Evidence of cellulose metabolism by the giant panda gut microbiome

Lifeng Zhu[a,1], Qi Wu[a,1], Jiayin Dai[a], Shanning Zhang[b], and Fuwen Wei[a,2]

[a]Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China and [b]China Wildlife Conservation Association, Beijing 100714, China

Edited by Rita R. Colwell, University of Maryland, College Park, MD, and approved September 7, 2011 (received for review December 2, 2010)

The giant panda genome codes for all necessary enzymes associated with a carnivorous digestive system but lacks genes for enzymes needed to digest cellulose, the principal component of their bamboo diet. It has been posited that this iconic species must therefore possess microbial symbionts capable of metabolizing cellulose, but these symbionts have remained undetected. Here we examined 5,522 prokaryotic ribosomal RNA gene sequences in wild and captive giant panda fecal samples. We found lower species richness of the panda microbiome than of mammalian microbiomes for herbivores and nonherbivorous carnivores. We detected 13 operational taxonomic units closely related to *Clostridium* groups I and XIVa, both of which contain taxa known to digest cellulose. Seven of these 13 operational taxonomic units were unique to pandas compared with other mammals. Metagenomic analysis using ∼37-Mbp contig sequences from gut microbes recovered putative genes coding two cellulose-digesting enzymes and one hemicellulose-digesting enzyme, cellulase, β-glucosidase, and xylan 1,4-β-xylosidase, in *Clostridium* group I. Comparing glycoside hydrolase profiles of pandas with those of herbivores and omnivores, we found a moderate abundance of oligosaccharide-degrading enzymes for pandas (36%), close to that for humans (37%), and the lowest abundance of cellulases and endohemicellulases (2%), which may reflect low digestibility of cellulose and hemicellulose in the panda's unique bamboo diet. The presence of putative cellulose-digesting microbes, in combination with adaptations related to feeding, physiology, and morphology, show that giant pandas have evolved a number of traits to overcome the anatomical and physiological challenge of digesting a diet high in fibrous matter.

Access to dietary resources shapes animal evolution (1). Early on, animals lost the ability to synthesize many key compounds, and instead this function is performed by symbionts (2). For example, microbial symbionts assist with extracting nutrients from food and key compounds from the environment, and also synthesize necessary metabolic compounds (1). Gut microbiota share specialized relationships with their hosts, and advances in genomics are revealing the dynamics of these relationships (3). Recent developments in culture-independent methodologies based on large-scale comparative analyses of microbial small-subunit ribosomal RNA genes (16S ribosomal RNA) and metagenomics have revealed the extent of microbial diversity and metabolic potential in greater detail (2–7). These techniques can now be applied to animals that have acquired a profoundly new diet, presenting an opportunity to investigate host physiological and microbial systems in an evolutionary context.

The giant panda (*Ailuropoda melanoleuca*) is well known for dietary oddities: a bamboo specialist within the mammalian order Carnivora possessing a gastrointestinal tract typical of carnivores. It consumes ∼12.5 kg of this highly fibrous plant each day (8), but because it lacks the long intestinal tract characteristic of other herbivores, extensive fermentation is not possible (9). Giant pandas digest only ∼17% of dry matter consumed (8), and have low digestion coefficients for bamboo hemicelluloses (27%) and celluloses (8%) (9). Indeed, the giant panda genome codes for all necessary enzymes associated with a carnivorous digestive system, but lacks the enzyme homologs needed for cellulose digestion (10). Although the giant panda can use noncellulosic material from the bamboo diet using enzymes coded in its own genome, digestion of cellulose and hemicellulose is impossible based on the panda's genetic composition, and must be dependent on gut microbiome. However, previous research using culture methods and small-scale sequencing identified three predominant bacteria from the panda gut—*Escherichia coli*, *Streptococcus*, and Enterobacteriaceae—none of which aids in cellulose digestion (11–13). Thus, an incomplete understanding of the gut microbial ecosystem in this interesting and high-profile species remains because of restrictions in methodology and past reliance on studies of captive animals.

Results and Discussion

We undertook a large-scale analysis of 16S rRNA gene sequences to profile microbial flora inhabiting the digestive system of giant pandas and used a metagenomic approach based on next-generation de novo sequencing to identify functional attributes encoded in the gut microbiome. A total of 5,636 near–full-length 16S rRNA gene segments were amplified from fecal samples of seven wild and eight captive giant pandas. After exclusion of 74 putative chimeric and 30 chloroplast sequences, 5,522 sequences were retained for analysis. Using a minimum identity of 97% as the threshold for any sequence pair, we identified 85 bacterial operational taxonomic units (OTUs), 14 of which were previously undescribed (Fig. 1*A* and *SI Appendix*, Table S1). Coverage of 99% was obtained across existing bacterial clone libraries, and we are confident that our dataset presents the most comprehensive assessment of gut microbes in this species based on the rarefaction method in DOTUR (14) (*SI Appendix*, Fig. S1*A*). The majority of microbes were members of the Firmicutes (62 OTUs, 4,633 sequences, 83.8% of the total of 5,522 sequences) and Proteobacteria (12 OTUs, 871 sequences, 15.8% of the total sequences), with the remainder belonging to the phyla Actinobacteria, Bacteroidetes, Cyanobacteria, and Acidobacteria (Fig. 1 and *SI Appendix*, Fig. S2 and Table S1). Within the Firmicutes, 33 OTUs (60.8% of the total of 5,522 sequences) were members of the class Clostridia and 29 OTUs (23.0% of total sequences) belonged to the class Bacilli. The high proportion of Firmicutes OTUs found in the gut of giant pandas differs from the findings in previous studies attempting to characterize giant panda gut microbes (11–13) and is notably similar

