JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

Cite This: J. Agric. Food Chem. 2018, 66, 9858−9867

pubs.acs.org/JAFC

Herbivore-Diet Analysis Based on Illumina MiSeq Sequencing: The Potential Use of an *ITS2*-Barcoding Approach to Establish Qualitative and Quantitative Predictions of Diet Composition of Mongolian Sheep

Yanping Guo,[†] Hao Zhang,[‡] Wenqing Chen,[§] and Yingjun Zhang*[,†]

[†]Key Laboratory of Grassland Management and Utilization, Ministry of Agriculture; Department of Grassland Science, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

[‡]College of Grassland Science, Gansu Agricultural University, Lanzhou 730070, Gansu, China

[§]Department of Grassland Science, College of Animal Science and Technology, Northwest A&F University, Yangling 712100, Shaanxi, China

ABSTRACT: DNA-barcoding approaches to estimate the diet compositions of grazing animals have received significant attention, and particularly when combined with next-generation sequencing, these techniques have substantially improved in recent years. In this study, the identity and species composition of plant material ingested by Mongolian sheep were estimated through the use of 350 bp *ITS2* gene sequences of the vegetation found in fecal samples. Four diets were formulated using varying amounts of eight plant species that are common in the grasslands of northern China. Sixteen Mongolian sheep were taken from pastures and randomly assigned to four groups, and each group received one of four diets. Each sheep was randomly assigned to one of 16 confinement pens and fed its respective diet for 12 consecutive days. Fecal samples were removed from each pen from days 7−12, preserved, and composited for each pen. All herbage species included in the daily diets were detected in each fecal sample, with the exception of *Phragmites australis*. Moreover, 12 additional different plant species were retrieved from feces of the experimental sheep. The obtained data provided preliminary support for the use of the *ITS2* barcode to determine which plants were consumed. Moreover, the proportions of the herbage DNA sequences recovered from sheep feces and those of the herbage masses in the daily diets did not completely match. These results indicate that the non-Gramineae DNA sequences amplified with *ITS2* primers (including those of *Chenopodium album*, *Artemisia scoparia*, *Artemisia tanacetifolia*, and *Medicago sativa*) far exceeded those of the Gramineae species (including *Leymus chinensis* and *Puccinellia distans*), which constitute the largest share of the experimental diets. A significant positive correlation (Spearman's $\rho = 0.376$, $P = 0.003$) between the actual herbage mass proportions in the experimental diets and the herbage-DNA-sequence proportions provided sufficiently favorable support for the further investigation of DNA barcoding for the quantification of plants in feces. A significant regression coefficient was found between the relative DNA-sequence proportions of *L. chinensis* ($R^2 = 0.82$, $P < 0.0001$), *P. distans* ($R^2 = 0.64$, $P = 0.0017$), and *C. album* ($R^2 = 0.98$, $P < 0.0001$) and their respective herbage mass proportions. The quantitative relationship can be expressed by the linear-regression equations $y = 0.90x - 0.22$, $y = 0.98x - 0.03$, and $y = 5.00x - 0.25$, respectively. Thus, these results demonstrate that dietary-DNA-barcoding methods exhibited potential in providing valuable quantitative information regarding food-item components. However, it should be noted that this explorative data needs to be further improved by using additional genes and by creating a sophisticated reference database, thus enhancing both quality and accuracy of the obtained results.

KEYWORDS: *ITS2 barcode, Mongolian sheep, herbivore diet, Illumina MiSeq sequencing*

INTRODUCTION

Herbage intake is strongly associated with the production performance[1] and nutritional status[2] of pasture stock. However, in nutritional studies, the precise estimation of herbage intake or diet composition of large herbivores remains difficult. Large herbivores generally consume an abundance of plants[3] but may exhibit dietary preferences for specific plants due to several factors, such as distinctive digestibility, tolerance of plant fibers or secondary metabolites, and energy demand.[4] The difference between the availability of herbage species and nutritional intake can be illustrated by the fact that grazing sheep typically consume a diet consisting of specific plant species despite being on pastures that offer a varied plant-species assemblage. These relationships between plants and herbivore production performance have been primarily investigated using animals fed in confinement, as it is more challenging to estimate the diet composition of grazing herbivores in open and diverse pastures.

Several techniques are commonly used to estimate the diet composition of pastured animals, including (1) direct observation of the consumed plants or of the animal foraging

Received: May 29, 2018
Revised: August 27, 2018
Accepted: August 30, 2018
Published: September 10, 2018

DOI: 10.1021/acs.jafc.8b02814
J. Agric. Food Chem. 2018, 66, 9858−9867

behavior in the field,[5] (2) utilization of internal markers (e.g., acid-insoluble ash, AIA),[6] (3) microscopic examination of either stomach contents or feces,[7] (4) near-infrared spectroscopy (NIRS),[8] and (5) a plant cuticular wax indicator method that uses saturated hydrocarbons (*n*-alkanes) as markers for estimating the species composition of consumed herbage.[1,9−12] All of these methods exhibit shortcomings under various conditions. For instance, Dove and Mayes[1,13,14] pointed out that the *n*-alkane approach, which had been widely accepted as the most reliable method, can only accurately distinguish among several herbage species but not among complex plant populations. Moreover, the concentrations as well as the fecal-recovery levels of *n*-alkanes differ across seasons, which affects the accuracy of composition estimation of seasonal diets. Over the past few years, DNA-barcoding studies have substantially developed toward diet-composition quantification.[15−17] Previous studies have indicated DNA barcoding as a relatively suitable tool (e.g., fast, simple, and very robust) for ascertaining the diet of herbivorous species (such as *Marmota caudata*, *Ursus arctos*, *Tetrao urogallus*, and *Chorthippus biguttulus*).[18,19] Furthermore, the practicality of modern high-throughput sequencing techniques has unlocked an array of possibilities for DNA-based dietary analysis. Moreover, improvements in public databases, which have enabled better identification of the recovered sequences, have been a further significant advancement for DNA-based diet research.[20−23]

However, unlike carnivores, genetic analyses of the diets of herbivores are subject to greater difficulties and complexities. Poinar et al.[24] amplified and sequenced a 183 bp *rbcL*-gene fragment to analyze the diet of the Gypsum Cave ground sloth. In contrast to this, Taberlet et al.[25] reported that a shorter fragment (the P6 loop, 10−143 bp) of the chloroplast *trnL* (UAA) intron can even be amplified from highly degraded DNA samples. Still, the low resolution of the whole *trnL* intron (67.3 or 72% of the identified species) as well as the P6 loop (only 19.5% identified) impeded their usage.[4,26] Chen et al. tested several candidate DNA regions (e.g., *psbA-trnH*, *matK*, *rbcL*, *ITS2*, and *ITS*) and proposed that *ITS2* can be used as a powerful universal DNA barcode for the identification of plant taxa, attributable to its high rate of successful identification (≥92.7% at the species level).[27,28] Bradley et al. further evaluated the efficacy of the *ITS2* (350 bp) region of nuclear ribosomal genes for plant identification using plant DNA obtained from feces of wild western gorillas and indicated that this approach can provide a foundational assessment of dietary variety and is appropriate for the evaluation of nutritional relationships in ecosystems.[29]

The main concern in many animal-diet-component-identification studies was whether the DNA-barcoding approach provides a sufficiently accurate estimate of the components that were consumed by the animal. However, to some extent, proportion bias is inevitable because of the variable digestibility of dietary components and the species eaten by herbivore animals. In this study, the feces of sheep housed in confinement were analyzed using an *ITS2*-barcoding approach combined with Illumina MiSeq sequencing to (1) test the discrimination ability of *ITS2* as a marker for distinguishing the numerous plant species consumed by sheep and (2) establish quantitative relationships between the recovered DNA sequences (operational taxonomic units, OTUs) and their proportions and the actual proportions of herbage species consumed by sheep.

MATERIALS AND METHODS

Feeding Experiment and Sample Collection. Sixteen Mongolian sheep (26−30 kg) were included in the experiment in July 2014 at the Guyuan National Field Research Station of the Grassland Ecosystem (China, 41° 44′ N, 115° 40′ E). All animal experiments received approval from the China Agricultural University Laboratory Animal Care Advisory committee.

