Intensive Bacterial Cultivation and Genome Assembly Reveal Previously Unknown Bacteria and Metabolic Potential in Karst Caves

Hai-Zhen Zhu[a,c,e], Zhi-Feng Zhang[d,e], Nan Zhou[a,c], Cheng-Ying Jiang[a,c], Bao-Jun Wang[a,c], Lei Cai[d], Hong-Mei Wang[b], Shuang-Jiang Liu[a,c,e#]

[a] State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, P. R. China;

[b] State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan, P. R. China;

[c] Environmental Microbiology Research Center, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, P. R. China;

[d] State Key Laboratory of Mycology at Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, P. R. China;

[e] University of Chinese Academy of Sciences, Beijing 100049, P. R. China

Corresponding author:

Shuang-Jiang Liu (liusj@im.ac.cn)

Institute of Microbiology, Chinese Academy of Sciences

Beichen Xilu No.1, Chaoyang District, Beijing 100101, China

Tel: +86-10-64807423; Fax: +86-10-64807421

Running title: Cultured microbiome from karst caves

Keywords: Bacterial cultivation, Karst cave microbiome, Biogeochemical cycling, 3-Oxoadipate-CoA-transferases, *Azospirillum*, *Oleomonas*.

Abstract: Karst caves are widely distributed subsurface systems, and the microbiomes therein are proposed to be the driving force for cave evolution and biogeochemical cycling. In past years, culture-independent studies on the microbiomes of cave systems have been conducted, yet intensive microbial cultivation is still needed to validate the sequence-derived hypothesis and to disclose the microbial functions in cave ecosystems. In this study, the microbiomes of two karst caves in Guizhou Province in southwest China were examined. A total of 3,562 bacterial strains were cultivated from rock, water, and sediment samples, and 329 species (including 14 newly described species) of 102 genera were found. We created a cave bacterial genome collection of 218 bacterial genomes from a karst cave microbiome through the extraction of 204 database-derived genomes and *de novo* sequencing of 14 new bacterial genomes. The cultivated genome collection obtained in this study and the metagenome data from previous studies were used to investigate the bacterial metabolism and potential involvement in the carbon, nitrogen, and sulfur biogeochemical cycles in the cave ecosystem. New N_2-fixing *Azospirillum* and alkane-oxidizing *Oleomonas* species were documented in the karst cave microbiome. Two *pcaIJ* clusters of the β-ketoadipate pathway that were abundant in both the cultivated microbiomes and the metagenomic data were identified, and their representatives from the cultivated bacterial genomes were functionally demonstrated. This large-scale cultivation of a cave microbiome represents the most intensive collection of cave bacterial resources to date and provides valuable information and diverse microbial resources for future cave biogeochemical research.

Importance: Karst caves are oligotrophic environments that are dark, humid, and have a relative stable annual temperature. The bacteria diversity and their metabolisms are crucial for understanding the biogeochemical cycling in cave ecosystems. We integrated large-scale bacterial cultivation with metagenomic data-mining to explore the composition and metabolisms of the microbiomes in two karst cave systems. Our results reveal the presence of a highly diversified cave bacterial community, and 14 new bacterial species were described and genome-sequenced. In this study, we obtained the most intensive collection of cultivated microbial resources from karst caves to date and predicted the various important routes for the biogeochemical cycling of elements in cave ecosystems.

Introduction

Karst caves are subterranean spaces that are mainly formed by the corrosion of soluble rocks such as limestone, dolomite, and gypsum. As relatively closed and extreme environments, caves are characterized by darkness, high humidity, comparably stable temperatures, and oligotrophic conditions (1). Nevertheless, rich and diversified microbiomes survive in caves (2-6). Culture-dependent and culture-independent studies have shown that *Proteobacteria* and *Actinobacteria* are abundant, and *Chloroflexi*, *Planctomycetes*, *Bacteroidetes*, *Firmicutes*, *Acidobacteria*, *Nitrospirae*, *Gemmatimonadetes*, and *Verrucomicrobia* also account for a significant proportion of the total microbial diversities in caves (7-9). Cave microbiomes play essential roles in the biogeochemical cycling of elements and in maintaining cave ecosystems. For example, *Acidithiobacillus thiooxidans* was dominant in the snottites

from Frasassi cave, and it is considered to provide the major energy and nutrient inputs for the sulfuric cave ecosystem (10). Other studies (11, 12) have revealed the diverse genes involved in nitrification, nitrate reduction, and denitrification. Recently, geobiological studies have suggested that caves contain abundant methanotrophic microbial communities and may be an atmospheric carbon sink because of the highly efficient methane oxidation performed by these microbes (13-15). Those conclusions are largely based on culture-independent studies. However, culture-dependent studies have shed light on cave microbial evolution and have provided new bioresources for the discovery of antibiotics. For example, the *Bacillus* species are involved in moonmilk and calcite formation (16, 17); the *Leptothrix* species are associated with ferromanganese deposits and have been cultivated from cave samples (18, 19); the *Streptomyces* strains from cave samples have exhibited strong inhibitory activities against gram-positive bacteria (20).

China has more than half a million caves that are integrated with the global subsurface system (21, 22). Many studies of microbial diversity have been conducted using culture-independent methods (3, 4, 7, 9, 10); however, the intensive cultivation of bacteria from the karst caves in China and around the world is rare. In this study, we studied two karst caves in southwestern China. Through intensive bacterial cultivation from rock, sediment, and water samples, we aimed to (i) discover previously unknown bacterial taxa and accumulate cave bioresources; and (ii) explore the bacterial metabolic potentials and involvements in cave biogeochemical cycles. We obtained 3,562 bacterial isolates and sequenced the genomes of 14 new bacterial

88 species. We integrated the newly cultured and available reference microbial genomes
89 and generated a cultured genome collection for karst cave microbiomes. Furthermore,
90 the involvement of the cultivated bacteria in biogeochemical C/N/S cycling in karst
91 cave environments was predicted through functional annotation of the cultured
92 genome collection and the mining of cultured-independent data from previous studies.
93 A new type of 3-oxoadipate-CoA-transferases, which was identified from the cultured
94 microbial genome collection, was biochemically and functionally characterized
95 through aromatic compound catabolism.

96 **Results**

97 **Bacterial cultivation and diversity**

98 Intensive and large-scale cultivation and identification of cave bacteria were
99 performed (Figure 1a). A total of 3,562 bacterial isolates were obtained, of which
100 1,408 and 2,154 isolates were obtained from Cave 1 and Cave 2, respectively (Dataset
101 S1). Cave 1 and Cave 2 are geographically close (500 m apart) and have similar
102 geological and climatic conditions. Through 16S rRNA gene sequencing and
103 phylogenetic analysis, the 3,562 bacterial isolates were assigned to 329 species in 102
104 genera (Dataset S2). Overall, 225 species and 201 species were obtained from Cave 1
105 and Cave 2, respectively, among which 97 species were found in both caves. The
106 Shannon index indicates that the cultured bacterial diversities of the two caves
107 exhibited no significant differences (Student's *t*-test, $p > 0.05$) (Figure 1b).
108 The bacterial isolates were also analyzed according to their origins in the cave
109 environments (i.e., rock, water, or sediments). The results revealed that 129 species

were isolated from rock samples, 155 were isolated from sediment samples, and 133 were isolated from water samples. The Shannon index analysis indicates that the species diversities were significantly different among the three environments (ANOVA, $F = 6.509$, $p < 0.01$), but similar distributions were observed when culture-independent methods were applied (7). The bacterial community in the sediment samples was more diverse than those in the rock samples (Tukey HSD, $p < 0.05$) and water samples (Tukey HSD, $p < 0.01$). Principal coordinates analysis based on the Brey-Curtis distance revealed that the community compositions of the three environments were statistically different (PERMANOVA, $F = 3.06$, $R^2 = 0.135$, $p = 0.001$, dotted line covers the 95% confidence interval) (Figure 1c).

Composition and representativeness of the cultured bacterial collections from the caves

Taking the isolates from both caves as a whole, the *Proteobacteria* were the most frequently isolated, followed by *Actinobacteria* and *Firmicutes*. *Bacteroidetes* and *Deinococcus–Thermus* were occasionally obtained (Dataset S2). At the genus level, the most abundant genera were *Brevundimonas* (13.7%), *Caulobacter* (6.3%), and *Bosea* (5.5%) of α-*Proteobacteria*; *Pseudomonas* (8.5%) of γ-*Proteobacteria*; *Streptomyces* (9.9%) and *Rhodococcus* (7.3%) of *Actinobacteria*; and *Bacillus* (8.8%) of *Firmicutes* (Figure 2a).

