatmap (v1.0.12) in R (v4.1.0). The upper part reflects the biological composition in the samples of Faeces Vespertilionis. Species were categorized at the order level, and samples were grouped by sample origin. The asterisks indicate that the species might have been introduced due to hygiene problems. Moreover, the lower part reflects the composition of bat species in each sample. Species were categorized at the family level.

Taxonomic identification of biological composition. In further evaluation of biological composition, 94 OTUs were identified at the order level. Among them, 68 were further identified at the family level and assigned to 27 taxa. Twenty-four were identified at the species level and assigned to 19 taxa (Table S2). The OTUs identified were divided into three classes (Fig. 5a) and 11 orders (Fig. 5b), of which 89 OTUs with 543,376 reads belonged to Insecta. The result was clearly dominated by Diptera OTUs. Lepidoptera was the second most abundant, followed by Hemiptera. Several taxa within Coleoptera, Orthoptera, Ephemeroptera, Psocoptera and Blattodea were also recorded at a much lower frequency (Fig. 5b). In contrast to most samples, Coleoptera accounted for the largest proportion in yms1, while Blattodea accounted for a larger proportion in yms14 (Fig. 5c).

Tabanidae sp3. represented a significant proportion of Diptera OTUs, while Nymphalidae sp. was representative of Lepidoptera (Fig. 2—upper panel). In addition, several Cicadidae OTUs were also detected, and Cicadae

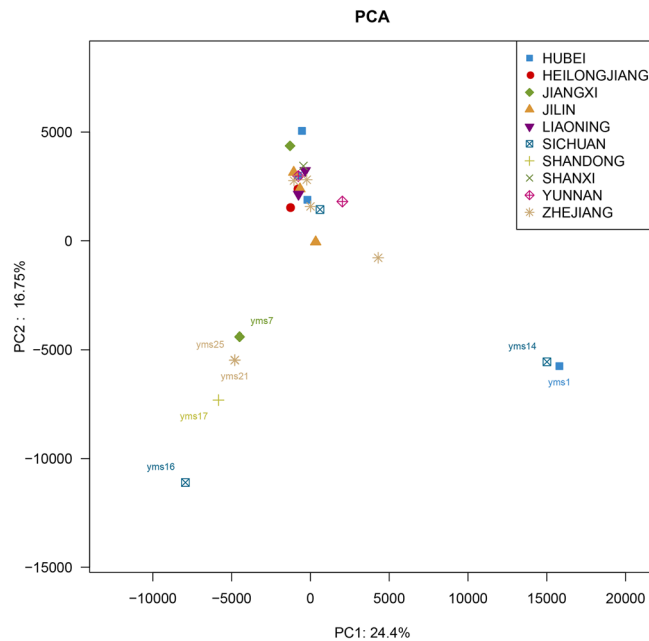


Figure 3. PCA. The default parameters of R (v2.12.1) were used for PCA and data visualization. Points of different colors or shapes represent samples of *Faeces Vespertilionis* from different origins.