A completely random experimental design was used in this study. The 16 Mongolian sheep were randomly assigned to one of four groups and then individually and randomly assigned to one of 16 custom-built pens. Each group was assigned to one of four treatment diets (listed below) and fed about 700 g (dry matter) twice daily for 12 days with sufficient water provided. Three to eight plant species that are common to the grasslands of northern China and universally eaten by sheep were included in the treatment diets: *Leymus chinensis* (LC), *Puccinellia distans* (PD), *Medicago sativa* (alfalfa), *Phragmites australis* (PA), *Chenopodium album* (CA), *Elymus nutans* (EN), *Artemisia scoparia* (AS), and *Artemisia tanacetifolia* (AT). Group 1 (sheep 1−4) was fed daily diet 1, consisting of 41.82% LC, 36.84% PD, and 21.34% alfalfa. Group 2 (sheep 5−8) was fed daily diet 2, consisting of 42.51% LC, 22.60% PD, 13.18% PA, and 21.71% alfalfa. Group 3 (sheep 9−12) was fed daily diet 3, consisting of 35.63% LC, 12.78% PD, 10.85% PA, 15.12% CA, 4.65% EN, and 20.97% alfalfa. Group 4 (sheep 13−16) was fed daily diet 4, consisting of 22.7% LC, 10.69% PD, 8.74% PA, 12.36% CA, 4.78% EN, 10.03% AS, 9.12% AT, and 21.58% alfalfa (the dry-matter intake of each group is shown in Table 1). As described in Brosh et al.,[30] the amounts fed to the sheep were adjusted during an adaptation period to ensure that less than 5% of the diets were refused.

Table 1. Dry-Matter Intake (g/kg DM) of Each Diet[a]

	diet 1	diet 2	diet 3	diet 4
LC	306.89	306.89	266.29	164.83
PD	270.35	163.1	95.51	77.63
PA	0	95.15	81.09	63.5
CA	0	0	112.99	89.84
EN	0	0	34.72	34.69
AS	0	0	0	72.82
AT	0	0	0	66.21
alfalfa	156.6	156.74	156.7	156.72
dry-matter intake (g/day)	733.84	721.88	747.29	726.24

[a]LC, *Leymus chinensis*; PD, *Puccinellia distans*; PA, *Phragmites australis*; CA, *Chenopodium album*; EN, *Elymus nutans*; AS, *Artemisia scoparia*; AT, *Artemisia tanacetifolia*; alfalfa, *Medicago sativa*.

The experiment consisted of an adaptation period of 5 days followed by a 7 day fecal-collection period. On day 6 of the feeding period, fecal-collection bags were placed at the bottoms of the pens for sampling. The total fecal output was collected at about 7:00 AM in the fecal-collection bags, which were emptied daily during the 7 day collection phase. A representative fecal sample of 20% was taken from the bags on each of the days of the collection phase and mixed as a final composite sample.[31] A total of 16 composite fecal samples were collected and stored at −20 °C for subsequent analysis. Additionally, samples of each of the four treatment diets were collected as reference samples and preserved in liquid nitrogen. However, during the sampling period, two sheep (one each from groups 2 and 4) refused to swallow sufficient plant materials, and two further sheep (one each from groups 1 and 3) got diarrhea. Therefore, these data were not included for the Illumina-sequencing procedure, resulting in a total of 12 sheep (3 in each group).

Genetic Analysis. Genetic analysis of fecal samples (n = 12) and herbage-mix samples (n = 4) included DNA extraction, amplification of the *ITS2* gene, and amplicon sequencing on the Illumina MiSeq platform.

DNA Extraction. Total DNA was extracted from sheep feces using the E.Z.N.A. Stool DNA Kit (Omega Biotek, Norcross, GA)

DOI: 10.1021/acs.jafc.8b02814
J. Agric. Food Chem. 2018, 66, 9858−9867

according to the manufacturer's protocol. The fecal samples were rapidly mashed, and approximately 200 mg of each sample was placed into a 2 mL microcentrifuge tube, containing 200 mg of Glass Beads X. Then, the tube was placed on ice. The following main extraction steps were used: (1) SLX-Mlus Buffer (540 μL) was added, and the sample was vortexed at maximum speed for 10 min or until the fecal sample was completely homogenized. (2) DS Buffer (60 μL) and Proteinase K Solution (20 μL) were added; the solution was either vortexed or pipetted up and down to achieve thorough mixing and then incubated at 70 °C for 10 min. (3) SP2 Buffer (200 μL) was added, and the solution was vortexed for 30 s and then allowed to settle on ice for 5 min. (4) The solution was centrifuged at maximum speed (≥13 000*g*) for 5 min, and then 400 μL of supernatant was carefully aspirated to a new 1.5 mL microcentrifuge tube. (5) cHTR reagent (200 μL) was added, and the solution was vortexed for 10 s. After that, the mixture was allowed to settle at room temperature for 2 min. (6) The solution was centrifuged at maximum speed for 2 min, and 250 μL of supernatant was transferred to a new 1.5 mL microcentrifuge tube. (7) BL Buffer (250 μL) and 100% ethanol (250 μL) were added, and the solution was vortexed at maximum speed for 10 s. (8) A HiBind DNA Mini Column was inserted into a 2 mL collection tube. (9) The entire sample from step 7 was transferred to the HiBind DNA Mini Column and then centrifuged at maximum speed for 1 min. Both the filtrate and the collection tube were discarded. (10) The HiBind DNA Mini Column was transferred into a new 2 mL collection tube, 500 μL of VHB Buffer was added, the tube was centrifuged at maximum speed for 30 s, the filtrate was discarded, and the collection tube was reused. (11) DNA Wash Buffer (700 μL) was added, and the tube was centrifuged for 1 min. (12) The filtrate was discarded, and the collection tube was reused. (13) Steps 11 and 12 were repeated for a second DNA Wash Buffer wash step. (14) The empty HiBind DNA Mini Column was centrifuged at maximum speed for 2 min at room temperature, and the column was transferred into a new 1.5 mL microcentrifuge tube. (15) Elution Buffer (100−200 μL) heated to 65 °C was add to the center of the HiBind matrix, and then the mixture was allowed to settle at room temperature for 2 min. (16) The tube was centrifuged at maximum speed for 1 min, and the obtained DNA was stored at −20 °C.

DNA extracts from the daily diets (herbage mixes) used 100 mg of sample and the DNAsecure Plant Kit (TIANGEN Biotech Company, Ltd., Beijing, China) according to the manufacturer's instructions. The following main extraction steps were used: (1) Samples of 100 mg of fresh plant tissue were ground with liquid nitrogen. (2) LP1 Buffer (400 μL) and RNase A (6 μL) were added, and the solution was vortexed for 1 min and allowed to settle at room temperature for 10 min. (3) LP2 Buffer (130 μL) was added and mixed thoroughly. Next, the solution was vortexed for 1 min. (4) The solution was centrifuged at a speed of 12 000 rpm (~13 400*g*) for 5 min, and the supernatant was aspirated to a new microcentrifuge tube. (5) LP3 Buffer (about 1.5 times the volume of the supernatant) was added, and the solution was immediately shaken for 15 s. (6) A Spin Column CB3 was inserted into a 2 mL collection tube, and the entire sample after step 5 was transferred to the Spin Column CB3, which was then centrifuged for 30 s. (7) The filtrate was discarded, and the collection tube was reused. (8) PW Buffer (600 μL) was added, the solution was centrifuged for 30 s, the filtrate was discarded, and the collection tube was reused. (9) Step 8 was repeated for a second DNA-washing step. (10) The empty Spin Column CB3 was centrifuged for 2 min, and the filtrate was discarded. Next, the mixture was allowed to settle at room temperature for 3 min. (11) The column was transferred into a new microcentrifuge tube, 50−200 μL of Elution Buffer TE was added, and then the mixture was allowed to settle at room temperature for 2−5 min. (12) The tube was centrifuged at a speed of 12 000 rpm for 2 min, and the DNA was stored at −20 °C. Mock extractions without samples were used to monitor contamination.