Taking 97% similarity in the 16S rRNA genes as the threshold for species differentiation, 166 isolates represented potential new bacterial taxa, accounting for 4.7% of all of the isolates (Table S1). These new taxa belonged to the following

genera: *Arthrobacter*, *Azospirillum*, *Brevundimonas*, *Deinococcus*, *Massilia*, *Methylibium*, *Nocardioides*, *Noviherbaspirillum*, *Oleomonas*, *Paenibacillus*, *Paenisporosarcina*, *Piscinibacter*, *Pseudogulbenkiania*, *Pseudomonas*, *Solimonas*, *Sphingomonas,* and *Zavarzinia* (Figure S1). Notably, the isolates representing *Azospirillum* and *Oleomonas* were repeatedly obtained (Table S1), suggesting that they were abundant in the cave environments. To further evaluate the representativeness of our isolates in terms of karst cave microbiomes, 4 culture-independent 16S rRNA gene amplicon datasets (NCBI accession Nos. PRJNA337918, PRJNA497480, PRJNA588777, and PRJNA630353; Dataset S6) from karst caves were collected and the samples were filtered for quality control. Among these datasets, samples of PRJNA497480 were collected from another 8 karst caves in southwestern China (7), and their geological backgrounds are very similar to those of the two caves investigated in this study. These 4 datasets include 153 samples, and the operational taxonomic units (OTUs) extracted from these samples were aligned with the 16S rRNA genes of the 3,562 cave isolates (species cut-off value set as a 97% 16S rRNA gene similarity). The result show that in terms of relative abundances, the 3,562 isolates represent 28.7% to 31.1% of the sequences on average and 75% for the highest sample in the 4 datasets (Figure 2b).

Morphology, genome annotation, and denomination of the new bacterial species

Twenty-four representative strains of the 166 potentially new isolates (Figure S1) were checked for purity, and 16S rRNA gene online alignment was performed using up-to-date databases (EzBiocloud and NCBI blast). Unfortunately, the bacterial

154 isolates representing 7 potential novel species were unable to propagate during the
155 subsequent cultivation. Isolates K2R10-124 and K2W31S-24 exhibited more than 98%
156 16S rRNA gene similarity to previously described species. Isolates K1W22B-3 and
157 K1W22B-8 exhibited 99% 16S rRNA gene similarity to each other, and they were
158 assigned as representative strains of one new species. The remaining 14 potential new
159 species were subjected to microscopic observations, phenotype determination using
160 BIOLOG testing, phylogenetic analysis, and genome sequencing. Their morphologies
161 and phylogenies are shown in Figures 3a and 3b, respectively, and their proposed
162 names are listed in Table 1. Detailed descriptions of the new species are provided in
163 Dataset S3, except for *Solimonas fluminis* K1W22B-7 and *Crenobacter cavernae*
164 K1W11S-77^T, which have been previously described (23, 24).
165 We tested the abilities of these potential new species to assimilate carbon sources. As
166 is shown in Figure 4a, short-chain fatty acids and amino acids were more frequently
167 assimilated than carbohydrates, particularly polysaccharides, although some
168 monosaccharides such as D-fructose, D-fructose-PO_4, and D-glucose were assimilated
169 by approximately half of the tested strains. Other carbon sources such as
170 glucuronamide, glycerol, and Tween 40 were also favored by the majority of the novel
171 cave bacteria. The general genome features of the new bacterial species are listed in
172 Table 2. As is shown in Table 2, the genome sizes of these potential new species range
173 from 2.5 to 6.5 Mb, coding 2,507 to 5,725 proteins. The Clusters of Orthologous
174 Groups (COG) database was used for the classification of the genes in the sequenced
175 genomes (Figure 4b). The results revealed that the highest number of genes contained

by these genomes are associated with transcription (COG-K), translation (COG-J), as well as DNA replication and repair (COG-L) for information storage and processing. For cellular processes and signaling, the genes involved in cell wall/membrane/envelope biogenesis and signal transduction are commonly abundant in sequenced genomes. Based on our analysis of the genes associated with metabolism, we found that the cave bacteria preferred carbon sources composed of amino acids (COG-E) and lipids (COG-I), which agreed with the results shown in Figure 4a. We observed that energy production and conversion (COG-C) and inorganic ion transport and metabolism (COG-P) were also abundant in the cave bacterial genomes. Noticeably, a large quantity of the genes in these new bacterial genomes are poorly characterized and their functions remain to be identified (COG-S).

Eleven of the 14 new species have flagella, and the genome data mining predicted that they have the capability for locomotive organ generation (Figure 4b). For bacteria living in complicated and nutrient-limited environments, the ability to migrate toward favorable environments (chemotaxis) is of importance for survival. We observed that the genes for chemoreceptors, histidine kinase CheA, and adaptor CheW occurred in 11 of the genomes of the new bacteria, and the number of chemoreceptor genes ranged from 2 (K2R01-6 and K1W22B-7) to as many as 46 (K2W22B-5). Biofilm formation has also been reported in regard to the survival of cave bacteria (25-27). Nine of the newly sequenced cave bacterial genomes have genes encoded for polysaccharide biosynthesis. Cross-talk between chemotaxis and biofilm formation

198 has also been reported recently (28), which indicates that coordination of bacterial
199 behavior may occur in cave microbiomes.

200 **Cultured bacterial genomes and metagenomic data predict metabolisms relevant**
201 **to biogeochemical cycling in karst caves**

202 To give an overview of the functional potential of the cultured bacteria from the karst
203 caves, a collection of cave bacterial genomes was established. The collection contains
204 14 newly sequenced bacterial genomes (Table 2) and 204 database-derived genomes,
205 representing the bacterial species found in the cave isolates in this study (Dataset S4).
206 These genomes covered 218 of the species found in the cultured bacterial collection,
207 and accounted for 72.3% of all of the isolates in terms of their relative
208 culture-frequencies. A total of 1,060,824 genes were recognized by CD-HIT and were
209 finally clustered as a non-redundant gene catalog containing 857,889 representative
210 sequences. The non-redundant cave gene catalog was annotated according to the
211 Kyoto Encyclopedia of Genes and Genomes (KEGG), and 7,476 KEGG orthologous
212 (KOs) were identified (Dataset S5). The genes involved in genetic information
213 processing (14.6%) accounted for the largest proportion, followed by signaling and
214 cellular process (11.5%), carbohydrate metabolism (9.4%), amino acid metabolism
215 (7.7%), energy metabolism (4.1%), and other metabolic processes. In addition, we
216 collected 8 metagenome datasets for karst cave sediment, speleothem, and rock
217 surface samples from previous studies (Dataset S6). The datasets were quality
218 controlled, re-annotated, and analyzed. The KOs related to the biogeochemical C/N/S
219 cycling in karst caves were checked in both the cultured genome collection and the

metagenomic data. Combined with the relative culture-frequencies of the bacterial isolates, the cultured genome collection and the metagenomic data were used to predict metabolic traits relevant to C/N/S cycling in karst caves.

Carbon metabolism. Analyses of the cultured genome collection and the gene catalog of the cave bacteria revealed that poly-β-hydroxybutyrate (PHB) and aromatic compounds may play important roles in biogeochemical carbon cycling in karst caves (Figure 5a). A total of 35.8% of the cultured bacteria in our genome collection contain genes for PHB synthesis, and 33.3% also contain genes for PHB depolymerization (Figure 5a). Previous studies have shown that stalagmite trapped poly aromatic hydrocarbons (29, 30), and aromatic compounds may serve as energy and carbon sources for cave systems. In our dataset, 4-hydroxybenzoate (4HB) degradation genes in the β-ketoadipate pathway were abundant, but the genes encoding 3-oxoadipate-CoA-transferases were missing (Figure 5a). We also found that 57 of the genomes, accounting for 26% of all of the isolates, were encoded with genes for carbon monoxide (CO) oxidation (Dataset S6). Although CO is toxic due to its ability to bind metalloproteins, it has a high potential as an electron donor, and thus, it may serve as a favorable carbon and/or energy source in extreme ecosystems (31-33) and in karst caves. Notably, the only cultured USCα bacterium (*Methylocapsa gorgona* MG08), which is a counterpart of the desired cave bacterial cluster (USCγ) in an acidic environment, has been proved to be able to use CO as an energy source (34). CO oxidation could be coupled with acetate or methane production under anaerobic conditions (35, 36), and under aerobic conditions, it could provide energy for CO_2