Author contributions: F.W. designed research; L.Z. and S.Z. performed research; L.Z., Q.W., J.D., S.Z., and F.W. analyzed data; and L.Z., Q.W. and F.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database. For a list of accession numbers, see *SI Appendix*. The metagenomic data project ID is 5936 in Integrated Microbial Genomes with Microbiome Samples (IMG/M).

[1]L.Z. and Q.W. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: weifw@ioz.ac.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1017956108/-/DCSupplemental.

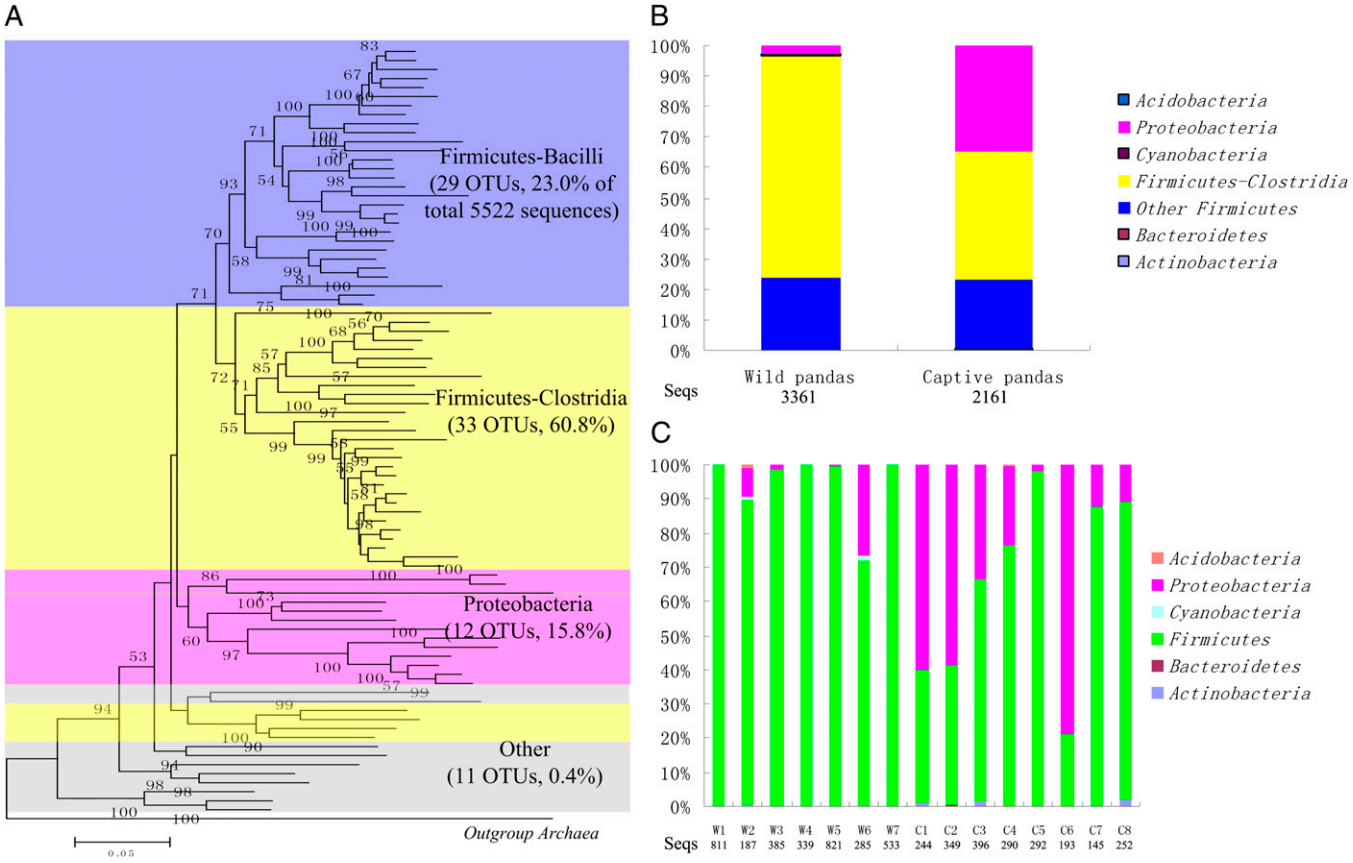

Fig. 1. Microbial flora of giant pandas. (*A*) Neighbor-joining tree containing one representative of each of 85 OTUs. Colors represent different phyla: yellow, Clostridia class of Firmicutes; blue, Bacilli class of Firmicutes; purple, Proteobacteria; gray, Actinobacteria, Bacteroidetes, Cyanobacteria, and Acidobacteria. seqs, sequences. All bootstrap values >80% are shown in this tree. (*B*) Relative abundance of sequences from wild and captive giant panda fecal samples; other Firmicutes belong to the Bacilli class. (*C*) Percentage of sequences from each fecal sample assigned to different phyla. W1–W7, wild giant panda samples; C1–C8, captive giant panda samples. Numbers below the sample number are total sequences from each individual.