PCR Amplification of the *ITS2* Gene. The *ITS2* region of the nuclear rDNA (~350 bp) was amplified via PCR with primers rD5-ITS2 (5′-barcode-TCCTCCGCTTATTGATATGC-3′) and rb1-ITS2f (5′-CGATACTTGGTGTGAATTGCAG-3′).[29] The barcode is an 8 bp sequence that is unique to each specimen. The PCR mixture included 4 μL of 5× FastPfu Buffer, 2 μL of 2.5 mM dNTPs, 0.8 μL of each primer (5 μM), 0.4 μL of FastPfu Polymerase, 0.2 μL of bovine serum albumin (BSA), and 10 ng of template DNA in a 20 μL volume. The thermocycling conditions consisted of 94 °C for 5 min, followed by 45 cycles of denaturation at 94 °C for 30 s, annealing at 59 °C for 60 s, and elongation at 72 °C for 60 s, with a final extension step at 72 °C for 10 min. Each PCR group included blank controls (containing no DNA) to test sample contamination. All PCR products were visually assessed via electrophoresis on 2.0% agarose gels. DL2000 markers (Takara Biotechnology, Dalian, China) were used for size comparison. Positive PCR products were selected for Illumina sequencing.

Illumina Sequencing of the *ITS2* Amplicons. PCR products were extracted from 2% agarose gels and purified using the AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA) and quantified using QuantiFluor-ST (Promega Corporation, Madison WI). Purified amplicons were paired-end sequenced (2 × 250) on an Illumina MiSeq platform (Majorbio Inc., Shanghai, China).

Sequencing-Data Analysis. The raw fastq files were demultiplexed and then quality-filtered using QIIME (version 1.17, www.qiime.org). The following criteria were used: (i) The 300 bp reads were trimmed at any site that obtained a mean quality score <20 over a 50 bp sliding window; truncated reads shorter than 50 bp were discarded. (ii) Exact barcode matching was implemented, with a two-nucleotide mismatch in primer matching being specified, and reads with ambiguous nucleotides were also removed. (iii) Only sequences with 10 bp or more of overlap were assembled on the basis of their overlapping sequences. Any unassembled reads were discarded.

All sequences were sorted from the output file, using the sample-specific tag present on the 5′ end of the primers. Thus, a new data set was obtained for each sample, which constituted all the sequences in possession of the respective tag. For all sequences, OTUs were clustered with a 97% similarity cutoff using UPARSE (version 7.1, http://drive5.com/uparse/), and chimeric sequences were identified and removed using UCHIME. An OTU-distance-unit cutoff of 0.03 was used to assess both indexes to evaluate the selectivity in the OTU definitions.

Statistical Analyses. α-Diversity measures and rarefaction curves were computed using MOTHUR version 1.30.1.[32] For an estimation of botanical-species richness, which is related to the number of observed OTUs, the Chao1, ACE, and Simpson's indexes were calculated for all samples.[33] The Shannon diversity index and Good's coverage were also computed. β-Diversity analyses, including principal-component analysis (PCA) and nonmetric multidimensional scaling (NMDS), were performed using QIIME, and all figures were generated in R (version 3.2.1)[34] using the package VEGAN.[35,36] R was used to describe OTU-accumulation curves (using the "specaccum" command of the VEGAN package)[37] to ascertain whether sufficient sequencing depth had been obtained. An OTU -distance-unit cutoff of 0.03 was considered.

Taxonomic assignment of OTU representative sequences was conducted using the nucleotide BLASTn against the GenBank database for "nucleotide collection (nr/nt)" of NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi).[38] Final taxonomic classification of each OTU sequence was based on the closest blast match. Furthermore, several other factors were considered, as described in Deagle et al.,[22] including the geographical locations of the species that were present in the closest blast hit, as well as the diversity of closely related species. A high similarity threshold was set in the assignation step (>95% for species-level identification). As previously described, if the same score was assigned to two or more taxa for a given sequence or when these conditions were not met, this sequence was then assigned to a higher taxonomic level (e.g., genus or family).[18]

Irrelevant OTUs were eliminated from the OTU table prior to the diet-composition analysis. Sequences that occurred more than twice were considered as practicable information. Sequences that occurred less than twice were largely assumed to be an *ITS2*-sequencing error and were thus not considered for the following analyses. In dietary-DNA barcoding, it remains unclear whether the gene information recovered reflects the actual herbage fraction of consumed food

DOI: 10.1021/acs.jafc.8b02814
J. Agric. Food Chem. 2018, 66, 9858−9867

items.[22] Thus, it was evaluated whether the proportion of DNA sequences correlated with the actual proportion of herbage mass in the daily diets using the Spearman's correlation coefficient. The DNA-sequence numbers were log-transformed to obtain a symmetric frequency distribution.[39]

All statistical analyses were conducted using SPSS software (version 17.0). Differences between treatment groups were analyzed using parametric analysis of variance (ANOVA). A *P*-value below 0.05 was considered as indicating statistically significant differences. Linear-regression analysis was used to determine the quantitative relationships between the proportion of DNA sequences that were retrieved from sheep feces and the actual proportion of herbage species in the ingested diets.

Nucleotide-Sequence Accession Numbers. The raw Illumina MiSeq sequencing reads were submitted to the Sequencing Read Archive (SRA) database under the accession ID SRP148544.

RESULTS

Overview of the *ITS2*-Gene Illumina-Sequencing Data. A final valid sequence-data set was generated for further analysis containing 484 598 filtered, high-quality, usable *ITS2* sequences comprising 94 OTUs. A total of 361 281 sequences were recovered from 12 fecal samples, and 123 317 were obtained from the dietary herbage-tissue mixtures, ranging from 20 643 to 39 754 sequences per sample. The average length of the PCR products was 361 bp (the size range was 221−460 bp). Good's coverage, which estimates the percentage of OTUs represented in fecal or plant samples, averaged 99%.

Botanical Richness and Biodiversity. A species-accumulation (SA) curve was used to assess whether the observed botanical richness in the sample cohort was indicative of the general botanical diversity in either sheep feces or daily diets. This method determines the number of new identified OTUs and labels them when the additional samples are cumulatively added to the sequencing process. The OTU number increased rapidly between 0 and 10 samples and then stabilized once sampling was complete (Figure 1).

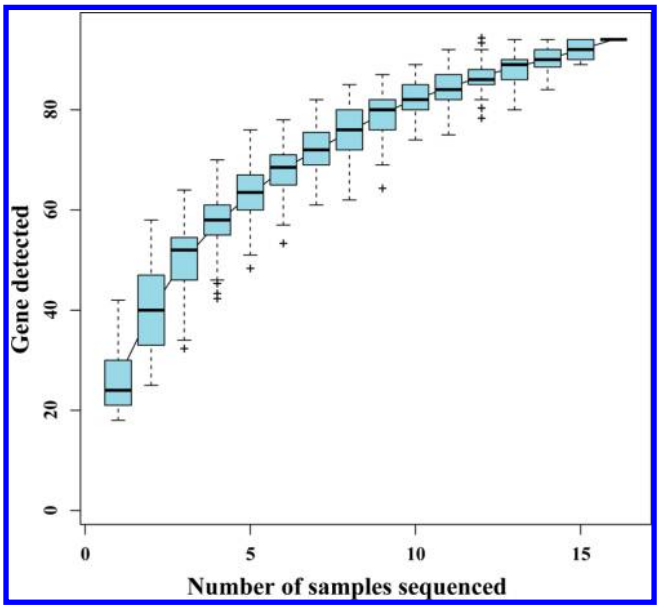

Figure 1. Species-accumulation (SA) analysis: SA plots, showing the increase in operational taxonomic units (OTUs) detected in response to the addition of each sample.

Forty-two representative OTUs were detected at most, but the botanical diversity of the fecal samples with different daily diets varied (Table 2). Moreover, the relative numbers of plant species in groups 3 and 4 were lower than those in groups 1 and 2. ANOVA identified significant differences ($P < 0.01$) in biodiversity between samples at an OTU cutoff of 0.03.

Sheep-Feces Botanical Identification. The NMDS and PCA results indicated that the identified taxa partitioned the daily diets into three distinct groups (Figure 2). At the genus level, a total of 19 floristic genera were detected in the sheep feces after disregarding fungal species (see details in Figures 3 and 4). Although a small number of fungi (0.12%) were amplified and sequenced with the *ITS2* primers, fungal sequences were excluded from the dietary analysis. Irrespective of diet, three genera accounted for 80.37% of the sequences. These included *Artemisia* (38.9%), *Chenopodium* (30.68%), and *Medicago* (10.79%). A total of 94 OTUs were identified across all sheep (Figure 4). The majority (56.38%) of these OTUs belonged to genera that appeared at abundances below 0.1%. The remaining 43.62% of the OTUs had abundances above 0.1% and accounted for 98.76% of all sequences. The most dominant OTUs across all sheep belonged to AS or AT (22.12%), followed by CA (17.54%) and alfalfa (10.14%).