242 fixation through the Calvin-Benson-Bassham (CBB) cycle (37-39). When we mined
243 the genome dataset for the existence of the CBB pathway (40), we found the *rbcL*
244 gene in 30 genomes, accounting for 14.9% of the relative abundance.
245 The CO oxidation gene (*cox*) in karst cave bacterial genomes can be exemplified by
246 our newly sequenced *Oleomonas cavernae* K1W22B-8^T (Table S2 and Figure S2).
247 The K1W22B-8^T genome harbors CO dehydrogenase genes (*coxMSL*), the
248 membrane-integral ATPase gene (*coxD*), and the *xdhC*-like genes (*coxF* and *coxI*)
249 involved in the Mo=S group (41). However, it lacks the genes (*coxB*, *coxC*, *coxH,* and
250 *coxK*) that were identified in *Oligotropha carboxidovorans* OM5 (42) and are needed
251 to anchor CO dehydrogenase to the cytoplasmic membrane, suggesting that the CO
252 dehydrogenase in the K1W22B-8^T strain may be located in the cytoplasm.
253 Interestingly, a soluble methane monooxygenase-like gene cluster (*smoXYB1C1Z),*
254 which has been prove to be active on C_2 to C_4 alkanes and alkenes in *Mycobacterium*
255 *chubuense* NBB4 (43), was also found in the genome of the K1W22B-8^T strain
256 (Figure S2).
257 In accordance with the cultured genome collection, the analyses of the metagenomic
258 data revealed that the genes involved in PHB synthesis and depolymerization, 4HB
259 degradation, and CO oxidation were not only prevalent but were also abundant in
260 cave samples (Figure 5b). In contrast to the cultured genome collection, in which all
261 three genes involved in the conversion from acetyl-CoA to PHB were detected in 80
262 bacterial genomes, acetoacetyl-CoA reductase (PhaB, K00023) was absent in all eight
263 cave metagenome datasets. The distribution of the CO dehydrogenase varied among

the cave metagenome datasets for the different samples, and the Portuguese cave samples (NCBI Access. Nos. ERR1514431, ERR1514432, and ERR1514433) exhibit a higher CO oxidation potential than the cave samples from USA (NCBI Access. Nos. SRR12350322, SRR11676647, SRR11676930, and SRR11678124) and India (NCBI Access. No. SRR9599867).

Nitrogen metabolism. Based on our analysis, the NtrC family two-component-system was distributed in 88 of the genomes in the cave bacterial genome collection, suggesting an intensive regulation of nitrogen metabolism. Eleven of the genomes in our dataset exhibited the potential to fix dinitrogen into biologically available ammonia (Figure 5c). The novel *Azospirillum cavernae* K2W22B-5^{T}, which was isolated from the water samples and has a high abundance, is representative of these 11 genomes. The genome of strain K2W22B-5^{T} contains all three key operons for nitrogen fixation, i.e., *nifHDK*, *nifENX*, and *nifUSV* (Figure S2), which are needed for encoding the structural part of nitrogenase, the nitrogenase molybdenum-cofactor, and the Fe-S cluster, respectively (44, 45). Similar to the genetic organization in other *Azospirillum* species, there is an *fdxB* gene (*nif*-specific ferredoxin III) downstream of the *nifENX* operon, and a *cysE* gene (serine O-acetyltransferase) between the *nifUSV* operon and the *nifW* gene (nitrogenase-stabilizing/protective protein) (46). Nitrogen fixation demands a large amount of adenosine triphosphate (ATP), and diazotrophic bacteria have several hydrogenase systems to oxidize the nitrogen fixation byproduct hydrogen (47). The oxygen tolerant (NiFe)-hydrogenase is wide spread in the domain of bacteria (48), and its coding genes (*hyaAB*) was also found in the genome of strain

286 K2W22B-5^T.

287 More than 50% of the cultured bacteria have the potential to perform one or two steps
288 of dissimilatory nitrate reduction (Figure 5c). The gene cluster responsible for the
289 reduction of dissimilatory nitrate to nitrite in the genome of strain K2W22B-5^T is
290 *napABCDE*, which encodes the enzyme needed to reduce nitrate in periplasm.
291 However, more of the genomes in our dataset contain *narGHI* genes, which encode a
292 membrane-bound nitrate reductase capable of directly producing a proton motive
293 force during the reduction process (49). The reduction of dissimilatory nitrite to
294 ammonia is encoded by *nirBD*; and in the genome of strain K2W22B-5^T, the genes
295 for nitrate/nitrite transport are encoded by *nrtABCD* (Figure S2 and Table S2).
296 In contrast, analysis of the metagenomic data did not find complete dissimilatory
297 nitrate reductase. Either the gamma subunit of the membrane-bound nitrate reductase
298 (Nar I, K00374) or the electron transfer subunit of the periplasmic nitrate reductase
299 (NapB, K02568) was missing. Nitrite reductases were prevalent and abundant in all 8
300 metagenomic datasets (Figure 5d). Nitrogenase exhibited different distributions in the
301 cave metagenome data, and it was being more abundant in Hawaiian cave samples
302 (NCBI Access. Nos. SRR12350322, SRR11676647, SRR11676930, and
303 SRR11678124) than in other samples.
304 **Sulfur metabolism.** Genes encoding dissimilatory sulfate reduction were rarely
305 detected; however, both the cultured cave bacterial genomes and the metagenomic
306 data contained encoded enzymes needed for assimilatory sulfate/sulfite reduction or
307 the reduction of thiosulfate to sulfide (Figures 5e and 5f). Based on our analysis, 71.1%

and 36.1% of the genome collection (Figure 5e) exhibited the potential for the oxidation of thiosulfate and sulfite, respectively, suggesting that thiosulfate and sulfite may be important molecules for the biogeochemical cycling of sulfur in karst caves.

Validation of the β-ketoadipate pathway and identification of the "missing" 3-oxoadipate-CoA-transferase in the cultured bacterial genomes

As was predicted above, the β-ketoadipate pathway for aromatic compound degradation (assigned as the xenobiotics metabolism in the KEGG; Figure 6b) was quite abundant in both the cultured bacterial genomes and the metagenome data. Surprisingly, the genes encoding 3-oxoadipate-CoA-transferase (*pcaIJ*) were not annotated by the Automatic Annotation Server (KAAS) tools in the cultured genome collection (Figure 5a), but they were annotated in the metagenomics data (Figure 5f; Data accession numbers K01031/K01032). Thus, we extracted the annotated genes (K01031/K01032) from the metagenomic data and performed a blast search against the cultured bacterial genomes. The top 16 hits showing sequence identities of ≥47% were collected and considered as candidate *pcaIJ* of the cultured bacterial genomes (Table S3). Based on the NCBI and KEGG annotations, we further manually screened the cultured genome data for any continuous genetic clusters of the β-ketoadipate pathway and any *pcaIJ* candidates. We obtained a total of 55 genomes that harbored candidate 3-oxoadipate-CoA-transferase genes within the genetic clusters of the β-ketoadipate pathway. Two representative genetic clusters from the genomes of strains K2W22B-5^{T} and K1R23-30^{T} are shown in Figure 6a. The sequences of *pcaI* and *pcaJ,* which encode the two subunits of 3-oxoadipate-CoA-transferase, were

extracted from 55 genomes and were concatenated for phylogenetic analysis. The results revealed that the candidate 3-oxoadipate-CoA-transferase genes were grouped into two clusters. Cluster I was composed of 26 candidate genes, which mainly originated from the *Pseudomonas* species that has been extensively investigated for aromatic compound degradation. Cluster II was composed of 29 candidate genes, and their hosts were very diverse (Figure 6b). We tested two strains (K2W22B-5^T and K1R23-30^T) and confirmed that both were able to grow with 4-hydroxybenzoate as the sole carbon source (Figure S3). We further cloned and expressed their candidate *pcaIJ* in *E. coli*. The expressed PcaIJ products were purified and 3-oxoadipate-CoA-transferase activities were demonstrated (Figure 6d).

Discussion

In this study, we performed large-scale, intensive cultivation of cave microbiomes, and 3,562 bacterial isolates representing 329 species were obtained. Previous studies of the cultivation of cave bacteria have suggested that the cultivation of cave microorganisms could be challenging because the conventional culture media used in labs would result in osmotic stress on cave bacterial cells that are adapted to nutrient-poor cave environments (50). To increase the cultivability of cave bacteria, we used the R2A medium, which has been demonstrated to be effective for oligotrophs (51-53). We also adopted a strategy that transferred all of the visible colonies for sequential cultivation. Although this strategy was laborious and contained a bias that could possibly be overcome by using diluted nutrient culture media, lower temperatures, or an extended cultivation time, we still obtained the largest collection

of cave bacteria to date. Based on the evaluation using the 16S rRNA gene abundance, our cave isolates represent 75% for the highest and about 28.7–31.1% on average of the 16S rRNA gene abundances from previous datasets for karst caves. This result verifies that our cultures representative the major microbial community in karst caves relatively well. Our culture collection is characterized by the predominance of the *Proteobacteria*, *Actinobacteria,* and *Firmicutes* members, but it also contains other bacterial groups found in cave habitats, including *Bacteroide* and *Deinococcus–Thermus*. Notably, *Proteobacteria* and *Actinobacteria* represent the most ubiquitous bacterial groups detected in cave environments (54-56). At the genus level, *Brevundimonas* of *Proteobacteria* was most frequently cultivated in this study, and it has been found to be abundant in other oligotrophic caves (57). The genus *Streptomyces* of *Actinomycetes* was also predominant in this study, and the members of the cave-originated *Streptomyces* have been used for the selection of new antibiotics (58). Although the two caves we studied have not been open to tourists, they both contained *Bacillus* and *Paenibacillus* of the phylum *Firmicutes,* which have also been found in a cave open to tourists, i.e., Kartchner Caverns (54).