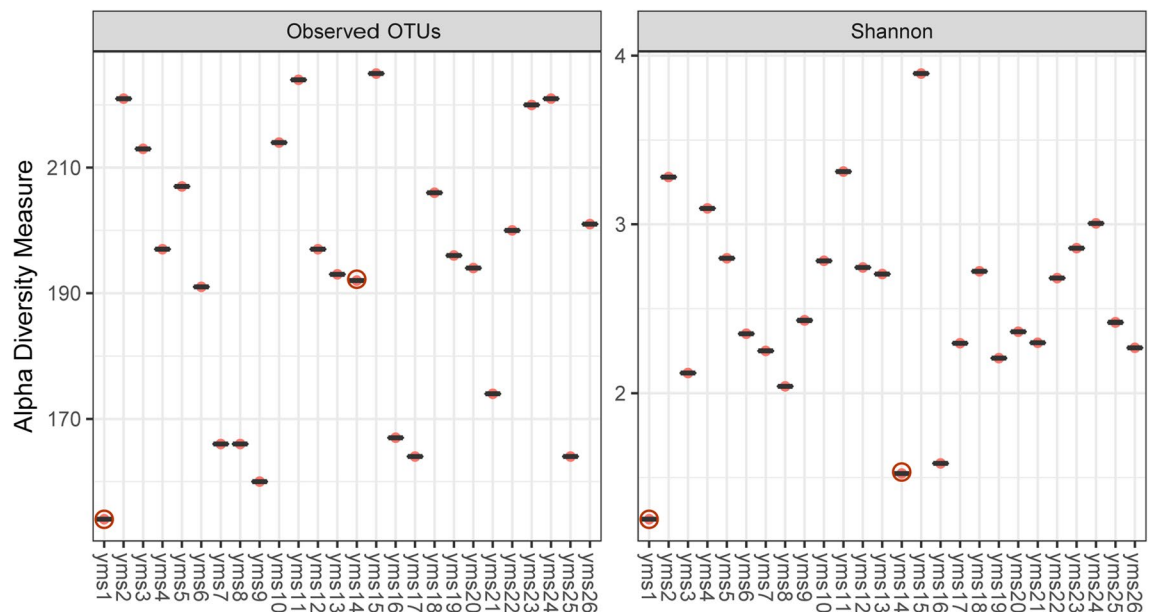


Figure 4. Alpha diversity analysis within each sample. The phyloseq (v1.36.0) and ggplot2 (v3.3.3) packages in R (v4.1.0) were used for alpha diversity analysis and data visualization. The red circles highlight that samples are different from others.

Periostracum is a traditional Chinese medicine traditionally has been used to improve eyesight and treat colds with fever².

Arachnida (11,222 reads) and Malacostraca (13,263 reads) species were also detected. We identified Arachnida in yms4 and Malacostraca in yms7, yms9, yms16, yms17, yms21, and yms25 (Figs. 2, 5c). The Arachnida OTUs included 3 mites that are often predatory on insects or nematodes (*Proctolaelaps* sp., *Macrocheles* sp., and *Macrochelidae* sp.). The cosmopolitan woodlouse *Porcellionides pruinosus* (Brandt, 1833) and *Decapoda* sp. corresponded to the Malacostraca OTUs.