Determination of Diet Composition. All treatment herbage species included in the daily diets were detected in each fecal sample, with the exception of PA (Figure 5). A total of 50 327 sequences were recovered from the herbage test species; however, the relative proportions of these sequences differed widely from each mass proportion in the daily diets (Table 3 and Figure 6). For sheep in group 1 (fed with treatment diet 1), the LC sequences contributed 15.83% of those corresponding to each sequence (vs 41.82% by mass in herbage mix), PD contributed 35.10% (vs 36.84% by mass in herbage mix), and alfalfa contributed 49.06% (vs 21.34% by mass in herbage mix). For sheep in group 2 (fed with treatment diet 2), LC sequences contributed 18.25% of those sequences (vs 42.51% by mass in herbage mix), PD contributed 29.29% (vs 22.59% by mass in herbage mix), and alfalfa contributed 52.46% (vs 21.71% by mass in herbage mix). For sheep in group 3 (fed with treatment diet 3) LC sequences contributed 5.17% of those sequences (vs 35.63% by mass in herbage mix), PD contributed 10.34% (vs 12.78% by mass in herbage mix), alfalfa contributed 32.48% (vs 20.97% by mass in herbage mix), CA contributed 46.84% (vs 15.12% by mass in herbage mix), and EN contributed 5.17% (vs 4.65% by mass in herbage mix). For sheep in group 4 (fed with treatment diet 4) PD sequences contributed 3.39% of those sequences (vs 10.69% by mass in herbage mix), alfalfa contributed 14.41% (vs 21.58% by mass in herbage mix), CA contributed 35.03% (vs 12.37% by mass in herbage mix), and AS and AT contributed 47.17% (vs 19.15% by mass in herbage mix). These results also showed that some sequences recovered from sheep feces matched other plant species, such as *Artemisia annua* (11.38%), *Potentilla anserina* (10.73%), *Sphallerocarpus gracilis* (4.06%), and *Amaranthus retroflexus* (2.98%), which are also common plants in the experimental areas.

Correlation and Quantitative Relationships. As shown in Figure 7, the proportion of log-transformed DNA sequences was significantly correlated with the proportion of herbage mass in the diets at the species level (Spearman's $\rho = 0.376$, $P = 0.003$; $n = 60$). The results of linear-regression analyses indicated a significant linear regression between herbage mass

DOI: 10.1021/acs.jafc.8b02814
J. Agric. Food Chem. 2018, 66, 9858−9867

Table 2. Richness and Diversity Indexes Relative to Each Sample: Number of Observed OTUs, Chao1 Index, and Shannon Index at an OTU Cutoff of 0.03

			α-diversity (threshold = 0.03)			
	sample ID	number of OTUs	Chao1	ACE	Shannon	Simpson
daily diets (herbage mixes)	G1	25	28	36	1.13	0.4303
	G2	24	35	36	1.27	0.437
	G3	21	27	65	0.95	0.495
	G4	19	22	24	1.43	0.2679
group 1	T1_1	31	32	23	2.62	0.0971
	T1_2	22	22	33	2.17	0.166
	T1_3	37	40	22	2.38	0.1431
group 2	T2_1	23	23	42	2.2	0.1673
	T2_2	30	31	25	2.46	0.1323
	T2_3	42	42	38	2.92	0.079
group 3	T3_1	26	26	42	1.36	0.3538
	T3_2	40	40	26	1.29	0.4611
	T3_3	19	21	40	0.89	0.6027
group 4	T4_1	18	20	26	0.74	0.5786
	T4_2	22	25	20	0.91	0.4506
	T4_3	25	30	36	0.91	0.4818
	P value	0.393	0.611	<0.01	<0.01	<0.01

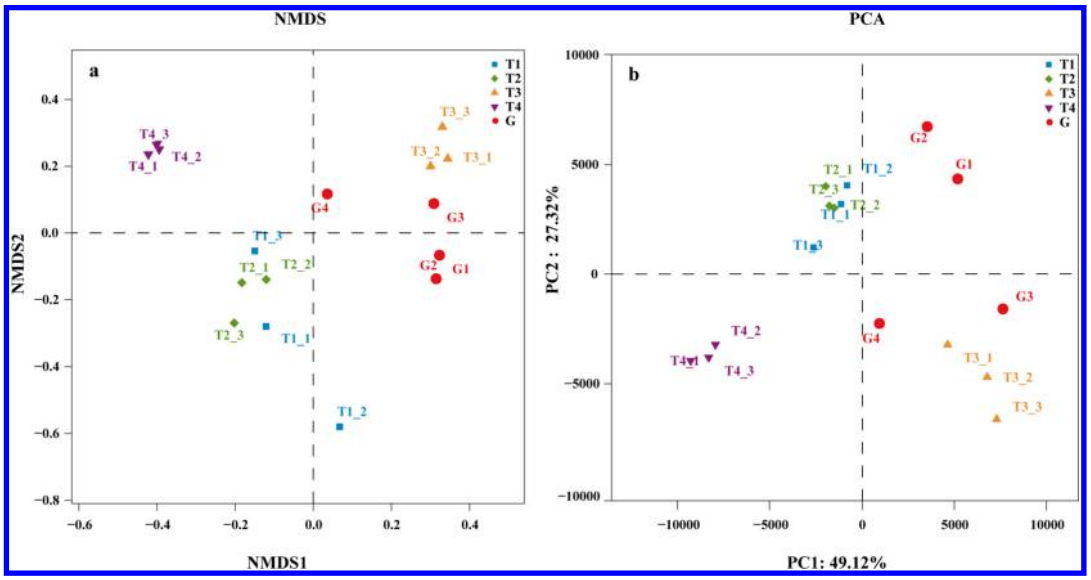

Figure 2. Nonmetric multidimensional scaling (NMDS) and principal-component analysis (PCA) scores for the dissimilarity-distance matrix between each sample. (a) NMDS plot and (b) PCA plot, highlighting the T3 (orange triangles) and T4 (purple triangles) samples. These plots were based on the weighted UniFrac distance for community dissimilarity. T1−T4, fecal metasamples of the four groups; G1−G4, four daily-diet herbage mixes.

proportions and the proportions of LC (R^2 = 0.82, P < 0.0001), PD (R^2 = 0.64, P < 0.01), and CA DNA sequences (R^2 = 0.98, P < 0.0001). The quantitative predictive relationships can be expressed via linear-regression equations $y = 0.90x - 0.22$ for LC, $y = 0.98x - 0.03$ for PD, and $y = 5.00x - 0.25$ for CA.

DISCUSSION

In this study, DNA barcoding was used in combination with Illumina sequencing of fecal-sample extracts from Mongolian sheep fed forages that are common to arid and semiarid grasslands. The goal was to ascertain the presence and amounts of the plants selected by the sheep. The study results appear to support the use of DNA barcoding for determining which plants were consumed. Furthermore, quantitative prediction models for the proportion of herbage mass consumed by sheep were established but still require further improvement.

***ITS2* Region as a DNA Barcode.** In the present study, all treatment herbage species that were included in the daily diets were detected in each fecal sample, with the exception of PA (Figure 5). Concentrating on the *ITS2* gene region, 19 different plant species were detected in the feces of the experimental sheep, all of which are common for the grasslands of northern China (Figures 3 and 4). Previously, a similar feeding trial was performed using *n*-alkanes, long-chain alcohols, and long-chain fatty acids as diet-composition markers. The results showed that the *n*-alkane technique identified a maximum of eight forage plants at the species level and exhibited limited application value for the identification of a complex grassland-species population.[40] Thus, the current