Microbial metabolisms are the major driving force of biogeochemical cycling in cave ecosystems. The results of culture-independent methods have predicted the general metabolic reactions of these microbial communities, but which organism plays what role remains to be specified. In this study, we collected 204 cultured bacterial reference genomes from public databases that corresponded to our bacterial isolates and sequenced 14 new bacterial species. These 218 bacterial genomes were analyzed

374 to dissect their specific metabolic traits that are relevant to the biogeochemical cycling
375 of C/N/S in cave environments. For examples, the CO oxidation and N_2 fixation
376 abilities of the newly cultivated *Oleomonas* and *Azospirillum* species, respectively,
377 may reduce carbon and nitrogen limitations in cave environments. In nutrient limited
378 habitats, microorganisms are forced to use any available nutrient to survive (59). A
379 range of bacteria in Movile Cave were able to grow on one-carbon (C_1) compounds
380 (60). In addition to *Oleomonas* species, we also obtained facultative methylotrophic
381 bacteria such as *Methylorubrum aminovorans*, *M. thiocyanatum*, *Methylobacterium*
382 *hispanicum,* and *Methylibium petroleiphilum*. Recently, a clade of uncultured
383 methanotrophs that are believed to have a high affinity for oxidizing atmospheric
384 methane in caves have received a great deal of attention (15). Although methane
385 oxidization was not confirmed, the *Oleomonas* species found in our study exhibit the
386 potential to oxidize C_2 to C_4 alkanes, providing a new perspective for research on
387 alkane oxidation in cave environments. Regarding nitrogen limitation, evidence has
388 been found for the existence of nitrogen fixation genes in other cave water niches (61).
389 We determined that more than 6% of all of the isolated strains, including the newly
390 cultivated *Azospirillum* species, have the potential to fix N_2 into ammonia. Notably,
391 the *Azospirillum griseum* in eutrophic river water (62), which is the closest
392 phylogenetic neighbor of the newly cultivated cave *Azospirillum* species, does not
393 contain any nitrogen fixing genes. Future studies of these two *Azospirillum* species
394 may provide hints as to the evolution of nitrogen fixation at the genomic level.
395 The β-ketoadipate pathway is widely distributed in soil bacteria and fungi (63), but it

has not been documented in the microbiomes in karsts caves. In this study, we observed abundant genes encoding the β-ketoadipate pathway in both the cultivated bacterial genomes and the previously reported metagenomic datasets (NCBI Access. Nos. ERR1514431, ERR1514432, ERR1514433, SRR9599867, SRR12350322, SRR11676647, SRR11676930, and SRR11678124). We further found that the *pcaIJ* genes from 55 of the cultivated genomes grouped into two clusters according to their sequences, and we experimentally identified the 3-oxoadipate-CoA-transferase activities of two of the newly cultivated representative bacterial strains. The results of this study demonstrate the power of studies conducted using a combination of culture-dependent and metagenomic methods, and the *pcaIJ* sequences of the two clusters provide highly valuable information for improving future *pcaIJ* annotation using the KAAS tools.

Materials and methods

Caves. All of the samples were collected from two unexploited karst caves designated as Cave 1 (28º12´37.74" N; 107º13´38.34" E) and Cave 2 (28º12´35.94" N; 107º13´39.66" E) in the Kuankuoshui Nature Reserve, Zunyi, Guizhou Province, China. The nature reserve was established in 2007 due to the subtropical forests and rare animals it contains. Except for its clasolite-based erosional landform in the central-southern areas, the nature reserve predominantly contains karst landforms developed from carbonate rocks. The annual average temperature of the nature reserve is 11.6–15.2°C, and the annual average relative humidity is more than 82% (64). Both Cave 1 and Cave 2 are horizontally zonal, and each has only one entrance

hidden on a hillside in the forest. Cave 1 is 908 m above sea level and 400 m in length, and the humidity and temperature at the time of the sampling were 75–80% and 21–22°C, respectively. Cave 2 is 930 m above sea level and 750 m in length, and the humidity and temperature at the time of sampling were 75–85% and 20–23°C, respectively.

Sample collection. The sampling procedure has been described by Zhang et al. (65). The samples were collected from the entrance to the deep part of the cave, and each sampling site was at least 100 m from the next site. Briefly, 10 ml of seeping or stream water were collected in 15 ml sterile centrifuge tubes at each site. Ten grams of shallow sediment (~1–5 cm) were collected from three sites after removing the surface layer (~1 cm). Rock samples were collected from five different orientations at each sampling site and were sealed in germfree zip-locked bags (66). All of the samples were kept at 4°C until further processing. A total of 42 samples were obtained from the two caves (Cave 1 and Cave 2), of which 20 samples were collected from Cave 1 (4 sediment, 8 water, and 8 rock samples) and 22 samples were collected from Cave 2 (6 sediment, 11 water, and 5 rock samples).

Bacterial isolation and cultivation. Two grams of sediment sample were suspended in 18 ml of sterile saline solution (NaCl, 0.85%, m/v) and were shaken for 30 min at room temperature. Two milliliters of a water sample were added to 18 ml of sterile saline solution and mixed thoroughly. The rock samples were weighed and placed in enough sterile saline solution to achieve a weight to volume ratio of 1:10, and then, they were shaken for 30 mins at room temperature. Ten-fold serial dilutions were

using sterile saline solution and 0.2 ml of the diluent with an appropriate concentration was spread on R2A medium (Reasoner's 2A agar) (67) in triplicate. The spread plates were incubated at 30°C for 48–72 h, and then, the colonies were picked and re-streaked to confirm their purity.

Identification of the cave bacteria. Amplification of 16S rRNA genes was accomplished using universal bacterial primers 27F (5'–AGAGTTTGATCTGGCTCAG–3', corresponding to positions 8 to 27 of *E. coli*) and 1492R (5'–GGTTACCTTGTTACGACTT–3', corresponding to positions 1510 to 1492 of *E. coli;*). The cells were collected from the agar plates and lysed in 2 µl of alkaline lysis solution (0.2 M NaOH, 1% SDS) for 5 min, and then, 98 µl of double distilled water was added to the lysis system and was mixed thoroughly as an amplification template.

The 50 µl of polymerase chain reaction (PCR) mixture contained 1 µl of template, 1 µl (10 pmol) of each primer, and 47 µl of 1.1 × Golden Star T6 Super PCR Mix (TsingKe Biotech. Beijing). The amplification conditions were as follows: initial denaturation (2 min at 94°C), 30 cycles of denaturing (30 sec at 94°C), annealing (30 sec at 55°C), extension (1 min at 72°C), and a final extension (72°C for 5 min). Five microliters of PCR products were visualized on a 1% agarose gel stained with YeaRed Nucleic Acid Gel Stain (Yeasen Biotech, Shanghai).

The amplified 16S rRNA genes were sequenced and then aligned using blast+ against NCBI's 16SMicrobial database (68). The biochemical characteristics of the novel species were determined using Biolog GEN III kits according to the manufacturer's

instructions. The average nucleotide identity (ANI) value between new species and their close relatives were calculated using ChunLab's online ANI Calculator (69). Digital DNA-DNA hybridization (dDDH) was performed on the novel species and their close relatives using the Genome-to-Genome Distance Calculator (GGDC2.1) (70).

Diversity and phylogenetic analysis. The diversity indices were calculated using the free license statistical software PAST (71). All of the statistical analyses of the data were performed in R version 3.4.2 (https://www.R-project.org/). The normal distributions of the data were checked using the Shapiro-Wilk test, and the homoscedasticity of variances was analyzed using Bartlett's test. The significant differences in the variances of the parameters were evaluated using the analysis of variance (ANOVA) test or the Student's t-test, and post hoc comparisons were conducted using Tukey's honest significant differences test. The principal coordinates analysis (PCoA) was conducted using the vegan package in R (https://CRAN.R-project.org/package=vegan). To statistically support the visual clustering of the bacterial communities in the PCoA analyses, the different cave substrates were compared using permutation-based hypothesis tests (PERMANOVA). Visualization of the diversity and distributions of the cave isolates was performed using the ggplot2 package in R unless otherwise stated (72). The Venn diagrams were plotted using the VennDiagram package in R (73).