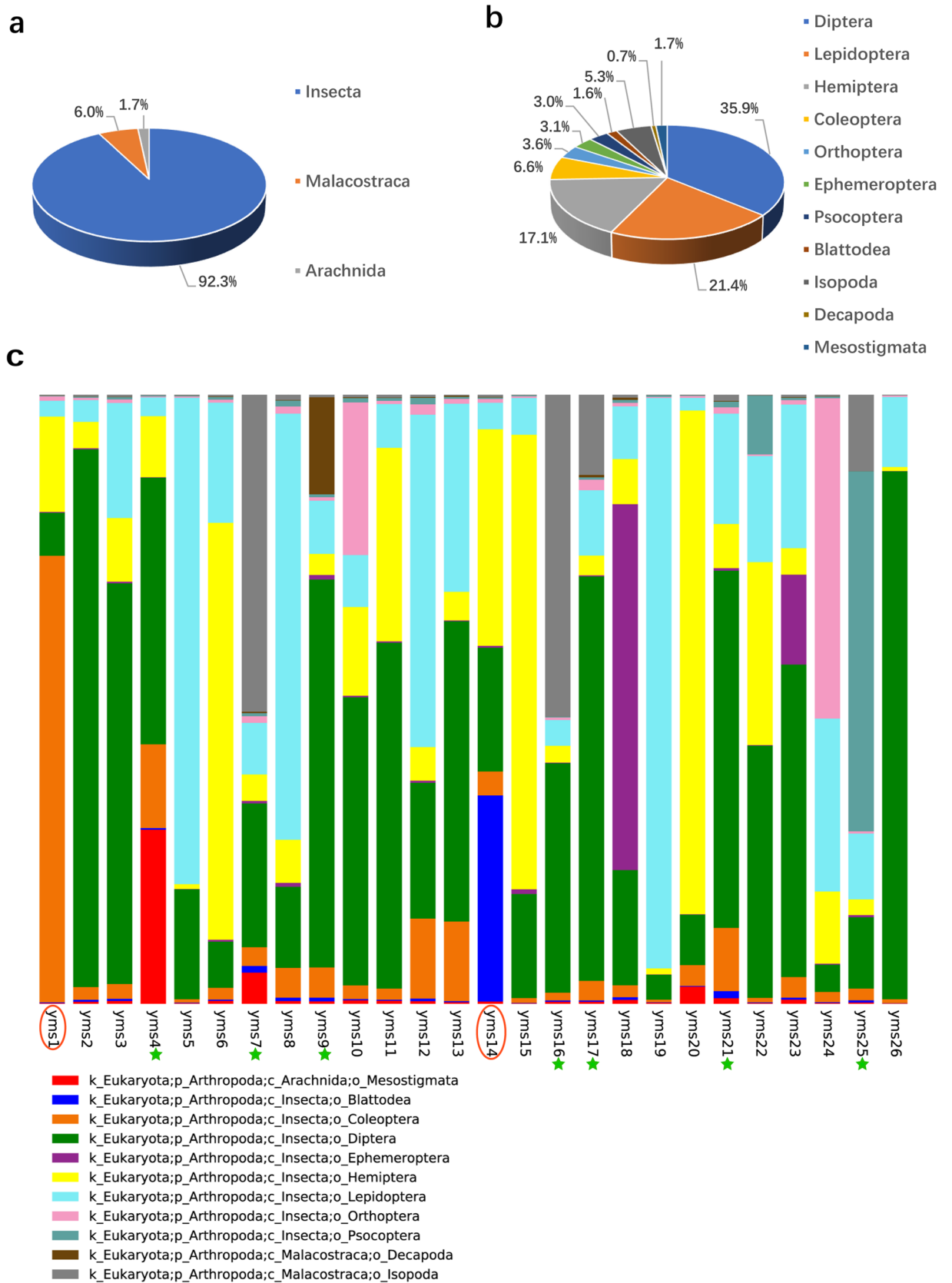


Figure 5. Taxonomic composition analysis. (a) Taxa at the class level. (b) Taxa at the order level. (c) Taxonomic composition analysis at the order level in each sample. The columns of different colors represent different taxa, and the lengths of the columns represent the relative abundances of the taxa. The red circles highlight that samples are different from others, and the asterisks indicate the samples containing non-Insecta taxa.

Origin	Sample ID					
Hubei	yms1	yms2	yms3	yms4		
Heilongjiang	yms5	yms6				
Jiangxi	yms7	yms8				
Jilin	yms9	yms10	yms11			
Lilaoning	yms12	yms13				
Sichuan	yms14	yms15	yms16			
Shangdong	yms17					
Shanxi	yms18					
Yunnan	yms19	yms20				
Zhejiang	yms21	yms22	yms23	yms24	yms25	yms26

Table 1. Sample origin information.

Discussion

As hypothesized, DNA metabarcoding was a powerful method for host identification. This is the first report that describes two bat species (*H. armiger* and *R. ferrumequinum*) as the main *Faeces Vespertilionis* hosts. *H. armiger*²⁶ and *R. ferrumequinum*²⁷ live in groups in caves, abandoned tunnels, roofs or abandoned houses. They are widely distributed in China and mainly prey on nocturnal flying insects, such as Diptera and Lepidoptera.

Our approach provided a detailed biological composition of the samples. At the order level, our results were congruent with previous studies, since Diptera and Lepidoptera were frequently observed in fecal samples^{28,29}. In fact, these two orders constituted the majority of samples in the whole study, regardless of read counts or OTU number. The non-Insecta OTUs, first reported by us, suggests that a broader range of taxa within *Faeces Vespertilionis* can be identified using DNA metabarcoding than by traditional identification methods. The three mites parasitizing insects or nematodes discovered also indicate the sensitivity of this method. It is noteworthy that the presence of taxa with bioactivity (e.g., Cicadae Periostracum) that we detected might be fundamental to *Faeces Vespertilionis* efficacy.

To ensure species-level taxonomy assignment accuracy, the following aspects were considered. First, our study utilized a portion of the mitochondrial cytochrome c oxidase subunit I (COI) “Folmer” region. This region is a widely applied animal barcode^{30,31}, especially for the identification of bat species and their diet^{21,32}. Second, since the balance between fidelity and amplification success rate must be considered for polymerase selection during biological composition analysis experiments, Tks Gflex DNA Polymerase, which has been successfully used in similar studies, was adopted in our research^{33,34}. Finally, the bioinformatics processing parameters we used have been adopted by previous related studies^{31,35,36}. In addition, we manually checked the taxonomic assignment results in MEGAN and investigated the geographical distribution and life history of taxa to ensure reliable species-level identification. Additionally, although we evaluated our selected primers through NCBI Primer-BLAST and found that the primers could amplify different Chiroptera families and different Insecta orders, metabarcoding markers (including the COI “Folmer” region) may have skewed our results due to the existence of primer bias^{37,38}.

