scientific data

Check for updates

OPEN

DATA DESCRIPTOR

The first large scale rDNA amplicon database of soil microbiomes from Pamir Plateau, China

Jing Zhu[1,5], Xiang Sun[2,5], Qi-Yong Tang[1], Mei-Ying Gu[1], Zhi-Dong Zhang[1]✉ & Jian-Wei Chen[3,4]✉

The Eastern Pamir, distinguished with high altitude, extremely arid and cold climate, limited nutrients and sparse vegetation, is a unique ecological reservoir. Microbial communities play a central role in maintaining Eastern Pamir's ecosystem functioning. Despite the ecological significance, due to the difficulty of sample collection and microbial isolation, the microbial diversity and its functionality at the Pamir Plateau have been rarely documented. To fill this gap, 80 soil samples from 17 sites across different elevations were collected, performed the rDNA amplicon sequencing to present the first large-scale overview of bacterial, archaeal, and fungal communities in the Eastern Pamir. Microbiome analysis revealed that the bacteria *Actinobacteria*, *Alphaproteobacteria* and *Bacteroidia*, alongside such as archaea *Nitrososphaeria* and *Halobacteria*, and fungi including *Dothideomycetes*, *Sordariomycetes* and *Eurotiomycetes* were dominant lineages at class level in soil microbial communities. The community structure and biodiversity of soil microorganisms provided by this dataset would be pivotal for future studies aimed at understanding the biogeographical distribution, ecological functions and environmental responses of microbial communities of the Pamir Plateau.

Background & Summary

The plateau ecosystems, characterized by their cold climates and extensive expanses of tundra, lakes, and glaciers, are integral to the terrestrial ecosystems. These high-latitude and high-elevation ecosystems are particularly susceptible to global climate change, as evidenced by Johnson *et al.*[1]. Microorganisms play a pivotal role in the biogeochemical cycling of carbon and nitrogen in such extreme environments, marked by such as high altitude, low temperatures, aridity, and low oxygen levels (only about 70% of that in the plains). Furthermore, microorganisms are essential in shaping and maintaining ecosystems under these harsh conditions[2].

The Pamir Plateau is the second largest plateau in the world, extending across southwestern Xinjiang in China, southeastern Tajikistan, and northeastern Afghanistan, covering an area of approximately 100,000 square kilometers with an average elevation of 4,500 meters above sea level[3,4]. The vast mountains, elevated terrain, and varying topography in the region foster an extraordinary arid and cold climate with severe temperature fluctuations[5,6]. The region also encompasses diverse natural habitats like lakes, water systems, and glaciers, which collectively shape a distinctive heterogeneous zone[7,8]. The region is known for its abundant and unique microbial resources, making it a significant constitution to the high-altitude bioresource repository and global genetic diversity and becoming a key zone for biodiversity conservation[9,10]. However, the region confronts challenges due to anthropogenic disturbances, natural disasters, and climate change, which have accelerated the loss and extinction of microbial species in fragile ecosystems[11]. Soil microbial diversity and community functionality are particularly vulnerable to these extreme environmental changes.

Soil microorganisms constitute one of the most diverse groups of organisms on Earth. They play a crucial role in soil biogeochemical processes, participate in almost all material transformations in the soil, and have a significant impact on ecosystems such as forests, grasslands, wetlands, and arable lands[12,13]. The extreme aridity, nutrient scarcity, diurnal temperature fluctuations, and high ultraviolet radiation in the Pamir Plateau present a challenging environment. These conditions disrupt soil productivity and ecosystem functioning[14,15]. To date, prior studies on

[1]Institute of Applied Microbiology, Xinjiang Academy of Agricultural Sciences/Xinjiang Laboratory of Special Environmental Microbiology, Urumqi, China. [2]School of Life Sciences, Hebei University, Baoding, China. [3]BGI Research, Qingdao, 266555, China. [4]Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, 2100, Copenhagen, Denmark. [5]These authors contributed equally: Jing Zhu, Xiang Sun. ✉e-mail: zhangzheedong@qq.com; chenjianwei@genomics.cn