DOI: 10.1021/acs.jafc.8b02814
J. Agric. Food Chem. 2018, 66, 9858−9867

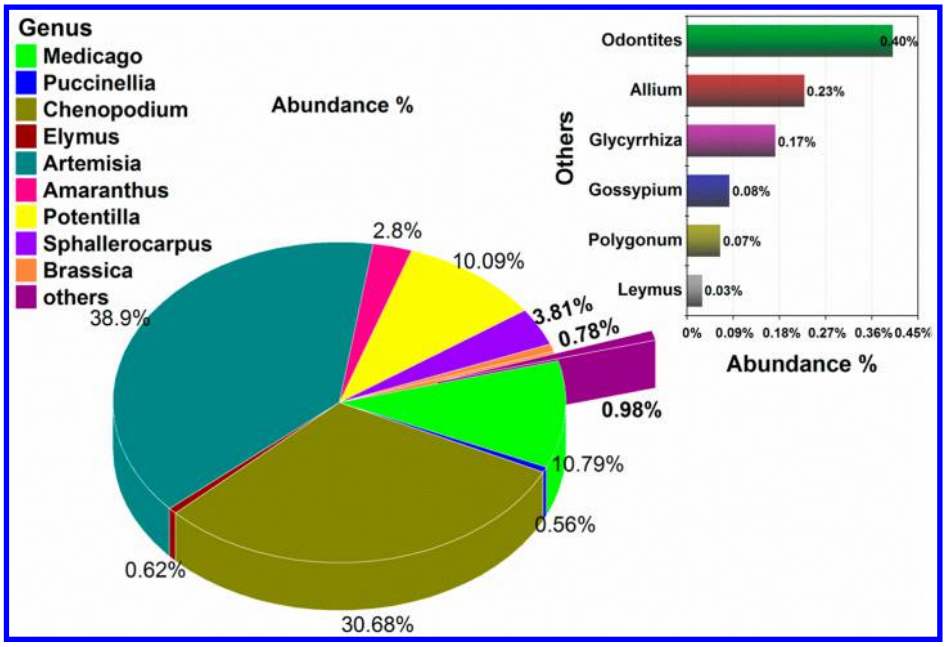

Figure 3. Relative content of plant genera in the fecal flora of the experimental sheep. The relative abundances of the genera were calculated using the number of sequences that were assigned to be of plant origin. These sequences were taxonomically assigned using the GenBank nr/nt database in NCBI, following OTU classification with UPARSE. Four genera with abundances below 0.01% are not shown in the figure.

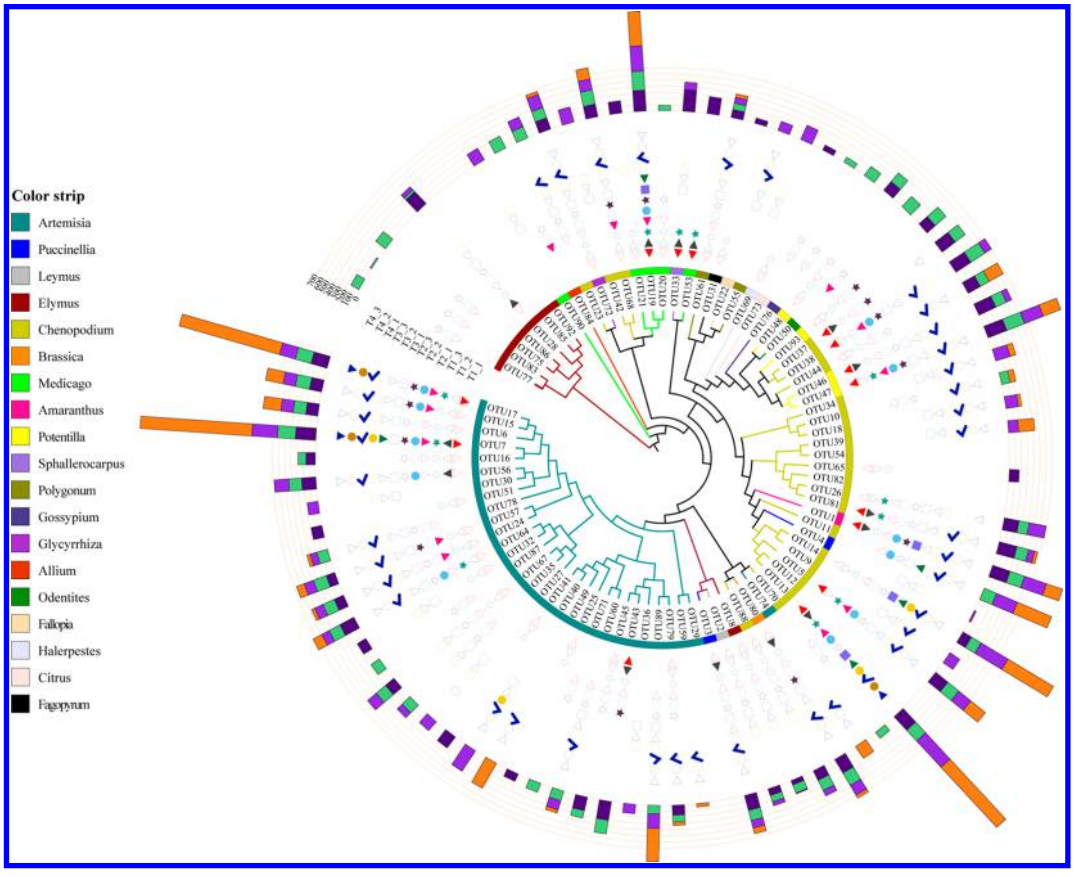

Figure 4. Phylogenetic tree showing the dietary OTUs from feces of all tested sheep. *ITS2* gene sequences were aligned using ClustalX. A phylogenetic tree was constructed on the basis of the maximum-likelihood algorithm using MEGA5. The 19 plant genera in the phylogenetic tree are indicated with colored loops (inner circle). The distribution of OTUs in each fecal sample is indicated using different graphics tags (middle circle). Solid, sequence number > 500; hollow, sequence number < 500; empty, sequence number = 0. The relative proportions of plant genera in the four different sample groups are represented by bar charts (outer circle). Irrelevant OTUs are not included on the tree.

DOI: 10.1021/acs.jafc.8b02814
J. Agric. Food Chem. 2018, 66, 9858−9867

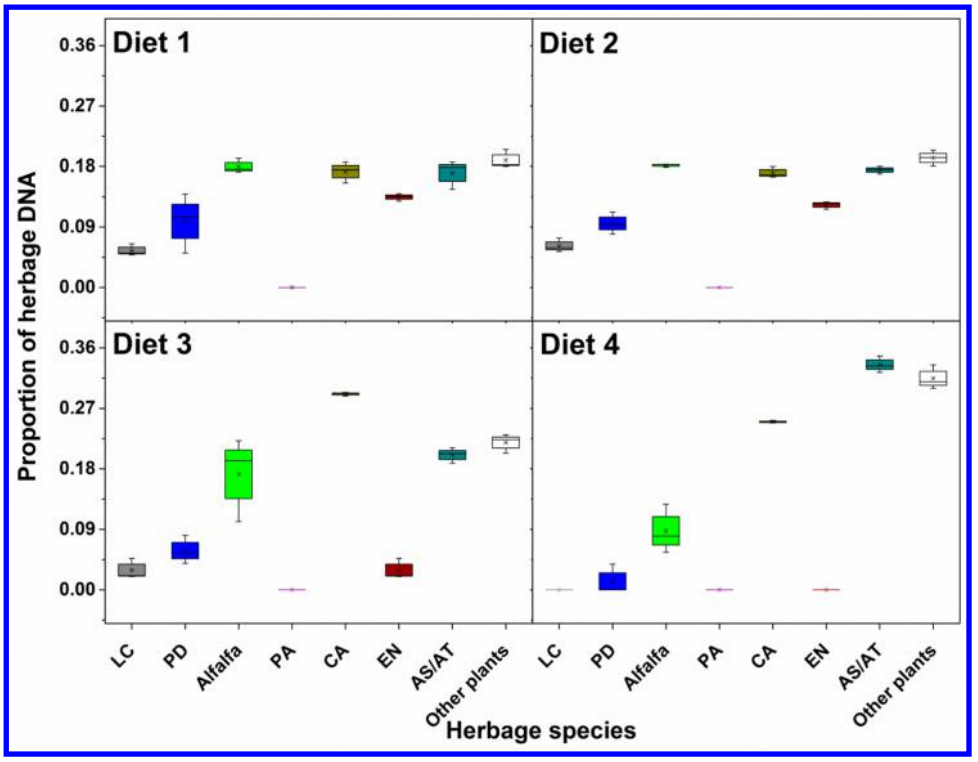

Figure 5. Illumina MiSeq sequencing estimates of herbage DNA in Mongolia sheep feces. Boxplots show the medians, ranges, and upper and lower quartiles of the percentages estimated by Illumina sequencing for each herbage species in the diet. LC, *Leymus chinensis*; PD, *Puccinellia distans*; alfalfa, *Medicago sativa*; PA, *Phragmites australis*; CA, *Chenopodium album*; EN, *Elymus Nutans*; AS, *Artemisia scoparia*; AT, *Artemisia tanacetifolia*.