The phylogenetic trees were established using the neighbor-joining algorithm. The relative evolutionary distances among the sequences were calculated using the

Kimura 2-parameter model, and the tree topology was statistically evaluated using 1000 bootstrap resampling (74). The phylogenetic trees were constructed using the MEGA7 software (75), and they were further modified using iTOL (Interactive Tree Of Life) (76).

Whole genome sequencing and functional annotation. The genomic DNA was extracted using a Wizard Genomic DNA Purification Kit (Promega, USA) according to the manufacturer's instructions, and then, it was sheared into 10 kb segments using a Covaris g-TUBE (Covaris, USA). AMPure XP beads (Beckman Coulter, USA) were used to purify the segmented DNA, and a PacBio SMRTbell Template Prep Kit (PacBio, USA) was used to prepare the segments for sequencing. The SMRTbell templates were annealed with primers and combined with polymerase using a PacBio DNA/Polymerase Kit (PacBio, USA), and finally, they were sequenced on a PacBio RS II platform.

The sequence assembly was performed in the PacBio SMRT Analysis version 2.3.0 platform using the RS_HGAP_Assembly.2 protocol (77). FinisherSC was subsequently used to further polish the assemblies (78). The final assemblies were annotated following the NCBI Prokaryotic Genome Annotation Pipeline (79), and their metabolic potentials were predicted using the KEGG Automatic Annotation Server (KAAS) (80) and the eggnog mapper (81). The completeness of each bacterial genome was evaluated using BUSCO (82). The non-redundant gene catalog of the cultured cave bacteria was obtained using CD-HIT (83). Amino acid sequences with more than 90% similarity and 80% coverage were assigned as one cluster.

16S rRNA gene amplicon and metagenome analysis. 16S rRNA amplicon and the metagenomes of the cave samples were downloaded from the NCBI Sequence Read Archive (SRA) using the sra-toolkit v2.8.2. For the 16S rRNA gene amplicon analysis, VSEARCH v0.9.11 was used to merge paired end sequences and for quality control (fastq_maxee = 0.01) (84). The singletons and chimeras were removed, and the OTUs were obtained using the UNOISE algorithm in USEARCH v11.0.667 (85, 86). Non-bacterial sequences and sequences representing OTUs with an average relative abundance of less than 0.00001 were filtered out using QIIME v1.9.1 (87). Blast+ v2.10.1 was used to construct the cultured cave bacteria 16S rRNA gene database and to align the amplicon data against this database (68).

The quality control of the metagenome data was performed using KneadData v0.7.4 (http://huttenhower.sph.harvard.edu/kneaddata), a sliding window was set as 4 bp to filter bases with a quality value of less than 20, and the filtered sequences with a length of less than 50 bp were dropped. Samples with less than 10,000 reads after quality control were removed. The resulting sequences were assembled using MEGAHIT v1.2.9 (88). Prokka v1.14.6 was used for the gene annotation (89), and then, CD-HIT v4.8.1 was used to construct a non-redundant gene catalog (83). The nucleotide sequences in the gene catalog were translated into amino acid sequences using EMBOSS v6.6.0 (90), and then, they were functionally annotated using eggnog-mapper v2.0.1 (81). Salmon v1.3.0 was used to quantify the genes in each sample (91).

3-oxoadipate-CoA transferase expression, purification, and activity assay. The

528 bacterial strains, plasmids, and primers used for the 3-oxoadipate-CoA transferase
529 expression are listed in Table 3. The genomic DNA of strains K2W22B-5^T and
530 K1R23-30^T was prepared as described above. PCR amplification of the target DNA
531 fragments was performed using Phusion High-Fidelity DNA polymerase (New
532 England Biolabs, USA). The vector plasmids and DNA fragments were digested using
533 restriction endonucleases Nde I and Hind III (New England Biolabs, USA), and then,
534 they were ligated using T4 DNA ligase (New England Biolabs, USA). After the
535 ligation, *pcaI* and *pcaJ* were given a 6xHis-tag at N-terminus and C-terminus,
536 respectively.
537 To prepare the 3-oxoadipate-CoA transferase of strains K2W22B-5^T and K1R23-30^T,
538 *E. coli* BL21 (DE3) carrying pET-28a-k5pcaIJ and pET-28a-30pcaIJ were grown in
539 Luria-Bertani (LB) broth supplemented with 50 μg/ml of kanamycin at 37°C until the
540 cell density (OD600) reached 0.3–0.4. The protein expression was induced using 0.3
541 mM IPTG at 16°C overnight. The cells were harvested through centrifugation and
542 then, they were lysed using ultrasonication. The protein purification was performed
543 with a Hisbind purification kit (Novagen, USA) following the manufacturer's
544 instructions. An Amicon Ultra-15 centrifugal filter (Merck Millipore, USA) was used
545 for the buffer desalting and protein concentration.
546 The 3-oxoadipate-CoA transferase assays were performed as described by MacLean et
547 al. (92). The assay mixture included 200 mM Tris-HCl (pH 8.0), 40 mM $MgCl_2$, 10
548 mM 3-oxoadipate, and 0.4 mM succinyl-CoA (Sigma-Aldrich, USA) with a final
549 volume of 200 μl (path length, 0.52 cm). Ninety-six well microtiter plates with UV

transparent, flat bottoms (Corning, USA) and a multimode plate reader (PerkinElmer, USA) were used to monitor the formation of 3-oxoadipyl-CoA, with Mg^{2+} at 305 nm over a temperature range of 23 to 24°C. The molar extinction coefficient of 16,300 M^{-1} cm^{-1} corresponding to the 3-oxoadipyl-CoA:Mg^{2+} complex was used to calculate the productivity (93).

Data availability. The 16S rRNA genes of the cave bacterial isolates in this study are presented in Dataset S1. The 14 newly sequenced cave bacterial genomes have been deposited in the NCBI GenBank and are available under BioProject PRJNA490657 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA490657). The accessions of all of the bacterial genomes analyzed in this study are presented in Dataset S3. The accessions and sample descriptions of the 16S rRNA gene amplicon and metagenome data used in this study are presented in Dataset S7. The representative strains of the previously described bacterial species obtained in this study are publicly available in the China General Microbiological Culture Collection Center (CGMCC), and the accession numbers of each strains are listed in Dataset S8.

Acknowledgements

This study was financially supported by the National Natural Science Foundation of China (Grant No. 91951208) and the Ministry of Science and Technology of China (Grant No. 2014FY120100). The funders had no role in the study's design, data collection, and interpretation or in the decision to submit the work for publication.

571 **References**

572 1. Gabriel CR, Northup DE. 2013. Microbial ecology: caves as an extreme
573 habitat, p 85-108. *In* Cheeptham N (ed), Cave microbiomes: a novel resource
574 for drug discovery. Springer, New York.
575 2. Kuzmina LY, Galimzianova NF, Abdullin SR, Ryabova AS. 2012. Microbiota
576 of the Kinderlinskaya cave (South Urals, Russia). Microbiology 81:251-258.
577 3. Ortiz M, Neilson JW, Nelson WM, Legatzki A, Byrne A, Yu Y, Wing RA,
578 Soderlund CA, Pryor BM, Rd PL. 2013. Profiling bacterial diversity and
579 taxonomic composition on speleothem surfaces in Kartchner Caverns, AZ.
580 Microb Ecol 65:371-383.
581 4. Hershey OS, Kallmeyer J, Wallace A, Barton MD, Barton HA. 2018. High
582 microbial diversity despite extremely low biomass in a deep karst aquifer.
583 Front Microbiol 9:13.
584 5. Barton HA, Jurado V. 2007. What's up down there? Microbial diversity in
585 caves, p 132-138. *In* Schaechter M (ed), Microbe. ASM Press, Washington.
586 6. Zhang Z-F, Cai L. 2019. Substrate and spatial variables are major determinants
587 of fungal community in karst caves in Southwest China. J Biogeogr 00:1-15.
588 7. Zhu H-Z, Zhang Z-F, Zhou N, Jiang C-Y, Wang B-J, Cai L, Liu S-J. 2019.
589 Diversity, distribution and co-occurrence patterns of bacterial communities in
590 a karst cave system. Front Microbiol 10:1726.
591 8. Rusznyák A, Akob DM, Nietzsche S, Eusterhues K, Totsche KU, Neu TR,
592 Frosch T, Popp J, Keiner R, Geletneky J, Katzschmann L, Schulze E-D, Küsel

K. 2012. Calcite biomineralization by bacterial isolates from the recently discovered pristine karstic Herrenberg cave. Appl Environ Microbiol 78:1157.

9. Wu Y, Tan L, Liu W, Wang B, Wang J, Cai Y, Lin X. 2015. Profiling bacterial diversity in a limestone cave of the western Loess Plateau of China. Front Microbiol 6:244.

10. Jones DS, Albrecht HL, Dawson KS, Schaperdoth I, Freeman KH, Pi Y, Pearson A, Macalady JL. 2012. Community genomic analysis of an extremely acidophilic sulfur-oxidizing biofilm. ISME J 6:158-170.