The stability of traditional medicines benefits from controlling the biological composition within them³⁹. However, the origins of *Faeces Vespertilionis* samples are not related to bat species or other biological compositions in their faeces, which suggests that the quality of this medical material could not be controlled by only limiting its sources. However, from the perspective of diversity, the samples were indeed clustered into three groups, which might be related to the presence of other orders from Insecta, such as Coleoptera and Blattodea, and might also be related to non-Insecta taxa. Regarding the abundant biodiversity of the samples, the sample size should be increased in the future to find key factors affecting the stability of the biological level of the samples to control the biodiversity and ensure the stability and uniformity of medicine quality.

Our results shed light on the biological composition of this commercial medicinal material, in which both dietary species and introduced species due to hygiene problems were observed. In our study, *Ctenoplusia albostriata* (Bremer & Grey, 1853)⁴⁰ from Lepidoptera was found to be distributed in all samples from ten provinces in China. *Coboldia fuscipes* (Meigen, 1830)⁴¹, an oyster mushroom fly from Diptera, was also found to be distributed in all samples from ten provinces in China. Additionally, the housefly *Musca domestica* (Linnaeus, 1758)^{42,43}, the carpet beetle *Anthrenus verbasci* (Linnaeus, 1767)⁴⁴, and the carrion beetle *Dermestes maculatus* (De Geer, 1774)⁴⁵ were detected in our samples, which might have been introduced due to hygiene problems.

In summary, exemplified by *Faeces Vespertilionis*, we have shown that the DNA metabarcoding approach is a practical solution to explore the species composition of traditional medicines with highly complex components. The quality control of natural medicines might benefit from our explorative research.

Methods

Samples and DNA extraction. In this experiment, we collected 26 samples of *Faeces Vespertilionis* from ten provinces in China (Table 1). Approximately 10 g of each sample was lysed with 80 ml of 1.5% sodium dodecyl sulfate, and the supernatant was collected. A TIANamp Stool DNA Kit (Tiangen Biotech Co., Ltd., Beijing, China) was used to extract genomic DNA according to the manufacturer’s protocol. Before each extraction, the

equipment was sterilized under ultraviolet (UV) light to prevent contamination from foreign DNA and PCR products²⁹.

PCR and high-throughput sequencing. The primers used in this study were an existing detection system⁴⁶ based on the mitochondrial COI sequence. Metabarcoding studies on bulk collections of animals usually target a region within the COI “Folmer” region^{30,47,48}. Therefore, the forward primer we used was LCO1490 (5'-GGTCAACAAATCATAAAGATATTGG-3'), and the reverse primer was HCO1777 (5'-ACTTATATTGTT TATACGAGGGAA-3')⁴⁶. Both primers were tagged with 8 bp tags at the 5' end to distinguish samples during data analysis (Table S3), and the final amplified product was 297 bp. PCR amplification was performed using Tks Gflex DNA Polymerase (Takara Bio Inc.). These PCRs were conducted in a 50 µl reaction volume containing 25 µl of 2× Gflex buffer, 2 µl of DNA, 1 µl each of the forward and reverse primers, 20 µl of ddH₂O, and 1 µl of Tks Gflex DNA Polymerase. The PCR conditions consisted of an initial denaturation step at 98 °C for 1 min, followed by 40 cycles of denaturation at 98 °C for 10 s, annealing at 45 °C for 15 s, and extension at 68 °C for 30 s, with a final extension step at 68 °C for 5 min. To avoid contamination, all PCR steps were conducted in a sterile laminar flow hood that was physically separated from locations where DNA extraction or post-PCR sample processing occurred. The laminar flow hood was treated with UV light for 30 min prior to PCR preparation. We included negative controls at the DNA extraction and PCR steps. PCR products were visualized with 1% agarose gel electrophoresis. The PCR product (297 bp) was selected by excision on agarose gel, and non-specific PCR products and primer dimers were discarded. Subsequently, PCR products were mixed in equimolar amounts. A sequencing library was generated using an NEBNext Ultra™ DNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's recommendations. The Illumina adapter sequences were 5'-AATGATACGGCG ACCACCGAGATCTACACATCGGAGATCTTTCCCTACACGACGCTCTTCCGATCT-3' and 5'-GATCGG AAGAGCACACGTCTGAACTCCAGTCACTCGACATCATCTCGTATGCCGTCTTCTGCTTG-3'. Sequencing was performed on the Illumina NovaSeq platform, and the read length was 2 × 250 bp.