to that found in herbivorous mammals (2, 6). However, the species richness in the giant panda was lower than that seen in previous work on mammalian herbivores and nonherbivorous carnivores (*SI Appendix*, Fig. S1*B*), perhaps due to the panda's unique bamboo diet and simple digestive system.

The proportion of 16S rRNA gene sequences belonging to the class Clostridia of Firmicutes were high for both wild (73% of the total of 3361 sequences) and captive (42% of the total of 2161 sequences) giant pandas (Fig. 1 *B* and *C*). In Clostridia, we detected 13 OTUs closely related to *Clostridium* group I (10 OTUs, 1,457 sequences, 26.4% of the total sequences) and XIVa (3 OTUs, 185 sequences, 3.3% of the total sequences) (*SI Appendix*, Fig. S3), groups that contain organisms known to digest cellulose (15–18). Moreover, we found that 7 of these 13 OTUs were unique to giant pandas compared with other mammals, including Artiodactyla, Perissodactyla, Rodentia, Diprotodontia, and primates (2, 19–21) (*SI Appendix*, Table S2).

Using gene functional classification in a metagenomic analysis of fecal samples from three wild individuals (W1, W2, and W5), we found sequences homologous to genes coding cellulase, β-glucosidase, xylan 1,4-β-xylosidase, and endo-1,4-β-xylanase. The de novo assembly (22) for Illumina Genome Analyzer sequence data resulted in a set of ~37 Mbp (*SI Appendix*, Table S4). We used MetaGene (23) to predict ORFs in our contigs and found 49,844 genes occupying nearly 32 Mbp. Of these genes, 44% possessed a complete ORF (*SI Appendix*, Tables S5 and S6). We determined the taxonomic level of the gene catalog with their protein sequences by the lowest common ancestor (LCA)-based algorithm implemented in MEGAN (24). The results again showed that the majority of microbes (71%) belong to the phylum Firmicutes (Fig. 2*A*). We also compared the catalog with the Kyoto Encyclopedia of Genes and Genomes (KEGG) database and the Clusters of Orthologous Groups (COG) databases to assess the functional capacities present in bacterial metagenome (Fig. 2 *B–D* and *SI Appendix*, Figs. S4 and S5). Our analysis revealed that half of these predicted genes coding for cellulose and hemicellulose-digesting enzymes are found in species within the *Clostridium* genus (*SI Appendix*, Table S7). For example, 2 of 12 putative cellulase genes in this study are homologous (E-value: 2.00E-115 and 6.00E-29, respectively) to genes from *Clostridium butyricum* of *Clostridium* group I (*SI Appendix*, Table S7), and 1,968 of 49,844 genes were assigned to this species. Of interest, OTU54 and OTU55 have a close phylogenetic relationship with *C. butyricum* (*SI Appendix*, Fig S3), and OTU54 was found in eight individuals [W1 (4.1%), W2 (13.4%), W3 (0.3%), W4 (1.5%), W5 (43.4%), W6 (3.5%), C3 (12.1%), and C8 (1.2%)] (*SI Appendix*, Table S1). We also found that 21 putative genes are homologous to genes coding xylan 1,4-β-xylosidase from *Clostridium* (*SI Appendix*, Table S7). KEGG metabolic pathways provide a highly integrated picture of global gut cell metabolism (Fig. 2*B*), suggesting that the gut microbiome of giant pandas has enriched activity for metabolism of carbohydrates, amino acids, xenobiotics, nucleotides, glycans, vitamins, and lipids (*SI Appendix*, Fig. S5). Moreover, comparative COG functional analysis of our data with several gastrointestinal or rumen metagenomes (6, 7, 20, 25) demonstrated