11. Kimble JC, Winter AS, Spilde MN, Sinsabaugh RL, Northup DE. 2018. A potential central role of *Thaumarchaeota* in N-cycling in a semi-arid environment, Fort Stanton Cave, Snowy River passage, New Mexico, USA. FEMS Microbiol Ecol 94:17.

12. Zhao R, Wang H, Yang H, Yun Y, Barton H. 2016. Ammonia oxidizing archaea dominant over bacterial ammonia oxidizers within alkaline cave sediments. Geomicrobiol J 34:511-523.

13. Nguyễn-Thuỳ D, Schimmelmann A, Nguyễn-Văn H, Drobniak A, Lennon JT, Tạ PH, Nguyễn NTÁ. 2017. Subterranean microbial oxidation of atmospheric methane in cavernous tropical karst. Chem Geol 466:229-238.

14. Webster KD, Drobniak A, Etiope G, Mastalerz M, Sauer PE, Schimmelmann A. 2018. Subterranean karst environments as a global sink for atmospheric methane. Earth Planet Sci Lett 485:9-18.

15. Zhao R, Wang H, Cheng X, Yun Y, Qiu X. 2018. Upland soil cluster gamma

dominates the methanotroph communities in the karst Heshang Cave. FEMS Microbiol Ecol 94:fiy192.

16. Baskar S, Baskar R, Mauclaire L, Mckenzie JA. 2006. Microbially induced calcite precipitation in culture experiments: possible origin for stalactites in Sahastradhara caves, Dehradun, India. Currentence 90:58-64.

17. Baskar S, Baskar R, Routh J. 2011. Biogenic evidences of moonmilk deposition in the Mawmluh cave, Meghalaya, India. Geomicrobiol J 28:252-265.

18. Peck SB. 1986. Bacterial deposition of iron and manganese oxides in North American caves. J Caves Karst Stud 48:26-30.

19. Hill CA. 2011. Origin of black deposits in caves. J Caves Karst Stud 44:15-19.

20. Adam D, Maciejewska M, Naome A, Martinet L, Coppieters W, Karim L, Baurain D, Rigali S. 2018. Isolation, characterization, and antibacterial activity of hard-to-culture Actinobacteria from cave moonmilk deposits. Antibiotics (Basel) 7:20.

21. Chen WH. 2006. An outline of speleology research progress. Geological Review 52:65–74. (In Chinese)

22. Zhang YH, Zhu DH. 2012. Large karst caves distribution and development in China. Journal of Guilin University of Technology 32:20–28. (In Chinese)

23. Lee Y, Lee B, Lee K, Jeon CO. 2018. *Solimonas fluminis* sp. nov., isolated from a freshwater river. Int J Syst Evol Microbiol 68:2755-2759.

24. Zhu H-Z, Jiang C-Y, Liu S-J. 2019. *Crenobacter cavernae* sp. nov., isolated

from a karst cave, and emended description of the genus *Crenobacter*. Int J Syst Evol Microbiol 69:476-480.

25. Macalady JL, Jones DS, Lyon EH. 2007. Extremely acidic, pendulous cave wall biofilms from the Frasassi cave system, Italy. Environ Microbiol 9:1402-1414.

26. Urzì C, De LF, Bruno L, Albertano P. 2010. Microbial diversity in paleolithic caves: a study case on the phototrophic biofilms of the Cave of Bats (Zuheros, Spain). Microb Ecol 60:116-129.

27. Banerjee S, Joshi SR. 2013. Insights into cave architecture and the role of bacterial biofilm. Proc Natl Acad Sci India Sect B Biol Sci 83:277-290.

28. Huang Z, Wang Y-H, Zhu H-Z, Andrianova EP, Jiang C-Y, Li D, Ma L, Feng J, Liu Z-P, Xiang H, Zhulin IB, Liu S-J. 2019. Cross talk between chemosensory pathways that modulate chemotaxis and biofilm formation. mBio 10:e02876-18.

29. Perrette Y, Poulenard J, Saber A-I, Fanget B, Guittonneau S, Ghaleb B, Garaudee S. 2008. Polycyclic aromatic hydrocarbons in stalagmites: occurrence and use for analyzing past environments. Chem Geol 251:67-76.

30. Marques ELS, Silva GS, Dias JCT, Gross E, Costa MS, Rezende RP. 2019. Cave drip water-related samples as a natural environment for aromatic hydrocarbon-degrading bacteria. Microorganisms 7:33.

31. Wu L, Wang R. 2005. Carbon monoxide: endogenous production, physiological functions, and pharmacological applications. Pharmacol Rev

57:585-630.

32. Grahame DA, DeMoll E. 1995. Substrate and accessory protein requirements and thermodynamics of acetyl-CoA synthesis and cleavage in *Methanosarcina barkeri*. Biochemistry 34:4617-4624.

33. Robb FT, Techtmann SM. 2018. Life on the fringe: microbial adaptation to growth on carbon monoxide. F1000Res 7.

34. Tveit AT, Hestnes AG, Robinson SL, Schintlmeister A, Dedysh SN, Jehmlich N, von Bergen M, Herbold C, Wagner M, Richter A, Svenning MM. 2019. Widespread soil bacterium that oxidizes atmospheric methane. Proc Natl Acad Sci U S A 116:8515-8524.

35. Sant'Anna FH, Lebedinsky AV, Sokolova TG, Robb FT, Gonzalez JM. 2015. Analysis of three genomes within the thermophilic bacterial species *Caldanaerobacter subterraneus* with a focus on carbon monoxide dehydrogenase evolution and hydrolase diversity. BMC Genomics 16:757.

36. Techtmann SM, Colman AS, Robb FT. 2009. 'That which does not kill us only makes us stronger': the role of carbon monoxide in thermophilic microbial consortia. Environ Microbiol 11:1027-1037.

37. Meyer O, Schlegel HG. 1983. Biology of aerobic carbon monoxide-oxidizing bacteria. Annu Rev Microbiol 37:277-310.

38. King GM. 2003. Molecular and culture-based analyses of aerobic carbon monoxide oxidizer diversity. Appl Environ Microbiol 69:7257-7265.

39. King GM, Weber CF. 2007. Distribution, diversity and ecology of aerobic

CO-oxidizing bacteria. Nat Rev Microbiol 5:107-118.

40. Sato T, Atomi H. 2010. Microbial inorganic carbon fixation, *In* Encyclopedia of life sciences. John Wiley & Sons, Ltd. Chichester.

41. Pelzmann A, Ferner M, Gnida M, Meyer-Klaucke W, Maisel T, Meyer O. 2009. The CoxD protein of oligotropha carboxidovorans is a predicted AAA+ ATPase chaperone involved in the biogenesis of the CO dehydrogenase [CuSMoO2] cluster. J Biol Chem 284:9578-9586.

42. Hille R, Dingwall S, Wilcoxen J. 2015. The aerobic CO dehydrogenase from oligotropha carboxidovorans. J Biol Inorg Chem 20:243-251.

43. Martin KE, Ozsvar J, Coleman NV. 2014. *SmoXYB1C*1Z of *Mycobacterium* sp. strain NBB4: a soluble methane monooxygenase (sMMO)-like enzyme, active on C-2 to C-4 alkanes and alkenes. Appl Environ Microbiol 80:5801-5806.

44. Frazzon J, Schrank IS. 1998. Sequencing and complementation analysis of the *nifUSV* genes from *Azospirillum brasilense*. FEMS Microbiol Lett 159:151-158.

45. Steenhoudt O, Vanderleyden J. 2000. *Azospirillum*, a free-living nitrogen-fixing bacterium closely associated with grasses: genetic, biochemical and ecological aspects. FEMS Microbiol Rev 24:487-506.

46. Sant'Anna FH, Almeida LG, Cecagno R, Reolon LA, Siqueira FM, Machado MR, Vasconcelos AT, Schrank IS. 2011. Genomic insights into the versatility of the plant growth-promoting bacterium *Azospirillum amazonense*. BMC Genomics 12:409.