Locations	Location Abbreviation	Latitude	Longitude	Altitude (m)	Moisture (%)	Annual Mean Air Temperature (°C)	Annual Precipitation (mm)	Annual Average Solar Radiation (kJ/m·day)
Oytagh Bridge 1st	OytBrdg1	39.02049	75.55944	1715	37.6	10.2	85.3	15995.26
Oytagh Glacier Park (sunny slope)	OytGlaSun	38.98163	75.52304	1772	37.7	9.7	84.4	15963.7
Oytagh Glacier Park (shady slope)	OytGlaShade	38.98283	75.48941	1845	27	9.1	88.2	15902.47
Akto County (sunny slope)	AktSun	38.84969	75.47602	2024	33.9	7.9	82.5	15820
Akto County (shady slope)	AktShade	38.8075	75.3985	2192	29.2	6.2	82	15756.35
Gez River site 2	Gez2	38.77064	75.1986	2743	36.5	2.1	94	15110.25
Gez River site 1	Gez1	38.74822	75.14589	3025	31.1	0.3	84.5	15219.63
Bulungkol Lake	BulLake	38.73108	75.01833	3311	32.6	1.0	79.8	15287.22
Subashi Bridge 8th	SbsBrdg8	38.07994	75.00631	3488	33.5	1.6	65.5	15248.35
Subashi Bridge 7th	SbsBrdg7	38.14237	74.97312	3493	20.3	0.3	67.75	15184.54
Kongur Tiube	KongT	38.4779	75.04658	3635	35.3	0.2	83	14910.54
Sate Baile Dikule Lake	SBDLake	38.47334	75.04553	3648	21.6	0.2	82.5	14910.58
Muztagata Camp	Muztagata	38.35576	74.96696	3743	35.4	−0.7	80.8	14968.42
Kulma Pass	Kulma	38.19964	74.92149	3792	31.4	−0.6	74.7	15061.81
Subashi Pass (shady slope)	SbsShade	38.28943	74.93091	3828	22.4	−1.3	76.6	15026.77
Subashi Pass (sunny slope)	SbsSun	38.26913	74.9166	4063	31.2	−2.2	86	14898.46
Ulugjewat Pass	Ulug	38.27456	74.91571	4069	36.8	−2.1	84.3	14923.88

Table 1. Geographic location of sampling sites and environmental conditions.