Bioinformatics processing. After quality control by FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/), the sequencing reads were trimmed using Trimmomatic⁴⁹, a trimming tool for Illumina sequencing data, to filter out the contaminating adapter sequence and low-quality reads (phred quality < 20). Sequencing reads were demultiplexed using fastq-multx⁵⁰ and assigned to each sample according to the unique tags. Primer and tag sequences were trimmed using bbdut from BBDMap tools⁵¹. Paired-end reads were merged using QIIME⁵². We dereplicated reads using the USEARCH^{53–55} “fastx_uniques” algorithm with the option “minuniquesize 12”. Then, we applied the USEARCH UNOISE3³⁶ algorithm to detect and remove chimeras, substitutions due to incorrect base calls and gaps due to omitted or spurious base calls. USEARCH was used to cluster OTUs at a 98% similarity³⁵ threshold. Finally, OTUs with a relative abundance less than 0.01% of total reads of the 26 samples were removed from the OTU table using QIIME (filter_otus_from_otu_table.py). A representative sequence from each OTU was then picked.

Taxonomic assignment. We imported the OTU representative sequences into Geneious Prime 2020.2 and translated them into their amino acid sequences. After translation, we removed the sequences containing stop codons, which were regarded as artificial sequences. Then, the OTU table and representative sequences were regenerated. BLASTN⁵⁶ was used to compare the OTU representative sequences against the NCBI GENBANK database, and the output was imported into MEGAN COMMUNITY EDITION version 6.10.8⁵⁷. The representative sequences were filtered with identity values of 80% to remove non-metazoans. Order-level taxonomy was assigned at > 95% identity values, family-level taxonomy was assigned at > 96.5%, and species-level taxonomy was assigned when the identity values between the query and reference sequences were above 98%^{31,35}. Finally, we manually checked the taxonomic assignment and investigated the distribution and life history of the identified species.

Statistical analysis. We respectively rarefied (QIIME script-single_rarefaction.py⁵²) the number of sequences per sample in the OTU table according to the sample with the smallest reads of bats and other biological compositions in their faeces to ensure comparability. Then, biodiversity analysis was performed after removing the host and non-metazoan sequences. The default parameters of R (v2.12.1) and the R package vegan (v2.0-1)⁵⁸ were used for PCA, NMDS analysis and data visualization. The default parameters of the R packages phyloseq (v1.36.0)⁵⁹ and ggplot2 (v3.3.3)⁶⁰ were used for alpha diversity analysis and data visualization. Heatmap was plotted using the R package pheatmap (v1.0.12)⁶¹. We calculated the relative abundance (i.e., the value of each taxon divided by the total reads per sample) and the average relative abundance (i.e., the total relative abundance of each taxon across all samples divided by 26) for sample composition analyses⁶².

Data availability

Illumina data sets have been deposited in the Sequence Read Archive (SRA) under BioProject PRJNA688101.

Received: 13 January 2021; Accepted: 20 December 2021

Published online: 10 January 2022

References

1. Fu, J. & Yang, M. *Yellow Emperor's Classic of Medicine, The—Essential Questions: Translation Of Huangdi Neijing Suwen* (World Scientific Publishing Company, 2019).
2. Li, J., Huang, L. & Qu, X. *Medicinal Fauna of China* (Fujian Science & Technology Publishing House, 2013).