703 47. Bothe H, Schmitz O, Yates MG, Newton WE. 2010. Nitrogen fixation and
704 hydrogen metabolism in *cyanobacteria*. Microbiol Mol Biol Rev 74:529-551.
705 48. Vignais PM, Billoud B. 2007. Occurrence, classification, and biological
706 function of hydrogenases: an overview. Chem Rev 107:4206-4272.
707 49. Moreno-Vivian C, Cabello P, Martinez-Luque M, Blasco R, Castillo F. 1999.
708 Prokaryotic nitrate reduction: molecular properties and functional distinction
709 among bacterial nitrate reductases. J Bacteriol 181:6573-6584.
710 50. Barton HA. 2006. Introduction to cave microbiology: a review for the
711 non-specialist. J Caves and Karst Stud 2:43–54.
712 51. Gibbs RA, Hayes CR. 1988. The use of R2A medium and the spread plate
713 method for the enumeration of heterotrophic bacteria in drinking water. Lett
714 Appl Microbiol 6:19-21.
715 52. Massa S, Caruso M, Trovatelli F, Tosques M. 1998. Comparison of plate count
716 agar and R2A medium for enumeration of heterotrophic bacteria in natural
717 mineral water. World J Microbiol Biotechnol 14:727-730.
718 53. van der Linde K, Lim BT, Rondeel JM, Antonissen LP, de Jong GM. 1999.
719 Improved bacteriological surveillance of haemodialysis fluids: a comparison
720 between tryptic soy agar and Reasoner's 2A media. Nephrol Dial Transplant
721 14:2433-2437.
722 54. Ikner LA, Toomey RS, Ginger N, Neilson JW, Pryor BM, Maier RM. 2007.
723 Culturable microbial diversity and the impact of tourism in Kartchner Caverns,
724 Arizona. Microb Ecol 53:30-42.

725 55. Iva T, Irina L, Anna T, Margarita K. 2013. Diversity and biosynthetic potential
726 of culturable aerobic heterotrophic bacteria isolated from Magura Cave,
727 Bulgaria. Int J Speleol 42:65-76.
728 56. Velikonja BH, Tkavc R, Pašić L. 2014. Diversity of cultivable bacteria
729 involved in the formation of macroscopic microbial colonies (cave silver) on
730 the walls of a cave in Slovenia. Int J Speleol 43:45-56.
731 57. Barton H, Taylor N, Kreate M, Springer A, Oehrle S, Bertog J. 2007. The
732 impact of host rock geochemistry on bacterial community structure in
733 oligotrophic cave environments. Int J Speleol 36:93-104.
734 58. Groth I, Vettermann R, Schuetze B, Schumann P, Saiz-Jimenez C. 1999.
735 Actinomycetes in karstic caves of northern Spain (Altamira and Tito Bustillo).
736 J Microbiol Methods 36:115-122.
737 59. Jamil SUU, Zada S, Khan I, Sajjad W, Rafiq M, Shah A, Hasan F. 2016.
738 Biodegradation of polyethylene by bacterial strains isolated from Kashmir
739 cave, Buner, Pakistan. J Caves Karst Stud 79:73-80.
740 60. Wischer D, Kumaresan D, Johnston A, El Khawand M, Stephenson J,
741 Hillebrand-Voiculescu AM, Chen Y, Colin Murrell J. 2015. Bacterial
742 metabolism of methylated amines and identification of novel methylotrophs in
743 Movile Cave. ISME J 9:195-206.
744 61. Desai MS, Assig K, Dattagupta S. 2013. Nitrogen fixation in distinct microbial
745 niches within a chemoautotrophy-driven cave ecosystem. ISME J
746 7:2411-2423.

747 62. Yang Y, Zhang R, Feng J, Wang C, Chen J. 2019. *Azospirillum griseum* sp.
748 nov., isolated from lakewater. Int J Syst Evol Microbiol 69:3676-3681.
749 63. Harwood CS, Parales RE. 1996. The beta-ketoadipate pathway and the
750 biology of self-identity. Ann Rev Microbiol 50:553-590.
751 64. China TsfaotPsRo. Natural conditions of Kuankuoshui nature reserve of China.
752 2014-04-03.
753 65. Zhang ZF, Liu F, Zhou X, Liu XZ, Liu SJ, Cai L. 2017. Culturable mycobiota
754 from karst caves in China, with descriptions of 20 new species. Persoonia
755 39:1-31.
756 66. Ruibal C, Platas G, Bills GF. 2005. Isolation and characterization of melanized
757 fungi from limestone formations in *Mallorca*. Mycol Prog 4:23-38.
758 67. Reasoner DJ, Geldreich EE. 1985. A new medium for the enumeration and
759 subculture of bacteria from potable water. Appl Environ Microbiol 49:1-7.
760 68. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K,
761 Madden TL. 2009. BLAST+: architecture and applications. BMC
762 Bioinformatics 10:421.
763 69. Yoon S-H, Ha S-m, Lim J, Kwon S, Chun J. 2017. A large-scale evaluation of
764 algorithms to calculate average nucleotide identity. Antonie van Leeuwenhoek
765 110:1281-1286.
766 70. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. 2013. Genome
767 sequence-based species delimitation with confidence intervals and improved
768 distance functions. BMC Bioinformatics 14:60.

769 71. Hammer Ø, Harper D, Ryan P. 2001. PAST: paleontological statistics software
770 package for education and data analysis. Palaeontologia Electronica 4:1-9.
771 72. Wickham H. 2009. Ggplot2: elegant graphics for data analysis. Springer. New
772 York.
773 73. Chen H, Boutros PC. 2011. VennDiagram: a package for the generation of
774 highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics
775 12:35.
776 74. Kimura M. 1980. A simple method for estimating evolutionary rates of base
777 substitutions through comparative studies of nucleotide sequences. J Mol Evol
778 16:111-120.
779 75. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary
780 genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870-1874.
781 76. Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. 2006.
782 Toward automatic reconstruction of a highly resolved tree of life. Science
783 311:1283.
784 77. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A,
785 Copeland A, Huddleston J, Eichler EE. 2013. Nonhybrid, finished microbial
786 genome assemblies from long-read SMRT sequencing data. Nat Methods
787 10:563.
788 78. Lam KK, LaButti K, Khalak A, Tse D. 2015. FinisherSC: a repeat-aware tool
789 for upgrading de novo assembly using long reads. Bioinformatics
790 31:3207-3209.

791 79. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky
792 L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic
793 genome annotation pipeline. Nucleic Acids Res 44:6614-6624.
794 80. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an
795 automatic genome annotation and pathway reconstruction server. Nucleic
796 Acids Res 35:W182-185.
797 81. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering
798 C, Bork P. 2017. Fast genome-wide functional annotation through orthology
799 assignment by eggNOG-mapper. Mol Biol Evol 34:2115-2122.
800 82. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015.
801 BUSCO: assessing genome assembly and annotation completeness with
802 single-copy orthologs. Bioinformatics 31:3210-3212.
803 83. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing
804 large sets of protein or nucleotide sequences. Bioinformatics 22:1658-1659.
805 84. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a
806 versatile open source tool for metagenomics. PeerJ 4:e2584.
807 85. Edgar RC. 2010. Search and clustering orders of magnitude faster than
808 BLAST. Bioinformatics 26:2460-1.
809 86. Edgar RC. 2016. UNOISE2: improved error-correction for Illumina 16S and
810 ITS amplicon sequencing. bioRxiv doi:10.1101/081257:081257.
811 87. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello
812 EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST,

Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335-6.

88. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674-1676.

89. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-9.

90. Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. Trends Genet 16:276-277.

91. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14:417-419.

92. MacLean AA, MacPherson G, Aneja P, Finan TM. 2006. Characterization of the beta-ketoadipate pathway in Sinorhizobium meliloti. Appl Environ Microbiol 72:5403-5413.

93. Kaschabek SR, Kuhn B, Müller D, Schmidt E, Reineke W. 2002. Degradation of aromatics and chloroaromatics by *Pseudomonas* sp. strain B13: purification and characterization of 3-oxoadipate:succinyl-coenzyme A (CoA) transferase and 3-oxoadipyl-CoA thiolase. J Bacteriol 184:207.

834 **Table 1** New bacterial species from karst caves 1 and 2 and their etymology and accession numbers in the international culture

835 collections.