ECOLOGY

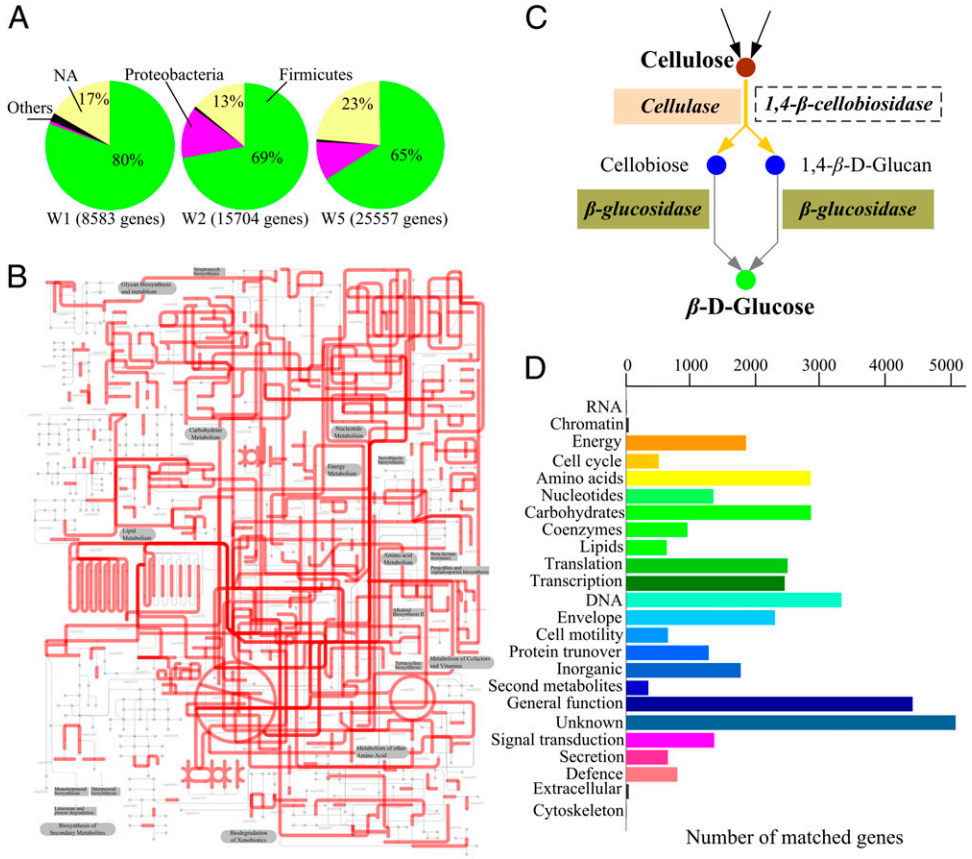

Fig. 2. The function of gut metagenomes from three wild giant pandas (W1, W2, and W5). (*A*) Gene taxonomic assignment. NA, no matched data or no taxon information for matched genes. "Others" includes bacteria in other classes, such as Bacteroidetes and Actinobacteria. (*B*) Projection of the gut genome on the KEGG pathways using iPath (26). (*C*) Pathway of cellulose digestion. The box indicates that 1,4-β-cellobiosidase was partially found in the unassembly short-read sequences set. (*D*) Functional composition of the gut metagenome in the COG database.

metabolic similarities among panda, wallaby, cow, termite, and human, associated mainly with carbohydrates, amino acids, and DNA metabolism (*SI Appendix*, Fig. S6).

Database searches for glycoside hydrolase (GH) genes were performed using HMMER HMM-based sequences search with Pfam hidden Markov models (HMMs), Pfam v25.0 and HMMER v3.0. These analyses found 448 genes from 44 different GH families in the Carbohydrate-Active Enzymes (CAZy) database (Table 1). Comparing GH profiles with those of other herbivores (wallaby, termite, cow) and an omnivore (human) (6, 7, 20, 25), we found that the abundance of oligosaccharide-degrading enzymes (36%) from giant pandas was close to that in human (37%) and lower than that in cow (57%); however, the abundance of cellulases and endohemicellulases was the lowest of the herbivores (*SI Appendix*, Table S8). These findings may reflect the low digestibility of the cellulose and hemicellulose present in the unique bamboo diet of giant pandas.

Our metagenomic analysis and 16S rRNA gene data confirm the presence of putative cellulose-metabolizing symbionts in this little-studied microbial environment and clarifies how giant pandas are able to partially digest bamboo fibers despite a genome lacking enzymes that can degrade cellulose. It is becoming increasingly clear that giant pandas possess a suite of evolutionary adaptations for the highly specialized herbivory. Our findings offer a more in-depth look at the microbiome of a species that occupies an interesting place in the evolutionary tree and has an unusually narrow diet. Thus, the putative harboring of cellulose and hemicellulose-digesting microbes in the gut of the giant panda, along with a suite of other traits, including pseudothumbs; well-developed teeth, mandible and skull morphology, and chewing muscles; high-volume food ingestion (12.5 kg/day); brief digestion time; and high mucous levels in the digestive tract, likely have arisen as a result of adapting to a highly fibrous bamboo diet within the constraints imposed by the panda's innate carnivore-like digestive system.

Methods

Sample Collection. Fecal samples from wild and captive giant pandas were collected immediately after defecation, snap-frozen in liquid N_2, and shipped to the laboratory on dry ice. All samples were obtained from inside the feces, where there was no contact with soil. Six fecal samples from wild giant pandas were collected from the Qinling Mountains, and one sample was collected from the Xiangling Mountains by following GPS-collared animals or by directly observing individuals. Eight fecal samples from captive giant pandas (six animals were born in captivity, and two have been kept in captivity for 9 y) were collected from breeding centers between April and May 2008 and between April and May 2009.