3. Yang, S. *The Divine Farmer's Materia Medica: A Translation of the Shen Nong Ben Cao Jing* (Blue Poppy Press, 1998).
4. Zhao, J. *et al.* Diterpenoids from the feces of *Trogopterus xanthipes*. *J. Nat. Prod.* **73**, 865–869. <https://doi.org/10.1021/np900814a> (2010).
5. Mei, H. *et al.* Possible mechanisms by which silkworm faeces extract ameliorates adenine-induced renal anaemia in rats. *J. Ethnopharmacol.* **266**, 113448. <https://doi.org/10.1016/j.jep.2020.113448> (2021).
6. Zhang, H. W. *et al.* *Cordyceps sinensis* (a traditional Chinese medicine) for treating chronic kidney disease. *Cochrane Database Syst. Rev.* <https://doi.org/10.1002/14651858.CD008353.pub2> (2014).
7. Gao, A. *et al.* Pharmaceutical research on Yemingsha (bat dung). *Jilin J. Traditional Chin. Med.* **32**, 1047–1049 (2012).
8. Zhao, X., Gu, C. & Pu, C. Astragalus, bat dung and flastem milkvetch seed in the treatment of diabetic retinopathy—Three prescription by professor TONG Xiaolin. *Jilin J. Traditional Chin. Med.* **040**, 995–997 (2020).
9. Luo, J., Gu, C. & Zhao, L. Bat dung, silkworm droppings and trogopterus dung in the alteration of intestinal flora—Three prescription by Professor TONG Xiaolin. *Jilin J. Traditional Chin. Med.* **040**, 8–11 (2020).
10. Zhang, S. & Zhu, W. Bat species and Yemingsha (bat dung) in Zhejiang Province. *J. Chin. Med. Mater.* **24**, 15–17 (2001).
11. Zhu, H., Hu, Z., Li, S. & Liao, M. Bat dung microscopic observation and preliminary measurement of Zhanjiang. *Shandong Chem. Ind.* **48**, 98–102+111 (2019).
12. Li, C. & Zhou, Y. Discussion on the status and related issues of the medicinal materials standards and processing practice of Yemingsha (bat dung). *Chin. J. New Drugs* **29**, 1851–1855 (2020).
13. Wassenaar, T. M. & Zou, Y. 2019_nCoV/SARS-CoV-2: Rapid classification of betacoronaviruses and identification of Traditional Chinese Medicine as potential origin of zoonotic coronaviruses. *Lett. Appl. Microbiol.* **70**, 342–348. <https://doi.org/10.1111/lam.13285> (2020).
14. Techen, N., Parveen, I., Pan, Z. & Khan, I. A. DNA barcoding of medicinal plant material for identification. *Curr. Opin. Biotechnol.* **25**, 103–110. <https://doi.org/10.1016/j.copbio.2013.09.010> (2014).
15. Xin, T. *et al.* Biomonitoring for traditional herbal medicinal products using DNA metabarcoding and single molecule, real-time sequencing. *Acta Pharm. Sin. B* **8**, 488–497. <https://doi.org/10.1016/j.apsb.2017.10.001> (2018).
16. Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21**, 2045–2050. <https://doi.org/10.1111/j.1365-294X.2012.05470.x> (2012).
17. Hebert, P. D., Cywinska, A., Ball, S. L. & de Waard, J. R. Biological identifications through DNA barcodes. *Proc. Biol. Sci.* **270**, 313–321. <https://doi.org/10.1098/rspb.2002.2218> (2003).
18. Binladen, J. *et al.* The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* **2**, e197. <https://doi.org/10.1371/journal.pone.0000197> (2007).
19. Bohmann, K. *et al.* Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* **29**, 358–367. <https://doi.org/10.1016/j.tree.2014.04.003> (2014).
20. Pompanon, F. *et al.* Who is eating what: Diet assessment using next generation sequencing. *Mol. Ecol.* **21**, 1931–1950. <https://doi.org/10.1111/j.1365-294X.2011.05403.x> (2012).
21. Galan, M. *et al.* Metabarcoding for the parallel identification of several hundred predators and their prey: Application to bat species diet analysis. *Mol. Ecol. Resour.* **18**, 474–489. <https://doi.org/10.1111/1755-0998.12749> (2018).
22. Swift, J. F. *et al.* Multifaceted DNA metabarcoding: Validation of a noninvasive, next-generation approach to studying bat populations. *Evol. Appl.* **11**, 1120–1138. <https://doi.org/10.1111/eva.12644> (2018).
23. Wang, Y. *et al.* Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. *Appl. Environ. Microbiol.* **78**, 8264–8271. <https://doi.org/10.1128/aem.01821-12> (2012).
24. Stefanni, S. *et al.* Multi-marker metabarcoding approach to study mesozooplankton at basin scale. *Sci. Rep.* **8**, 12085. <https://doi.org/10.1038/s41598-018-30157-7> (2018).