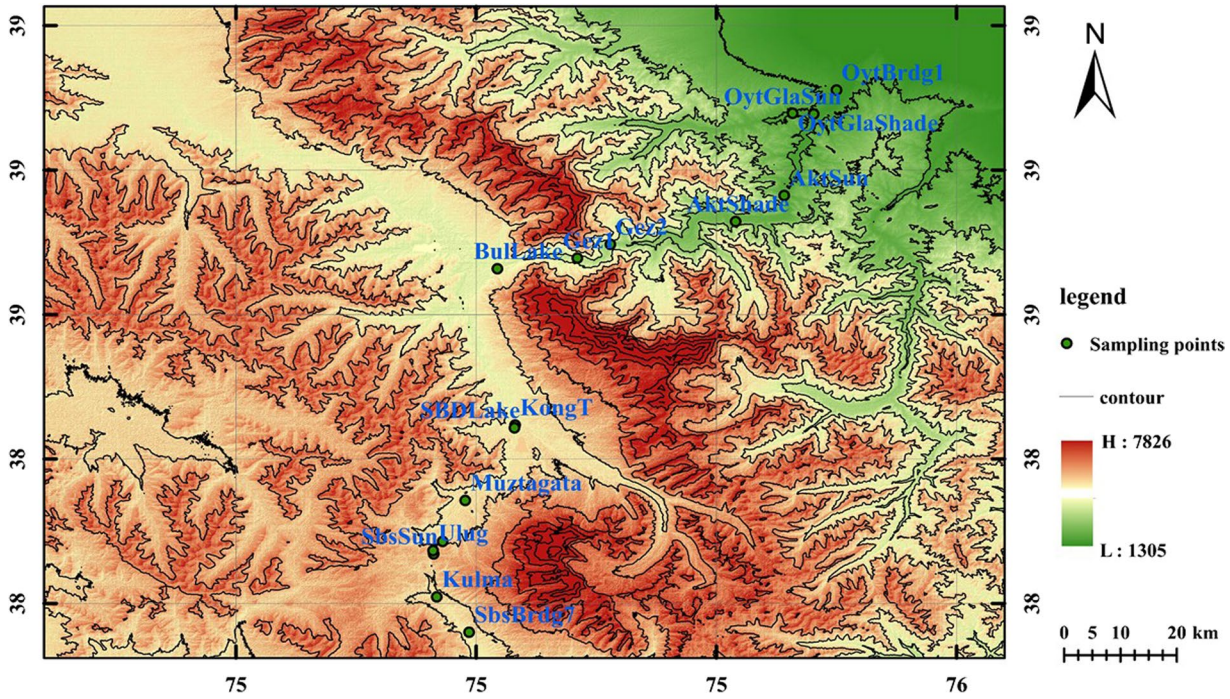

Fig. 1 Contour map of sampling site distribution in the Eastern Pamir.

microorganisms in plateau regions have focused on the Tibetan, Loess, and Mongolian Plateaus. These studies have involved variations in soil enzyme activities and microbial communities along elevation gradients[16], reflections of functional microorganisms in soil to natural factors such as nitrogen deposition and precipitation[17,18], or different responses of soil microorganisms from different plant communities to changes in environmental stresses[19].

There were sporadic reports on the microorganisms of the Pamir Plateau. As early as the 1960s, researchers studied soil actinomycetes and their antagonistic properties on the Pamir Plateau[20,21]. Aksenov *et al.*[22] studied the adaptive mechanisms of *Cryptococcus* in the Pamir region under very low humidity conditions. In the 1970s, Szymon *et al.*[23] isolated 16 species of fungi from 12 species of endemic plants of the Pamir Aly Mountains, which were found to be predominantly cotyledonous mycorrhizal fungi. Nataliia *et al.*[24,25] carried out studies on microbial diversity and colonization strategies in rock of cold desert ecosystems in the Eastern Pamir Mountains of Tajikistan, and Bu *et al.*[26] isolated numerous cold-adapted bacteria as well as archaeal resources from the Pamir Plateau. However, a systematic investigation of the microbial resources of the Pamir Plateau, comprehensive