Table 3. Diet Composition, Fecal DNA Sequences Recovered, and Estimates of Amount of the Plant Species Consumed Based on Fecal-DNA Analysis

		LC	PD	PA	CA	EN	AS or AT	alfalfa	other plants	total
mass of daily diet 1 (%)		41.82	36.84	—	—	—	—	21.34	—	
feces of group 1	no. of sequences	16	467	—	—	—	—	5383	17 060	22 927
	DD[a] (%)	15.83	35.10					49.06		
mass of daily diet 2 (%)		42.51	22.59	13.18	—	—	—	21.71	—	
feces of group 2	no. of sequences	20	123	0	—	—	—	5508	16 701	22 352
	DD (%)	18.25	29.29	0.00				52.46		
mass of daily diet 3 (%)		35.63	12.78	10.85	15.12	4.65	—	20.97	—	
feces of group 3	no. of sequences	3	9	0	21 115	3	—	997	2766	24 893
	DD (%)	5.17	10.34	0.00	46.84	5.17		32.48		
mass of daily diet 4 (%)		22.70	10.69	8.74	12.37	4.78	19.15	21.58	—	
feces of group 4	no. of sequences	0	2	0	1287	0	15 375	19	8055	24 738
	DD (%)	0.00	3.39	0.00	35.03	0.00	47.17	14.41		

[a]Daily-diet estimates based on DNA sequences. DD (%) was calculated via the proportion of the logarithm of the number of DNA sequences.

results support, with certain limitations, the utility and specification of the *ITS2* gene as a marker for determining the diet components of sheep from fecal samples.

The *ITS2* primer pairs, first used by Bradley et al.,[29] proved to be efficient for the identification of herbage species. This finding has further contributed to the analysis of grazing-livestock diets, particularly in pastures with more complex plant-species populations, yet interestingly, it was found that the non-Gramineae DNA sequences amplified by the *ITS2* primers (CA, AS, AT, and alfalfa) far exceeded those of the Gramineae species (LC and PD), which supplied the largest portion of the experimental diets. One possible explanation is that there was specific preferential DNA amplification in the molecular analysis.[29] Thus, the *ITS2* region might be better suited for the identification of non-Graminaceous plants. Coincidently, this was confirmed by a previous study, reporting a higher success rate for identifying dicotyledons (76.1%) than that for monocotyledons (74.2%).[41]

Compared with other potential candidate-DNA gene regions (such as *psbA-trnH*, *matK*, *rbcL*, and *rpoC1*) and apart from its preponderant discrimination ability at the plant level (up to 92.7% at the species level), the *ITS2* barcode is also capable of distinguishing closely related taxa. This corroborates the concept that *ITS2* in plants should be the gold-standard DNA barcode (analogous to *CO1* in animals) for identifying plants at different classification levels.[27,42] According to Chen et al.,[27] the chloroplast *psbA3-trnH* intergenic region has been recommended as a complementary barcode to *ITS2* because of its dependability for species authentication.

DOI: 10.1021/acs.jafc.8b02814
J. Agric. Food Chem. 2018, 66, 9858−9867

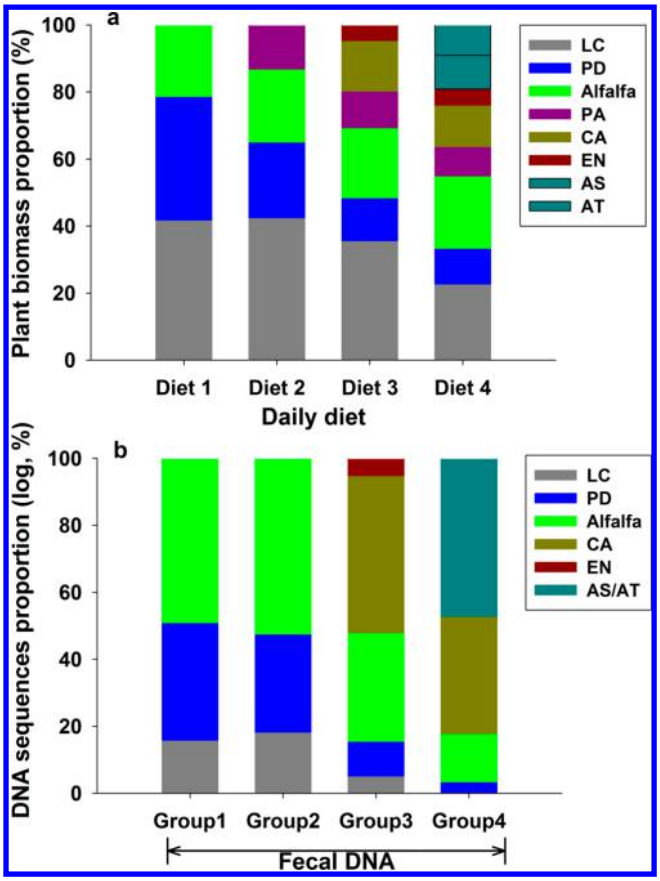

Figure 6. Proportions of herbage mass and herbage DNA sequences obtained from the captive feeding experiment. (a) Mass proportions of the eight herbages in the daily diet fed to the sheep. (b) Proportions of herbage DNA sequences recovered from sheep feces.

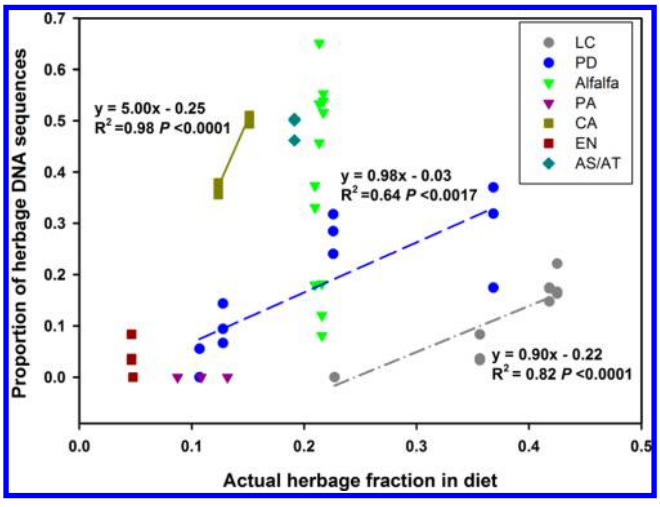

Figure 7. Correlation between the plant-species mass proportions in the experimental diets and the plant-species mass proportions estimated using fecal-DNA analysis. Regression relationships between the diet plant-species mass proportions and the plant-species-mass-proportion estimates for LC (gray dash−dot line), PD (dark-blue long-dash line), and CA DNA (dark-yellow solid line) based on DNA Illumina sequencing. The plotted association is based on the species level. The different lines represent the best-fit lines.

Quantitative Estimation of Sheep-Diet Composition. This study focused on fecal-DNA analysis as the method for the identifying of both the plants consumed by grazing herbivores and their amounts. The aim was to determine whether the relative sequence abundances could reliably reflect the diets of the sheep. Similar to other studies, it remains uncertain whether the proportion of recovered DNA sequences reflects the actual proportion of eaten material. Kartzinel et al.[43] analyzed the fecal DNA of seven large African mammalian herbivores and quantified their diets using high-throughput sequencing. The authors emphasized that grass relative read abundance can realistically reflect grass consumption. In the present study, the correlation between the proportion of DNA sequences and the proportion of herbage materials in the experimental diets was examined using Spearman's ρ correlation coefficient. The results indicate a significant correlation between both (Spearman $\rho = 0.376$, $P = 0.003$), which indicates that the quantity of consumed herbage biomass affects how many DNA sequences are recovered after DNA barcoding. Such a finding is important, especially for wild free-ranging animals with complicated diets.[44] Similar to our findings, a positive Pearson correlation was found between the actual proportions of forbs fed to sheep and the proportions of forbs estimated with the *trnL*-barcoding approach. However, this study investigated only diets that included two plant species.[45]

Researchers are primarily interested in obtaining a reliable quantitative indication (absolute or relative quantification) of the food items that have been ingested by the animals in a pasture. To confirm whether this *ITS2* approach could correctly estimate herbage proportions in various diets, a linear-regression analysis of the herbage-DNA-sequence proportions was performed via Illumina sequencing against the actual mass proportions of the herbage fed to the sheep. As expected, a series of analyses showed a strong linear-regression relationship between the proportions of the herbage DNA sequences (LC, PD, and CA) retrieved from sheep feces and the actual proportions of herbage in the daily diets (Figure 7). However, the regression analysis results of alfalfa (approximately accounting for the same mass proportion of 21% in the four diets) were not consistent with the expected results. This may be due to the fact that alfalfa pellets were used in the diet formulation rather than fresh-cut alfalfa. The pelleting process applies heat that may distort nucleic acid information. In addition, another limitation of this study was the small sample size of the herbage species involved in the experimental diets. A follow-up study will include more herbage species to strengthen the utility of the obtained prediction model.