25. Willis, A. D. Rarefaction, alpha diversity, and statistics. *Front. Microbiol.* **10**, 2407. <https://doi.org/10.3389/fmicb.2019.02407> (2019).
26. Xu, L. *et al.* Phylogeography and population genetic structure of the great leaf-nosed bat (*Hipposideros armiger*) in China. *J. Hered.* **101**, 562–572. <https://doi.org/10.1093/jhered/esq039> (2010).
27. Schöber, W. & Grimberger, E. *The Bats of Europe & North America: Knowing Them, Identifying Them, Protecting Them* (Tfh Publications Incorporated, 1997).
28. Clare, E. L. *et al.* The diet of *Myotis lucifugus* across Canada: Assessing foraging quality and diet variability. *Mol. Ecol.* **23**, 3618–3632. <https://doi.org/10.1111/mec.12542> (2014).
29. Vesterinen, E. J. *et al.* What you need is what you eat? Prey selection by the bat *Myotis daubentonii*. *Mol. Ecol.* **25**, 1581–1594. <https://doi.org/10.1111/mec.13564> (2016).
30. Andújar, C., Arribas, P., Yu, D. W., Vogler, A. P. & Emerson, B. C. Why the COI barcode should be the community DNA metabarcode for the metazoa. *Mol. Ecol.* **27**, 3968–3975. <https://doi.org/10.1111/mec.14844> (2018).
31. Alberdi, A., Aizpurua, O., Gilbert, M. T. P. & Bohmann, K. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods Ecol. Evol.* **9**, 134–147 (2018).
32. Aldasoro, M. *et al.* Gaining ecological insight on dietary allocation among horseshoe bats through molecular primer combination. *PLoS ONE* **14**, e0220081. <https://doi.org/10.1371/journal.pone.0220081> (2019).
33. Yang, J. *et al.* *Cutibacterium acnes* (*Propionibacterium acnes*) 16S rRNA genotyping of microbial samples from possessions contributors to owner identification. *mSystems* <https://doi.org/10.1128/mSystems.00594-19> (2019).
34. Sekihara, S., Shibata, T., Hyakkendani, M. & Kawabata, S. I. RNA interference directed against the transglutaminase gene triggers dysbiosis of gut microbiota in drosophila. *J. Biol. Chem.* **291**, 25077–25087. <https://doi.org/10.1074/jbc.M116.761791> (2016).
35. Aizpurua, O. *et al.* Agriculture shapes the trophic niche of a bat preying on multiple pest arthropods across Europe: Evidence from DNA metabarcoding. *Mol. Ecol.* **27**, 815–825. <https://doi.org/10.1111/mec.14474> (2018).
36. Vesterinen, E. J., Puiisto, A. I. E., Blomberg, A. S. & Lilley, T. M. Table for five, please: Dietary partitioning in boreal bats. *Ecol. Evol.* **8**, 10914–10937. <https://doi.org/10.1002/ece3.4559> (2018).
37. Piñol, J., Mir, G., Gomez-Polo, P. & Agustí, N. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Mol. Ecol. Resour.* **15**, 819–830. <https://doi.org/10.1111/1755-0998.12355> (2015).
38. Hajibabaei, M. *et al.* A minimalist barcode can identify a specimen whose DNA is degraded: BARCODING. *Mol. Ecol. Notes* **6**, 959–964. <https://doi.org/10.1111/j.1471-8286.2006.01470.x> (2006).
39. Yang, W. *et al.* Approaches to establish Q-markers for the quality standards of traditional Chinese medicines. *Acta Pharm. Sin. B* **7**, 439–446. <https://doi.org/10.1016/j.apsb.2017.04.012> (2017).
40. Xue, S., Zhang, Y., Gao, S. & Zhang, M. Characterisation and phylogenetic analysis of the complete mitochondrial genome of *Ctenoplia albobstriata* (Lepidoptera: Noctuidae: Plusiinae). *Mitochondrial DNA B. Resour.* **4**, 3509–3510 (2019).
41. Yi, J. H., Perumalsamy, H., Sankarapandian, K., Choi, B. R. & Ahn, Y. J. Fumigant toxicity of phenylpropanoids identified in *Asarum sieboldii* aerial parts to *Lycoriella ingenua* (diptera: sciaridae) and *Coboldia fuscipes* (diptera: scatopsidae). *J. Econ. Entomol.* **108**, 1208–1214. <https://doi.org/10.1093/jee/tov064> (2015).
42. Chu, F., Jin, X. & Ma, H. Anti-diarrhea effects and identification of *Musca domestica* larvae low molecular weight peptides (LMWP). *J. Pharm. Biomed. Anal.* **173**, 162–168. <https://doi.org/10.1016/j.jpba.2019.05.032> (2019).
43. Niu, Y. *et al.* A novel bioconversion for value-added products from food waste using *Musca domestica*. *Waste Manag.* **61**, 455–460. <https://doi.org/10.1016/j.wasman.2016.10.054> (2017).