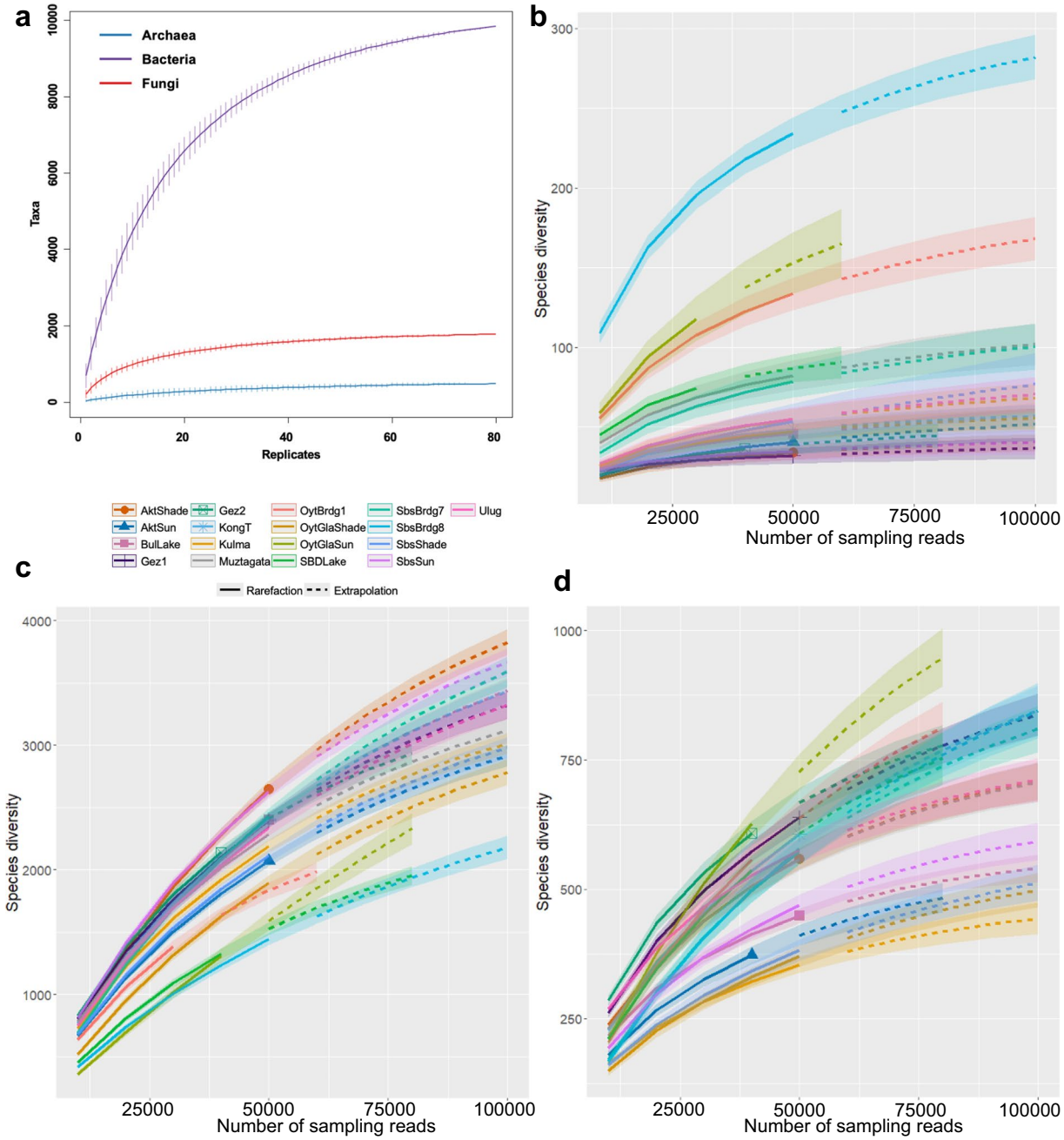

Fig. 2 Microbial diversity of the soil microbiota profile in Eastern Pamir. (**a**) Rarefaction curves of the bacterial (purple), archaeal (blue) and fungal (red) communities. (**b**–**d**) The richness of the archaeal (**b**), bacterial (**c**), and fungal (**d**) diversity index for each sampling site in Pamir, respectively.

data on the microbial resources of the Pamir region, including the distribution, composition, and function of the microbial communities is still an uncharted area. Therefore, research into soil microorganisms and the data collection on microbial resources in the Pamir Plateau are of great importance for understanding regional and global climatic and environmental changes, biodiversity, and carbon cycling.

In this study, we collected 85 soil samples from 17 sites along the G314 highway, stretching approximately 300 kilometers from Kashgar City to Tashkurgan County, with altitudes ranging from 1715 meters to 4069 meters (Table 1 and Fig. 1). Since five of the soil samples were unsuccessful in DNA extraction, resulting in a final dataset of 80 soil samples. Amplicon sequencing using Illumina NovaSeq sequencing platform yielded 7,644,450, 6,821,012, and 6,627,693 amplicon reads for soil archaeal, bacterial, and fungal datasets, respectively. Subsequent analysis yielded a count of 480 taxa in the archaeal dataset, 9,829 taxa in the bacterial dataset, and 1,778 taxa in the fungal dataset across all soil samples (Fig. 2 and Table 2). In Fig. 2, the rarefaction curves of all samples exhibit a plateauing trend, suggesting that the current sampling strategy is adequate to capture all the microbial taxa in soil communities. However, it is noteworthy that the species richness levels for archaea and fungi in this study were relatively low.