DNA Extraction and PCR Amplification. DNA was extracted from fecal samples using the Qiagen QIAamp DNA Stool Mini Kit according to the protocol for isolation of DNA for pathogen detection. DNA was eluted in a final volume of 250 μL using elution buffer and then stored at −20 °C. Tubes containing only the QIAamp DNA Stool Mini Kit extractions were included throughout the lysis and PCR steps to serve as negative controls. PCR amplification of 16S rRNA gene was performed as described previously (2). Five replicate PCRs were performed for each DNA sample. Each 25-μL reaction contained 60–100 ng of purified DNA, 12.5 μL of HotStarTaq Plus DNA Polymerase (Qiagen; 0.63 U DNA polymerase, 0.2 mM dNTPs, 2 mM Mg^{2+}, 10 × PCR buffer), 0.5 μM Bact-8F (5′-AGAGTTTGATCCTGGCTCAG-3′), 0.5 μM Bact-1391R (5′-GACGGGCGGTGWGTRCA-3′), and 1 μg/μL BSA (Promega). Cycling conditions

Table 1. Inventory of putative GHs identified in the giant panda gut microbiome

CAZy family	Pfam HMM name	Pfam accession	Known activities	Giant panda gut community
GH1	Glyco_hydro_1	PF00232	β-Glucosidase, β-galactosidase, β-mannosidase, others	101
GH2	Glyco_hydro_2_C	PF00703	β-Galactosidase, β-mannosidase, others	1
GH3	Glyco_hydro_3	PF00933	β-1,4-Glucosidase, β-1,4-xylosidase, β-1,3-glucosidase, α-larabinofuranosidase, others	18
GH4	Glyco_hydro_4	PF02056	α-Glucosidase, a-galactosidase, α-glucuronidase, others	6
GH5	Cellulase	PF00150	Cellulase, β-1,4-endoglucanase, β-1,3-glucosidase, β-1,4-endoxylanase, β-1,4-endomannanase, others	3
GH8	Glyco_hydro_8	PF01270	Cellulase, β-1,3-glucosidase, β-1,4-endoxylanase, β-1,4-endomannanase, others	2
GH10	Glyco_hydro_10	PF00331	Xylanase, β-1,3-endoxylanase	2
GH13	Alpha-amylase	PF00128	α-Amylase, catalytic domain, and related enzymes	35
GH16	Glyco_hydro_16	PF00722	β-1,3(4)-Endoglucanase, others	9
GH18	Glyco_hydro_18	PF00704	Chitinase, endo-β-*N*-acetylglucosaminidase, noncatalytic proteins	1
GH20	Glyco_hydro_20	PF00728	β-Hexosaminidase, lacto-*N*-biosidase	7
GH23	SLT	PF01464	G-type lysozyme, peptidoglycan lytic transglycosylase	20
GH24	Phage_lysozyme	PF00959	Lysozyme	1
GH25	Glyco_hydro_25	PF01183	Lysozyme	9
GH27	Melibiase	PF02065	α-Galactosidase, α-*N*-acetylgalactosaminidase, isomalto-dextranase	36
GH28	Glyco_hydro_28	PF00295	Polygalacturonase, rhamnogalacturonase, others	1
GH29	Alpha_L_fucos	PF01120	α-L-fucosidase	1
GH31	Glyco_hydro_31	PF01055	α-Glucosidase, α-xylosidase, others	9
GH32	Glyco_hydro_32N	PF00251	Levanase, invertase, others	13
GH35	Glyco_hydro_35	PF01301	β-Galactosidase	4
GH36	No Pfam	No Pfam	α-Galactosidase, α-*N*-acetylgalactosaminidase	33
GH37	Trehaalse	PF01204	α,α-trehalase	2
GH38	Glyco_hydro_38	PF01074	α-Mannosidase	10
GH39	Glyco_hydro_39	PF01229	β-Xylosidase, α-L-iduronidase	9
GH42	Glyco_hydro_42	PF02449	β-Galactosidase	18
GH51	No Pfam	No Pfam	Endoglucanase, α-L-arabinofuranosidase	5
GH55	No Pfam	No Pfam	Exo-1,3-glucanase, endo-1,3-glucanase	1
GH58	No Pfam	No Pfam	Endo-*N*-acetylneuraminidase or endo-sialidase	1
GH65	Glyco_hydro_65m	PF03632	Trehalase, maltose phosphorylase, trehalose phosphorylase	1
GH67	Glyco_hydro_67M	PF03648	α-Glucuronidase, others	2
GH70	Glyco_hydro_70	PF02324	Dextransucrase, alternansucrase	9
GH72	Glyco_hydro_72	PF03198	β-1,3-glucanosyltransglycosylase	3
GH73	Glucosaminidase	PF01832	Peptidoglycan hydrolase with endo-β-*N*-acetylglucosaminidase specificity	28
GH77	Glyco_hydro_77	PF02446	4-α-Glucanotransferase, amylomaltase	12
GH78	Bac_rhamnosid	PF05592	α-L-Rhamnosidase	2
GH85	Glyco_hydro_85	PF03644	Endo-β-*N*-acetylglucosaminidase	4
GH88	Glyco_hydro_88	PF07470	D-4,5 Unsaturated β-glucuronyl hydrolase	1
GH89	NAGLU	PF05089	Alpha-*N*-acetylglucosaminidase	2
GH90	No Pfam	No Pfam	Endorhamnosidase	1
GH94	No Pfam	No Pfam	Cellobiose phosphorylase, chitobiose phosphorylase, cellodextrin phosphorylase	7
GH95	No Pfam	No Pfam	α-L-Fucosidase	4
GH101	No Pfam	No Pfam	Endo-α-*N*-acetylgalactosaminidase	6
GH104	Phage_lysozyme	PF00959	Peptidoglycan lytic transglycosylase	1
GH109	No Pfam	No Pfam	α-*N*-acetylgalactosaminidase	7