Taxonomy	Rank	Etymology	Type Designation	CGMCC/KCTC/ NBRC Accessions
Azospirillum cavernae	sp. nov.	ca.ver′nae. L. gen. n. *cavernae*, of a cave	K2W22B-5^T	CGMCC 1.13529 / NBRC 113558
Deinococcus cavernous	sp. nov.	ca.ver′nus. L. gen. masc. n. *cavernous*, of a cave	K2S05-167^T	CGMCC 1.13537 / KCTC 43236
Massilia cavernae	sp. nov.	ca.ver′nae. L. gen. n. *cavernae*, of a cave	K1S02-61^T	CGMCC 1.13526 / KCTC 82189
Nocardioides caviwatar	sp. nov.	*cavum*, L. hole; *watar*, Gk, water, *caviwatar*, from cave water	K1W22B-1^T	CGMCC 1.13535 / KCTC 49465
Noviherbaspirillum cavernae	sp. nov.	ca.ver′nae. L. gen. n. *cavernae*, of a cave	K2R10-39^T	CGMCC 1.13602
Noviherbaspirillum rocha	sp. nov.	ro'cha. ML. gen. n. *rocha*, from rock	K1R23-30^T	CGMCC 1.13534
Noviherbaspirillum sedimentum	sp. nov.	sedi'mentum. L. gen. pl. n. *sedimentum*, from sediment	K1S02-23^T	CGMCC 1.13533
Oleomonas cavernae	sp. nov.	ca.ver′nae. L. gen. n. *cavernae*, of a cave	K1W22B-8^T	CGMCC 1.13560 / KCTC 82188
Paenisporosarcina cavernae	sp. nov.	ca.ver′nae. L. gen. n. *cavernae*, of a cave	K2R23-3^T	CGMCC 1.13561 / NBRC 113453
Pseudomonas cavernecola	sp. nov.	ca.verne′co.la. L. n. *cavernae* cave; L. suff. *-cola*, dweller; N.L. n. *cavernecola* cave-dweller	K1S02-6^T	CGMCC 1.13525 / KCTC 82190
Pseudomonas cavernae	sp. nov.	ca.ver′nae. L. gen. n. *cavernae*, of a cave	K2W31S-8^T	CGMCC 1.13586 / KCTC 82191
Sphingomonas cavernae	sp. nov.	ca.ver′nae. L. gen. n. *cavernae*, of a cave	K2R01-6^T	CGMCC 1.13538 /

KCTC 82187

836 **Table 2** General features of the newly cultivated and novel bacterial genomes

Organism	GenBank accession number	Contigs	Size (Mb)	Genes	Proteins	G+C%	Completeness
Azospirillum cavernae strain K2W22B-5^T	GCA_003590795.1	9	6.461	5850	5595	66.0%	94.2%
Deinococcus cavernous strain K2S05-167^T	GCA_003590815.1	32	4.566	4571	4192	64.0%	77.7%
Massilia cavernae strain K1S02-61^T	GCA_003590855.1	201	5.439	5022	4473	63.6%	87.3%
Nocardioides caviwatar strain K1W22B-1^T	GCA_003600895.1	2	3.467	3334	3236	69.4%	94.3%
Noviherbaspirillum cavernae strain K2R10-39^T	GCA_003590875.1	4	4.665	4376	4207	59.9%	98.3%
Noviherbaspirillum rocha strain K1R23-30^T	GCA_003591035.1	3	6.495	5936	5725	57.5%	98.9%
Noviherbaspirillum sedimentum strain K1S02-23^T	GCA_003590835.1	4	5.038	4666	4484	59.4%	98.8%
Oleomonas cavernae strain K1W22B-8^T	GCA_003590945.1	29	5.643	5559	5077	66.7%	83.8%
Paenisporosarcina cavernae strain K2R23-3^T	GCA_003595195.1	1	2.537	2658	2507	39.8%	95.6%
Pseudomonas cavernecola strain K1S02-6^T	GCA_003596405.1	8	5.626	5241	4830	60.6%	98.7%
Pseudomonas cavernae strain K2W31S-8^T	GCA_003595175.1	1	4.950	4514	4308	64.5%	98.9%
Sphingomonas cavernae strain K2R01-6^T	GCA_003590775.1	5	4.244	4033	3878	63.9%	91.0%
Crenobacter cavernae strain K1W11S-77^T	GCA_003355495.1	1	3.271	3167	2980	65.3%	96.9%

Solimonas fluminis strain K1W22B-7	GCA_003428335.1	1	5.373	4807	4699	67.1%	92.0%

837 **Table 3** Bacterial strains, plasmids, and primers used in this study

Strain/Plasmid/Primer	Description	Source/Sequences
Strains		
Azospirillum K2W22B-5^T	4HB degrading strain	This study
Noviherbaspirillum K1R23-30^T	4HB degrading strain	This study
E. coli BL21 (DE3)	Protein expression host	TransGen
Plasmids		
pET-28a (+)	Gene expression vector	Novagen
pET-28a-k5pcaIJ	pET-28a (+) carrying *pcaI* and *pcaJ* of strain K2W22B-5^T	This study
pET-28a-30pcaIJ	pET-28a (+) carrying *pcaI* and *pcaJ* of strain K1R23-30^T	This study
Primers		
k5pcaIJ-F	For PCR of *pcaI* and *pcaJ* of K2W22B-5^T	GACGCATATGGCGCTCATCACACCC
k5pcaIJ-R	For PCR of *pcaI* and *pcaJ* of K2W22B-5^T	CCCAAGCTTACCCTCCGAACTGGTGCT
30pcaIJ-F	For PCR of *pcaI* and *pcaJ* of K1R23-30^T	GCGGCATATGATCAATAAAATTTGCACTTCC
30pcaIJ-R	For PCR of *pcaI* and *pcaJ* of K1R23-30^T	ATCCAAGCTTATTGGGGATATACGTCAGCG

838

839 **Figure legends**

840 **Figure 1** Workflow of the isolation procedure and the diversity of the cultured cave
841 bacteria. The boxplots in panels (b) and (c) show the Shannon indices of the cultivated
842 bacterial strains from the two caves and the three cave niches (rock, sediment, and water).
843 The pie charts in panels (b) and (c) show the taxonomy distribution of the cave isolates
844 from the two caves and the three cave niches. The PCoA plot in panel (b) shows the
845 β-diversity of the cultured cave bacteria based on the Bray-Curtis dissimilarity. The Venn
846 diagram in panel (b) shows the intersection of the cave isolates from the cave niches at
847 the species level.

848 **Figure 2** Taxonomic distribution of the cultured cave bacteria collection and its
849 representativeness in 16S rRNA gene amplicon datasets. Panel (a) shows the taxonomic
850 distributions at the phylum and genus levels. *Proteo.* - *Proteobacteria*; *Actino.* -
851 *Actinobacteria*; and *Firmi.* - *Firmicutes*; The boxplots in panel (b) show the percentage of
852 the sequences in the amplicon datasets that are represented by the cultured isolates, and
853 the triangles in each boxplot indicated the mean representativeness of each dataset.

854 **Figure 3** Morphologies and phylogenetic affiliations of the new species isolated from the
855 cave samples. The morphology in panel (a) is from transmission electron microscopy.
856 The phylogenetic tree in panel (b) was constructed based on the 16S rRNA genes using
857 the neighbor-joining algorithm.

858 **Figure 4** Metabolic overview of the newly isolated bacterial species from the caves.
859 Panel (a) shows the assimilation of the carbon sources according to the Biolog® GENIII

system; purple indicates positive and white indicates negative. Panel (b) shows the distributions of the COGs in the 14 newly sequenced genomes, the COGs are color-coded, with the highest number of genes shown in pink, and the genes with the lowest number shown in green.

Figure 5 Overview of the metabolisms of (a, c, e) the cave cultured genome collection and (b, d, f) the public cave metagenome data and their relationships to the C/N/S cycles. The numbers and percentages on the arrows in panels (a), (c), and (e) represent the number of species that are able to perform the conversion and their relative abundances; the width of the arrow is in proportion to the number of species that are able to perform the transformation. The color ranges in panels (b), (d), and (f) indicate the TPM value (transcripts per million) of each KO in the metagenome data (accession numbers are shown as x-axis labels).

Figure 6 (a) Representative genetic cluster and (b) β-ketoadipate pathway, and (c) the two 3-oxoadipate-CoA-transferase gene clusters and (d) their enzymatic activity in the pathway. The red percentages in panel (a) indicate the amino acid similarities between the cave isolates and strain ATCC35469. The controls in panel (d) summarize three conditions: the assay mixture without enzyme, or with K5PcaIJ but without succinyl-CoA, or with 30PcaIJ but without succinyl-CoA.

879 **Supplementary information**

880 **Table S1** Distribution of the novel taxa in the two caves and among the three substrates

881 **Table S2** Carbon and nitrogen metabolism related genes in strains K1W22B-8^T and

882 K2W22B-5^T

883 **Table S3** Blast results for the annotated PcaIJ in the metagenomic data versus the

884 cultured bacterial genomes

885 **Figure S1** Phylogenetic trees based on the 16S rRNA genes of the cave bacterial isolates.

886 **Figure S2** Organizations of the carbon and nitrogen cycling related genes in strains

887 K2W22B-5^T and K1W22B-8^T.

888 **Figure S3** Growth of strain K2W22B-5^T with 4HB as the sole carbon source.

889 **Dataset S1** 16S rRNA gene sequences and alignment results for the isolated cave bacteria

890 **Dataset S2** Statistics for the isolated cave bacteria species

891 **Dataset S3** Description of the new bacterial species from the karst caves

892 **Dataset S4** Accession and completeness of the cultured cave bacterial genomes

893 **Dataset S5** Annotation of the non-redundant cave gene catalog (using the KEGG

894 database)

895 **Dataset S6** Check list of the C/N/S metabolism related genes in the cave bacterial

896 genomes

897 **Dataset S7** Metadata of the culture-independent data used in this study

898 **Dataset S8** Accession numbers of the representative bacterial strains obtained in this

899 study