No.	Location Abbreviation	Bacteria			Fungi			Archaea		
		Raw Data	Clean Data	Taxa	Raw Data	Clean Data	Taxa	Raw Data	Clean Data	Taxa
1.1	OytBrdg1	81127	37046	624	106181	70228	243	71994	29606	52
1.2		NA	NA	NA	97441	71222	187	102508	72556	59
1.3		96763	38291	684	101993	78320	201	102342	69549	29
1.4		NA	NA	NA	NA	NA	NA	113473	44395	109
1.5		93155	42075	608	96983	69924	188	109918	74512	29
2.1	OytGlaSun	96041	43972	583	105809	77964	177	NA	NA	NA
2.2		66457	45946	155	107575	70399	146	104308	37907	87
2.3		56148	31365	116	102689	73311	220	105817	81798	34
2.4		NA	NA	NA	NA	NA	NA	100273	49003	55
2.5		82945	29755	592	98953	38008	300	NA	NA	NA
3.1	AktSun	98450	39646	653	NA	NA	NA	68061	35078	18
3.2		93462	34872	619	99320	72083	202	96392	46233	23
3.3		86179	40949	692	102975	42904	189	83731	57515	19
3.4		82393	36739	677	101004	66329	207	60065	38912	20
3.5		80986	35024	698	100784	62676	124	85668	39382	17
4.1	AktShade	86603	38797	685	109611	57296	202	80458	32611	14
4.2		99861	42097	848	96145	32663	208	106091	41123	17
4.3		93110	39756	771	105440	40561	242	98180	61756	18
4.4		85457	37454	776	101892	64562	230	102724	42016	21
4.5		92872	38900	695	108790	56616	307	93930	59586	19
5.1	Gez1	80538	35217	757	87137	64059	293	68965	18251	24
5.2		92557	38848	804	104914	71405	251	100995	31464	28
5.3		93009	39541	794	99152	55096	191	103662	29404	19
5.4		95288	41661	834	73897	58164	207	106732	25386	22
5.5		89905	39038	822	103324	67032	361	99264	32307	19
6.1		95153	39902	820	101409	61367	324	102757	33989	13
6.2	Gez2	98946	40289	828	59501	40570	282	97811	27883	18
6.3		92736	41161	791	100780	66384	276	108241	37623	27
6.4		92707	40660	890	101765	77743	260	96429	21677	22
6.5		NA	NA	NA	NA	NA	NA	NA	NA	NA
7.1	BulLake	82784	39317	707	103654	84889	177	104962	52355	16
7.2		93309	37702	739	100586	73873	192	99953	34696	25
7.3		93138	42157	767	108176	78538	237	99110	35857	28
7.4		83447	38237	771	97369	65186	220	57988	19234	19
7.5		92861	39680	660	107833	75591	263	100077	42338	35
8.1	KongT	94350	42409	816	99669	59198	247	94400	28866	32
8.2		97770	42533	731	103514	77261	179	107780	40015	28
8.3		97851	44258	855	93623	61889	211	105292	54314	20
8.4		84483	34393	650	108747	81320	192	109687	61840	19
8.5		95077	42218	786	106769	83772	308	106241	65936	28
9.1	Ulug	91544	37740	712	107448	57594	267	107204	24581	20
9.2		99665	43802	825	96619	54878	247	100230	40095	21
9.3		83700	32439	702	94006	39803	260	104717	32030	33
9.4		87298	38941	759	102264	71696	299	103261	44939	31
9.5		86640	39360	760	105541	67208	269	105980	37566	30
10.1	SbsBrdg8	98940	42530	469	101126	36841	102	76088	28679	85
10.2		87046	28168	385	96806	68489	165	111395	18499	97
10.3		87918	32005	307	59710	26443	165	101113	20049	130
10.4		90073	37236	514	105560	75376	288	108576	30421	125
10.5		87358	37206	405	96344	33724	116	99615	31571	109
11.1	SbsBrdg7	96804	43791	781	88225	56071	252	82555	52902	17
11.2		81827	39433	742	78325	51099	120	106978	57662	32
11.3		86957	39381	746	90041	65233	121	106383	62481	37
11.4		99407	45173	730	96509	8598	93	101142	48389	48
11.5		84119	40198	467	108047	69131	269	100687	41758	35