CAZy family and known activities are from the Carbohydrate Active Enzymes database (http://www.cazy.org/) (32). Pfam HMM-based sequences name and Pfam accession are from the Pfam database (http://pfam.sanger.ac.uk/) (33), a large collection of protein families, each represented by multiple sequence alignments and hidden Markov models. Giant panda gut community: Number of predicted genes match the Pfam domains or "non-Pfam" representative sequences.

were as follows: 95 °C for 5 min, followed by 35 cycles of 95 °C for 40 s, 55 °C for 45 s, and 72 °C for 2 min, with a final extension period of 10 min at 72 °C. Replicate PCRs were pooled and gel-purified with the Wizard SV Gel and PCR Clean-Up System (Promega). Amplicons were cloned into PMD18-T vector (TaKaRa) and transformed into *E. coli* TOP10 (Tiangen). Approximately 192–900 colonies from each sample PCR product were chosen at random. Plasmid inserts were sequenced bidirectionally using BigDye Terminator (Applied Biosystems) and vector-specific primers (M13F: 5′-GTAAAACGACGGCCAG-3′; M13R: 5′-CAGGAAACAGCTATGAC-3′). Sequences were analyzed with DNASTAR (DNAStar) and trimmed to remove vector sequences. After trimming and adjusting for quality values, the average single sequence read length was ∼700 nucleotides. Bidirectional sequence reads of clone inserts provided near–full-length 16S rRNA gene sequences of ∼1,400 bp.

Chimera Checking. Each sequence was edited manually in conjunction with its chromatogram and secondary structure information. Only unambiguous

nucleotide positions were included in the analysis, and primer sequences were excluded. A multiple sequence alignment was generated with the NAST online tool (27), and chimeras were identified with Bellerophon version 3 (28), implemented at the Greengenes Web site (http://greengenes.lbl.gov) with the following (default) parameters: Sequences were compared with others within the same host species and with the Greengenes Core Set, identity to the core set was set to 97%, the match length to sequence threshold was set to 1,050 bp and 1,150 bp, respectively, the window size was set to 300, the count of similar sequences to search for each window was 7, the parent-to-fragment ratio was 90%, and the divergence ratio threshold was set at 1.1.

Determining OTU and Taxonomy Assignments. *Determining OTU for each sample*. OTUs were defined as terminal nodes in phylogenetic analysis. Sequences remaining after the chimeras were checked for each individual were aligned using CLUSTAL in MEGA4 (29). OTU determination was performed according to DOTUR analysis (14). In DOTUR, sequences were grouped into OTU using distance-similarity matrices (by PHYLIP; http://evolution.genetics.washington.edu/phylip.html), such that the least similar pair within the OTU shared at least 97% identity.

Determining OTUs for all samples. To increase validity and decrease computing time, we determined all OTUs and shared OTUs in the sequences for all samples in two steps. First, we selected one wild and one captive sample sequence as the central library. We combined each sample sequence with this central library independently, and then determined the shared OTUs between the central libraries with each sample. Second, we combined unshared sequences for each sample into one new library, then used the foregoing method to determine the OTUs in this remaining large library. We chose a representative sequence for each OTU in each of the 97% identity groups at random, then used the BLAST method to find the closest GenBank neighbor for these representative sequences. Each representative sequence for each OTU was submitted to GenBank and granted an accession number. The result of EzTaxon (30), which was also used to determine the closest taxon for each OTU, served as the reference. Further, each representative sequence for each OTU with <95% identity to any GenBank sequence was defined as a previously undescribed OTU.

Phylogenetic Analysis. Based on the foregoing taxon assignment results, we downloaded the representative of the 16S rRNA gene sequences of the bacteria from the National Center for Biotechnology Information (NCBI) Nucleotide Database, combining all OTUs to construct the neighbor-joining phylogenetic tree (1,000 bootstraps) using MEGA4 (29). We selected the representative sequence for each OTU at random. We used this tree to determine the phylogenetic position for each out, and used the representative sequence on each OTU to construct the neighbor-joining phylogenetic tree (1,000 bootstraps) with MEGA4 (29).