44. Miyazaki, Y., Nisimura, T. & Numata, H. Circannual rhythm in the varied carpet beetle, *Anthrenus verbasci*. *Prog. Brain Res.* **199**, 439–456. <https://doi.org/10.1016/b978-0-444-59427-3.00025-3> (2012).
45. Samish, M., Argaman, Q. & Perelman, D. Research note: The hide beetle, *Dermestes maculatus* DeGeer (Dermestidae), feeds on live turkeys. *Poult. Sci.* **71**, 388–390. <https://doi.org/10.3382/ps.0710388> (1992).
46. Brown, D. S., Jarman, S. N. & Symondson, W. O. Pyrosequencing of prey DNA in reptile faeces: Analysis of earthworm consumption by slow worms. *Mol. Ecol. Resour.* **12**, 259–266. <https://doi.org/10.1111/j.1755-0998.2011.03098.x> (2012).
47. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299 (1994).
48. Elbrecht, V. *et al.* Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ* **7**, e7745. <https://doi.org/10.7717/peerj.7745> (2019).
49. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> (2014).
50. Aronesty, E. Comparison of sequencing utility programs. *Open Bioinform. J.* **7**, 1–8 (2013).
51. Bushnell, B., Rood, J. & Singer, E. BBMerge—Accurate paired shotgun read merging via overlap. *PLoS ONE* **12**, e0185056. <https://doi.org/10.1371/journal.pone.0185056> (2017).
52. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336. <https://doi.org/10.1038/nmeth.f.303> (2010).
53. Edgar, R. C. UPPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998. <https://doi.org/10.1038/nmeth.2604> (2013).
54. Edgar, R. C. UNOISE2: Improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv*. <https://doi.org/10.1101/081257> (2016).
55. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461> (2010).
56. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402. <https://doi.org/10.1093/nar/25.17.3389> (1997).
57. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386. <https://doi.org/10.1101/gr.5969107> (2007).
58. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
59. McMurdie, P. J. & Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217. <https://doi.org/10.1371/journal.pone.0061217> (2013).
60. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis* (Springer Publishing Company, 2009).
61. Guo, R. *et al.* Guiding T lymphopoiesis from pluripotent stem cells by defined transcription factors. *Cell Res.* **30**, 21–33. <https://doi.org/10.1038/s41422-019-0251-7> (2020).
62. Duvall, C., Gibbons, S. M., Gurry, T., Irizarry, R. A. & Alm, E. J. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat. Commun.* **8**, 1784. <https://doi.org/10.1038/s41467-017-01973-8> (2017).

Acknowledgements

This work is supported by the Grants from State Key Laboratory of Component-based Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, 301617, China.

Author contributions

X.-X.T. and Y.C. designed the study; X.-Y.Z. and W.-W.W. performed the experiments; X.-Y.Z., X.-L.Y., Y.-X.L., W.-H.L. and H.-X.Y. analyzed the results; X.-Y.Z. drafted the manuscript; X.-X.T. and Y.C. revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-04387-1>.

Correspondence and requests for materials should be addressed to Y.C. or X.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022