Illumina MiSeq Sequencing of Plant DNA from Sheep Feces. In general, the application of Illumina MiSeq sequencing demonstrated good potential for estimating the plant composition of herbivorous animal diets. Because of their good resolution, next-generation-sequencing techniques are appropriate for quantifying the dietary compositions of herbivores and for evaluating the diet niche partitioning.[46] Robeson et al.[47] sequenced DNA from the fecal material of wild pigs using blocking primers; the authors emphasized that the DNA-sequencing-based method is suitable for studying the diet compositions of animals from different locations. However, in the present study, AS and AT (accounting for 19.15% of the total in daily diet 4), which belong to homologous species, were not distinguishable. This could be attributable to the incompleteness of the NCBI database.[48] Moreover, the establishment of a complete *ITS2*-fragment reference database for a variety of herbage species that occur in a particular region typically allows for the identification of

DOI: 10.1021/acs.jafc.8b02814
J. Agric. Food Chem. 2018, 66, 9858−9867

approximately 50% of the different species at the species level and 90% at the genus level.[18]

Several issues during the experiment need to be reported. First, a noticeable deviation was observed between the proportions of DNA sequences and the actual proportions of herbage in the diets (Figure 6). Apart from specific preferential DNA amplification, to some extent this may be caused by interference from other plant species consumed by the sheep prior to being confined. It is important to note that the retrieved DNA sequences also matched other plant species (Figure 3). This result implies that a degree of genetic information may remain in the feces for at least 6 days after ingestion. In this context, more than 6 days should be an appropriate duration for sheep maintenance in confinement prior to fecal sampling. Second, it was noted that DNA sequences extracted from sheep fecal samples contained information on plant species not included in their diets (e.g., *Artemisia annua* accounted for 11.38% of the sequences), which was likely caused by the imprecise sampling. Third, in the current study, sheep were fed experimental diets formulated to contain three to eight different plant species for 6 days prior to sampling. Inexplicably, DNA sequences from PA, which is known to be the major plant consumed by pasture sheep, was not detected in any of the fecal samples. Deagle et al.[49] concluded that the use of group-specific markers may alleviate this problem. Bowles et al.[50] used real-time PCR in combination with species-specific primers to quantify the prey compositions of Steller sea lions (*Eumetopias jubatus*) on the basis of fecal DNA and reported that this method can accurately describe the relative quantities of prey species consumed. The precision of their fecal-DNA method remained within 12−17% of the actual quantity once the relative mitochondrial contents of the prey species had been corrected. This real-time-PCR approach faces several challenges; however, it is worth investigating in future studies.

AUTHOR INFORMATION

Corresponding Author

*E-mail: zhangyj@cau.edu.cn.

ORCID

Yanping Guo: 0000-0001-7637-3733

Author Contributions

Y.J.Z. and Y.P.G. conceived and designed the experiment; Y.P.G., H.Z., and W.Q.C. performed both the sheep-feeding experiment and the field sampling. All molecular-analysis work was performed by Y.P.G., and Y.P.G. also analyzed the data; compiled tables and figures; and wrote the manuscript, which was revised and improved by H.Z., W.Q.C., and Y.J.Z.

Funding

The work was funded by the 973 project (2014CB138805) and the project for Agro-Industry Technology Research System (CARS-34).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are grateful to the National Field Research Station of Grassland Ecosystem in Guyuan, China for providing the research site and for the generous support from the staff who work there. We thank Dr. Chao Yang for his constructive comments on the manuscript.

REFERENCES

(1) Dove, H.; Mayes, R. The use of plant wax alkanes as marker substances in studies of the nutrition of herbivores: a review. *Aust. J. Agric. Res.* **1991**, *42*, 913−952.

(2) Mayes, R. W.; Dove, H. Measurement of dietary nutrient intake in free-ranging mammalian herbivores. *Nutr. Res. Rev.* **2000**, *13*, 107−138.

(3) Hofmann, R. R. Evolutionary steps of ecophysiological adaptation and diversification of ruminants: a comparative view of their digestive system. *Oecologia* **1989**, *78*, 443−457.

(4) Rayé, G.; Miquel, C.; Coissac, E.; Redjadj, C.; Loison, A.; Taberlet, P. New insights on diet variability revealed by DNA barcoding and high-throughput pyrosequencing: chamois diet in autumn as a case study. *Ecol. Res.* **2011**, *26*, 265−276.

(5) Sibbald, A. M.; Oom, S. P.; Hooper, R. J.; Anderson, R. M. Effects of social behaviour on the spatial distribution of sheep grazing a complex vegetation mosaic. *Applied Animal Behaviour Science* **2008**, *115*, 149−159.

(6) Miraglia, N.; Bergero, D.; Bassano, B.; Tarantola, M.; Ladetto, G. Studies of apparent digestibility in horses and the use of internal markers. *Livest. Prod. Sci.* **1999**, *60*, 21−25.

(7) NORBURY, G. L.; SANSON, G. D. Problems with measuring diet selection of terrestrial, mammalian herbivores. *Australian Journal of Ecology* **1992**, *17*, 1−7.

(8) Boval, M.; Coates, D. B.; Lecomte, P.; Decruyenaere, V.; Archimède, H. Faecal near infrared reflectance spectroscopy (NIRS) to assess chemical composition, in vivo digestibility and intake of tropical grass by Creole cattle. *Anim. Feed Sci. Technol.* **2004**, *114*, 19−29.

(9) Mayes, R.; Lamb, C.; Colgrove, P. M. The use of dosed and herbage n-alkanes as markers for the determination of herbage intake. *J. Agric. Sci.* **1986**, *107*, 161−170.

(10) Kelman, W.; Bugalho, M.; Dove, H. Cuticular wax alkanes and alcohols used as markers to estimate diet composition of sheep (Ovis aries). *Biochem. Syst. Ecol.* **2003**, *31*, 919−927.

(11) Lin, L.; Luo, H.; Zhang, Y.; Wang, H.; Shu, B.; Hong, F. The potential use of long-chain alcohols and fatty acids as diet composition markers: factors influencing faecal recovery rates and diet composition estimates in sheep. *Animal* **2009**, *3*, 1605−1612.

(12) Ali, H.; Mayes, R.; Hector, B.; Orskov, E. Assessment of n-alkanes, long-chain fatty alcohols and long-chain fatty acids as diet composition markers: the concentrations of these compounds in rangeland species from Sudan. *Anim. Feed Sci. Technol.* **2005**, *121*, 257−271.

(13) Dove, H.; Mayes, R. Plant wax components: a new approach to estimating intake and diet composition in herbivores. *J. Nutr.* **1996**, *126*, 13.

(14) Dove, H.; Mayes, R. Using n-alkanes and other plant wax components to estimate intake, digestibility and diet composition of grazing/browsing sheep and goats. *Small Ruminant Research* **2005**, *59*, 123−139.

(15) Ando, H.; Setsuko, S.; Horikoshi, K.; Suzuki, H.; Umehara, S.; Inoue-Murayama, M.; Isagi, Y. Diet analysis by next-generation sequencing indicates the frequent consumption of introduced plants by the critically endangered red-headed wood pigeon (Columba janthina nitens) in oceanic island habitats. *Ecology and Evolution* **2013**, *3*, 4057−4069.

(16) 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.* **2012**, *12*, 259−266.

(17) Pompanon, F.; Deagle, B. E.; Symondson, W. O.; Brown, D. S.; Jarman, S. N.; Taberlet, P. Who is eating what: diet assessment using next generation sequencing. *Mol. Ecol.* **2012**, *21*, 1931−1950.

(18) Valentini, A.; Miquel, C.; Nawaz, M. A.; Bellemain, E. V. A.; Coissac, E.; Pompanon, F.; Gielly, L.; Cruaud, C.; Nascetti, G.; Wincker, P.; Swenson, J. E.; Taberlet, P. New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Mol. Ecol. Resour.* **2009**, *9*, 51−60.