Relationship Between All OTUs and Known *Clostridium* Clusters. We downloaded sequences from GenBank according to known *Clostridium* clusters (15). We then combined these sequences with all of the OTUs in this study and constructed the neighbor-joining phylogenetic tree (1,000 bootstraps) using MEGA4 (29). Based on phylogenetic information (in the same clade), we determined the relationships between all OTUs and known *Clostridium* clusters.

Estimation of Microbial Diversity. We calculated the Good coverage estimate as $[1 - (n/N)] \times 100$, where n is the number of singleton sequences and N is the total number of sequences for the sample analyzed (3). We used the rarefaction method in DOTUR (14) to explore the diversity of our clone libraries and found that the number of observed OTUs increased with additional sampling effort; however, the 99% coverage obtained over existing bacterial clone libraries indicated that only 1 new OTU would be expected for every 100 additional sequenced clones. Species richness was defined as the number of gut microbe species detected in a sample. DOTUR was used to calculate various diversity indices along with species richness. The classic Shannon–Weaver diversity indices (with 95% confidence interval) of each sample were calculated as described previously (14).

Metagenomics Analysis: DNA Extraction, DNA Library Construction, and Sequencing. DNA was extracted from fecal samples of three individual wild giant pandas using the Qiagen QIAamp DNA Stool Mini Kit according to the protocol for isolation of DNA for pathogen detection. Sequencing and general data analyses were performed by BGI (Shenzhen, China). DNA library construction and sequencing followed BGI's previous work on human gut microbe metagenomic sequencing (7). A library with 350-bp clone insert size was constructed for our sample. Approximately 83 million high-quality reads, with read lengths of 90 bp, were generated for the samples; thus, the total data volume of high-quality reads was nearly 7.5 Gbp (*SI Appendix*, Table S3).

Public Data Use. Public data used in the metagenomic analysis included the human gut contig set (version December 10, 2009), the NCBI sequenced bacteria genomes database (version January 13, 2010), the NCBI sequenced fungi genomes database (version March 17, 2010), the NCBI sequenced protozoa genomes database (version July 25, 2007), the Ribosomal Database Project (version July 14, 2010), and the integrated NCBI-NR database (version March 2011).

Illumina Genome Analyzer (GA) Short Reads de novo Assembly. We compared the raw short reads with giant panda genome data to remove the host sequence. The clean reads thus obtained were assembled to obtain long contig sequences by the SOAPdenovo assembler (22), as used in human gut microbe metagenomic analyses. We tried different Kmer frequencies to obtain different assembly results, and used N50 lengths to access the best assembly result. The longest contig length and highest read utilization rate were obtained in Kmer55 for W1, Kmer31 for W2, and Kmer53 for W5. Thus, we used the contigs of Kmer55 for W1, Kmer31 for W2, and Kmer53 for W5 as the final assembly result (*SI Appendix*, Table S4).

Gene Prediction and Taxonomic Assignment. We used the assembly contig sequences and applied MetaGene software, with only ORFs longer than 100 bp preserved. We translated the ORFs into protein sequences using NCBI Genetic Code 11. We carried out BLASTP (31) alignment to query the predicted protein sequences against the integrated NR protein database. For each predicted gene, hits with E-values $>1 \times 10^{-5}$ were filtered. Then a significant-matches set was retained to distinguish taxonomic groups, which were defined for hits with E-values <10 times the top hit E-value. Next, the LCA-based algorithm implemented in MEGAN (24) was introduced to determine the taxonomic level of each gene. The LCA-based algorithm assigns genes to taxa so that the taxonomic level of the assigned taxon reflects the level of conservation of the gene. For example, if the hits in the significant matches set belonged to one species, then the predicted gene was considered conserved within the species and assigned to the species. If the hits belonged to several species within one genus, then the predicted gene was assigned to the genus, which was considered the LCA of the predicted gene. If the hits belonged to more species from different genera and all of the genera belonged to one family (or higher taxonomic unit), then the predicted gene was assigned to the family, which was then considered the LCA. The remaining higher taxa can be determined in the same manner.

Gene Functional Classification. We performed predicted gene functional classification by querying protein sequences of the genes against the eggNOG database (an integration of the COG and KOG databases) and the KEGG database using BLASTP with E-values $<1 \times 10^{-5}$. Genes were annotated as a function of the NOG or KEGG homologs with the lowest E-value. In the COG database, genes were classified into COG categories, whereas in KEGG, genes were assigned to KEGG pathways.

Glycoside Hydrolase (GH) Family Annotation and Analysis. We performed database searches for GHs using HMMER HMM-based sequences search with Pfam hidden Markov models (Pfam v25.0 and HMMER v3.0) as described previously (20, 25). We used the Pfam_ls HMMs to find complete matches with the family by global alignment. All hits with E-values $<10^{-4}$ were counted, and their sequences were analyzed further. GHs were named in accordance with the CAZy nomenclature scheme (32). For those GH families for which there is currently no Pfam HMM-based sequences, the representative sequences selected from the CAZy Web site were used in BLAST searches of the metagenomic data to identify these GH families using an E-value cutoff of 10^{-6} (20, 25).

ACKNOWLEDGMENTS. We thank Drs. Chaodong Zhu and Hengling Cui for assistance with data analysis, Yonggang Nie for assistance in the field, the anonymous reviewers for their constructive comments, Prof. James Elser for the English polishing and giant panda breeding centers for sampling assistance. This study was supported by the National Basic Research Program of China (973 Program; 2007CB411600), the National Natural Science Foundation of China (30830020 and 31070331), and the Knowledge Innovation Program of the Chinese Academy of Sciences (KSCX2-EW-Z-4).

1. Moran NA (2006) Symbiosis (a primer). *Curr Biol* 16:866–871.
2. Ley RE, et al. (2008) Evolution of mammals and their gut microbes. *Science* 320: 1647–1651.
3. Eckburg PB, et al. (2005) Diversity of the human intestinal microbial flora. *Science* 308: 1635–1638.
4. Weng L, Rubin EM, Bristow J (2006) Application of sequence-based methods in human microbial ecology. *Genome Res* 16:316–322.
5. Swanson KS, et al. (2011) Phylogenetic and gene-centric metagenomics of the canine intestinal microbiome reveals similarities with humans and mice. *ISME J* 5:639–649.
6. Brulc JM, et al. (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc Natl Acad Sci USA* 106:1948–1953.
7. Qin JJ, et al.; MetaHIT Consortium (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59–65.
8. Schaller GB, Hu JC, Pan WS, Zhu J (1984) *The Giant Pandas of Wolong* (Univ. of Chicago Press, Chicago).
9. Dierenfeld ES, Hintz HF, Robertson JB, Van Soest PJ, Oftedal OT (1982) Utilization of bamboo by the giant panda. *J Nutr* 112:636–641.
10. Li RQ, et al. (2010) The sequence and de novo assembly of the giant panda genome. *Nature* 463:311–317.
11. Zhang ZH, He GX, Wang XL, Zhong SL (1988) The study on the giant panda's intestinal flora. *Acta Ther Sin* 15:170–175.
12. Hirayama K, Kawamura S, Mitsuoka T, Tashiro K (1989) The faecal flora of the giant panda (*Ailuropoda melanoleuca*). *J Appl Bacteriol* 67:411–415.
13. Wei G, et al. (2007) The microbial community in the feces of the giant panda (*Ailuropoda melanoleuca*) as determined by PCR-TGGE profiling and clone library analysis. *Microb Ecol* 54:194–202.
14. Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* 71:1501–1506.
15. Collins MD, et al. (1994) The phylogeny of the genus *Clostridium*: Proposal of five new genera and eleven new species combinations. *Int J Syst Bacteriol* 44:812–826.
16. Whitford MF, Forster RJ, Beard CE, Gong JH, Teather RM (1998) Phylogenetic analysis of rumen bacteria by comparative sequence analysis of cloned 16S rRNA genes. *Anaerobe* 4:153–163.
17. Burrell PC, O'Sullivan C, Song H, Clarke WP, Blackall LL (2004) Identification, detection, and spatial resolution of *Clostridium* populations responsible for cellulose degradation in a methanogenic landfill leachate bioreactor. *Appl Environ Microbiol* 70:2414–2419.
18. Van Dyke MI, McCarthy AJ (2002) Molecular biological detection and characterization of *Clostridium* populations in municipal landfill sites. *Appl Environ Microbiol* 68: 2049–2053.
19. Yang SL, et al. (2010) Bacterial diversity in the rumen of Gayals (*Bos frontalis*), Swamp buffaloes (*Bubalus bubalis*) and Holstein cow as revealed by cloned 16S rRNA gene sequences. *Mol Biol Rep* 37:2063–2073.
20. Pope PB, et al. (2010) Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. *Proc Natl Acad Sci USA* 107:14793–14798.
21. Gill SR, et al. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312:1355–1359.
22. Li R, et al. (2010) De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* 20:265–272.
23. Noguchi H, Park J, Takagi T (2006) MetaGene: Prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* 34:5623–5630.
24. Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. *Genome Res* 17:377–386.
25. Warnecke FW, et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450:560–565.
26. Letunic I, Yamada T, Kanehisa M, Bork P (2008) iPath: Interactive exploration of biochemical pathways and networks. *Trends Biochem Sci* 33:101–103.
27. DeSantis TZ, Jr., et al. (2006) NAST: A multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* 34(Web Server Issue): W394–W399.
28. Huber T, Faulkner G, Hugenholtz P (2004) Bellerophon: A program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* 20:2317–2319.
29. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
30. Chun J, et al. (2007) EzTaxon: A web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. *Int J Syst Evol Microbiol* 57:2259–2261.
31. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
32. Cantarel BL, et al. (2009) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res* 37(Database Issue):D233–D238.
33. Finn RD, et al. (2010) The Pfam protein families database. *Nucleic Acids Res* 38(Database issue):D211–D222.