DOI: 10.1021/acs.jafc.8b02814
J. Agric. Food Chem. 2018, 66, 9858−9867

(19) Valentini, A.; Pompanon, F.; Taberlet, P. DNA barcoding for ecologists. *Trends in Ecology & Evolution* **2009**, *24*, 110−117.
(20) Deagle, B. E.; Kirkwood, R.; Jarman, S. N. Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Mol. Ecol.* **2009**, *18*, 2022−2038.
(21) Pegard, A.; Miquel, C.; Valentini, A.; Coissac, E.; Bouvier, F. d. r.; François, D.; Taberlet, P.; Engel, E.; Pompanon, F. o. Universal DNA-based methods for assessing the diet of grazing livestock and wildlife from feces. *J. Agric. Food Chem.* **2009**, *57*, 5700−5706.
(22) Deagle, B. E.; Chiaradia, A.; McInnes, J.; Jarman, S. N. Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? *Conserv. Genet.* **2010**, *11*, 2039−2048.
(23) Ankenbrand, M. J.; Keller, A.; Wolf, M.; Schultz, J.; Förster, F. ITS2 database V: Twice as much. *Mol. Biol. Evol.* **2015**, *32*, 3030−3032.
(24) Poinar, H. N.; Hofreiter, M.; Spaulding, W. G.; Martin, P. S.; Stankiewicz, B. A.; Bland, H.; Evershed, R. P.; Possnert, G.; Pääbo, S. Molecular coproscopy: dung and diet of the extinct ground sloth Nothrotheriops shastensis. *Science* **1998**, *281*, 402−406.
(25) Taberlet, P.; Coissac, E.; Pompanon, F.; Gielly, L.; Miquel, C.; Valentini, A.; Vermat, T.; Corthier, G.; Brochmann, C.; Willerslev, E. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* **2007**, *35*, e14−e14.
(26) Taberlet, P.; Coissac, E.; Pompanon, F.; Gielly, L.; Miquel, C.; Valentini, A.; Vermat, T.; Corthier, G.; Brochmann, C.; Willerslev, E. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* **2007**, *35*, No. e14.
(27) Chen, S.; Yao, H.; Han, J.; Liu, C.; Song, J.; Shi, L.; Zhu, Y.; Ma, X.; Gao, T.; Pang, X.; et al. Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. *PLoS One* **2010**, *5*, No. e8613.
(28) Pang, X.; Shi, L.; Song, J.; Chen, X.; Chen, S. Use of the potential DNA barcode ITS2 to identify herbal materials. *Journal of natural medicines* **2013**, *67*, 571−575.
(29) Bradley, B. J.; Stiller, M.; Doran-Sheehy, D. M.; Harris, T.; Chapman, C. A.; Vigilant, L.; Poinar, H. Plant DNA sequences from feces: potential means for assessing diets of wild primates. *Am. J. Primatol.* **2007**, *69*, 699−705.
(30) Brosh, A.; Henkin, Z.; Rothman, S. J.; Aharoni, Y.; Orlov, A.; Arieli, A. ORLOV Effects of faecal n-alkane recovery in estimates of diet composition. *J. Agric. Sci.* **2003**, *140*, 93−100.
(31) Elwert, C.; Kluth, H.; Rodehutscord, M. Effect of variable intake of alfalfa and wheat on faecal alkane recoveries and estimates of roughage intake in sheep. *J. Agric. Sci.* **1999**, *142*, 213−223.
(32) Amato, K. R.; Yeoman, C. J.; Kent, A.; Righini, N.; Carbonero, F.; Estrada, A.; Rex Gaskins, H.; Stumpf, R. M.; Yildirim, S.; Torralba, M.; Gillis, M.; Wilson, B. A.; Nelson, K. E.; White, B. A.; Leigh, S. R. Habitat degradation impacts black howler monkey (Alouatta pigra) gastrointestinal microbiomes. *ISME J.* **2013**, *7*, 1344−1353.
(33) Schloss, P. D.; Gevers, D.; Westcott, S. L. Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. *PLoS One* **2011**, *6*, No. e27310.
(34) R Core Team. *R: a language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria, 2014. Available at http://www.R-project.org/.
(35) Wang, Y.; Sheng, H.-F.; He, Y.; Wu, J.-Y.; Jiang, Y.-X.; Tam, N. F.-Y.; Zhou, H.-W. Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. *Appl. Environ. Microbiol.* **2012**, *78*, 8264−8271.
(36) Noval Rivas, M.; Burton, O. T.; Wise, P.; Zhang, Y.-q.; Hobson, S. A.; Garcia Lloret, M.; 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. A microbiota signature associated with experimental food allergy promotes allergic sensitization and anaphylaxis. *J. Allergy Clin. Immunol.* **2013**, *131*, 201−212.
(37) Maughan, H.; Wang, P. W.; Diaz Caballero, J.; Fung, P.; Gong, Y.; Donaldson, S. L.; Yuan, L.; Keshavjee, S.; Zhang, Y.; Yau, Y. C. W.; et al. Analysis of the Cystic Fibrosis Lung Microbiota via Serial Illumina Sequencing of Bacterial 16S rRNA Hypervariable Regions. *PLoS One* **2012**, *7*, No. e45791.
(38) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403−410.
(39) Sokal, R. R.; Rohlf, F. J. *Biometry: the principles and practice of statistics in Biological Research*; W. H. Freeman: New York, 1995.
(40) Lin, L. J.; Zhu, X. Y.; Jiang, C.; Luo, H. L.; Wang, H.; Zhang, Y. J.; Hong, F. Z. The potential use of n-alkanes, long-chain alcohols and long-chain fatty acids as diet composition markers: indoor validation with sheep and herbage species from the rangeland of Inner Mongolia of China. *Animal* **2012**, *6*, 449−458.
(41) Yao, H.; Song, J.; Liu, C.; Luo, K.; Han, J.; Li, Y.; Pang, X.; Xu, H.; Zhu, Y.; Xiao, P.; et al. Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS One* **2010**, *5*, No. e13102.
(42) Müller, T.; Philippi, N.; Dandekar, T.; Schultz, J.; Wolf, M. Distinguishing species. *RNA* **2007**, *13*, 1469.
(43) Kartzinel, T. R.; Chen, P. A.; Coverdale, T. C.; Erickson, D. L.; Kress, W. J.; Kuzmina, M. L.; Rubenstein, D. I.; Wang, W.; Pringle, R. M. DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 8019−8024.
(44) Nichols, R. V.; Åkesson, M.; Kjellander, P. Diet Assessment Based on Rumen Contents: A Comparison between DNA Metabarcoding and Macroscopy. *PLoS One* **2016**, *11*, No. e0157977.
(45) Willerslev, E.; Davison, J.; Moora, M.; Zobel, M.; Coissac, E.; Edwards, M. E.; Lorenzen, E. D.; Vestergård, M.; Gussarova, G.; Haile, J.; et al. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* **2014**, *506*, 47−51.
(46) Tilman, D.; Borer, E. T. African mammals, foodwebs, and coexistence. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 7890−7891.
(47) Robeson, M. S.; Khanipov, K.; Golovko, G.; Wisely, S. M.; White, M. D.; Bodenchuck, M.; Smyser, T. J.; Fofanov, Y.; Fierer, N.; Piaggio, A. J. Assessing the utility of metabarcoding for diet analyses of the omnivorous wild pig (Sus scrofa). *Ecology and Evolution* **2018**, *8*, 185−196.
(48) Jo, H.; Ventura, M.; Vidal, N.; Gim, J.-S.; Buchaca, T.; Barmuta, L. A.; Jeppesen, E.; Joo, G.-J. Discovering hidden biodiversity: the use of complementary monitoring of fish diet based on DNA barcoding in freshwater ecosystems. *Ecology and Evolution* **2016**, *6*, 219−232.
(49) Deagle, B. E.; Jarman, S. N.; Coissac, E.; Pompanon, F.; Taberlet, P. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biol. Lett.* **2014**, *10*, 20140562.
(50) Bowles, E.; Schulte, P. M.; Tollit, D. J.; Deagle, B. E.; Trites, A. W. Proportion of prey consumed can be determined from faecal DNA using real-time PCR. *Mol. Ecol. Resour.* **2011**, *11*, 530−540.

DOI: 10.1021/acs.jafc.8b02814
J. Agric. Food Chem. 2018, 66, 9858−9867