Continued

No.	Location Abbreviation	Bacteria			Fungi			Archaea		
		Raw Data	Clean Data	Taxa	Raw Data	Clean Data	Taxa	Raw Data	Clean Data	Taxa
12.1	Kulma	89756	39844	737	91067	75556	166	100238	70058	23
12.2		95343	42014	697	98032	75644	150	106011	34652	37
12.3		90373	40396	732	106359	83818	169	102570	62457	23
12.4		92810	38061	738	107623	85373	174	102845	68556	22
12.5		98819	42569	730	110523	90111	156	100170	58957	22
13.1	SbsSun	82955	36927	782	73772	52785	170	105102	67569	26
13.2		90166	42547	898	70773	23683	206	89428	47278	22
13.3		92805	40953	802	72543	29191	211	109099	55110	27
13.4		80138	36949	821	108335	59513	215	101583	35751	22
13.5		94503	40283	756	72819	60415	160	102685	38785	20
14.1	SbsShade	84621	37152	665	86146	70877	140	95770	57832	21
14.2		94208	39175	733	82297	38227	199	96050	45042	17
14.3		87595	36151	713	89026	69658	157	98091	55849	15
14.4		89134	38332	659	76132	52794	184	103492	38897	15
14.5		88583	39214	624	66874	49040	121	106644	48336	42
15.1	Muztagata	99451	40691	789	81664	51021	232	96985	46916	55
15.2		93884	39904	770	67192	41985	188	107963	49119	23
15.3		93292	39845	831	54698	40413	224	104981	53142	34
15.4		83925	36503	780	102541	62155	270	78994	32384	35
15.5		81575	35545	727	95595	66562	237	98689	68570	53
16.1	SBDLake	90060	35091	571	92143	69039	216	93060	46678	67
16.2		91537	61700	58	NA	NA	NA	NA	NA	NA
16.3		NA	NA	NA	99829	69259	316	NA	NA	NA
16.4		93617	41625	634	96269	66037	207	109626	61580	17
16.5		83779	31905	550	90873	65003	102	98892	53283	51
17.1	OytGlaShade	78399	56105	205	101311	78789	147	104636	75109	21
17.2		92113	44251	705	104515	34438	115	100544	70146	18
17.3		89096	43143	787	105164	77054	143	105325	76928	23
17.4		53678	45397	63	105793	31322	142	102962	77373	28
17.5		92456	41945	838	108489	49105	195	106569	44224	25

Table 2. Sequence data statistics of each sample. Note: NA indicates that soil sample DNA could not be extracted and sequencing results could not be obtained.

In addition, 224 taxa could not be taxonomically classified within the fungal community at the phylum level, representing 12.60% of the overall detected fungal taxa with the cumulative relative abundance. Similarly, 422 bacterial taxa could not be classified at the phylum level, comprising 4.29% of the total detected bacterial taxa. Moreover, all archaeal taxa were annotated into 14 families, but 191 archaeal taxa remained unclassified at the genus level, accounting for a substantial 39.79% of the total detected archaeal taxa. This suggests that there may be a considerable volume of unrecognized microbial resources in this region waiting for further exploration.

The analysis of the bacterial communities revealed that the *Actinobacteria* (15.61%) and *Alphaproteobacteria* (11.99%) emerged as the dominant groups, alongside notable prevalence of *Bacteroidia* (8.72%) and *Gammaproteobacteria* (7.01%) (Fig. 3a). Fungal communities were dominantly represented by *Dothideomycetes* (32.06%), *Sordariomycetes* (23.66%), and *Eurotiomycetes* (9.35%) (Fig. 3b). In archaeal communities, *Nitrososphaeria* class was predominant, representing 70.46% of the relative abundance, followed by *Halobacteria* at 27.90% (Fig. 4a). Further analysis of the composition of archaeal community at the genus level showed that *Candidatus Nitrososphaera* (26.48%) was the predominant genus, followed by *Candidatus Nitrocosmicus* (25.9%), *Nitrososphaeraceae* (18.08%), *Haloferacaceae* (8.58%), and *Halalkalicoccus* (3.74%) (Fig. 4b). Notably, the highest species richness for archaea is observed at Subashi Bridge 8th (SbsBrdg8), whereas for bacteria and fungi, it was at Gez River site 2 (Gez2) (Table 3). The results of the amplicon data provided insights into the composition of the microbial community and its spatial distribution patterns in the region.

Multiple Regression on Distance Matrices (MRM) analysis explored the influence of environmental factors on the composition of the three types of microbial communities (Table 4). It also reveals that site geographical distances significantly influenced the compositions of all three microbial communities. Archaeal communities are best explained by the environmental factors ($R^2 = 0.4121$), followed by fungal ($R^2 = 0.2517$) and bacterial communities ($R^2 = 0.1446$). Soluble salt emerges as the most influential factor for archaeal communities, followed by total nitrogen, pH, precipitation seasonality, and available nitrogen. For bacterial communities, precipitation seasonality and soil moisture were significantly influential, while total nitrogen, moisture, total organic matter, soluble salt, and available nitrogen were key factors shaping fungal communities. Annual average solar radiation, annual mean temperature, and temperature seasonality strongly correlate with altitude, suggesting elevation is a key determinant of temperature and solar radiation in this region (Fig. 5).

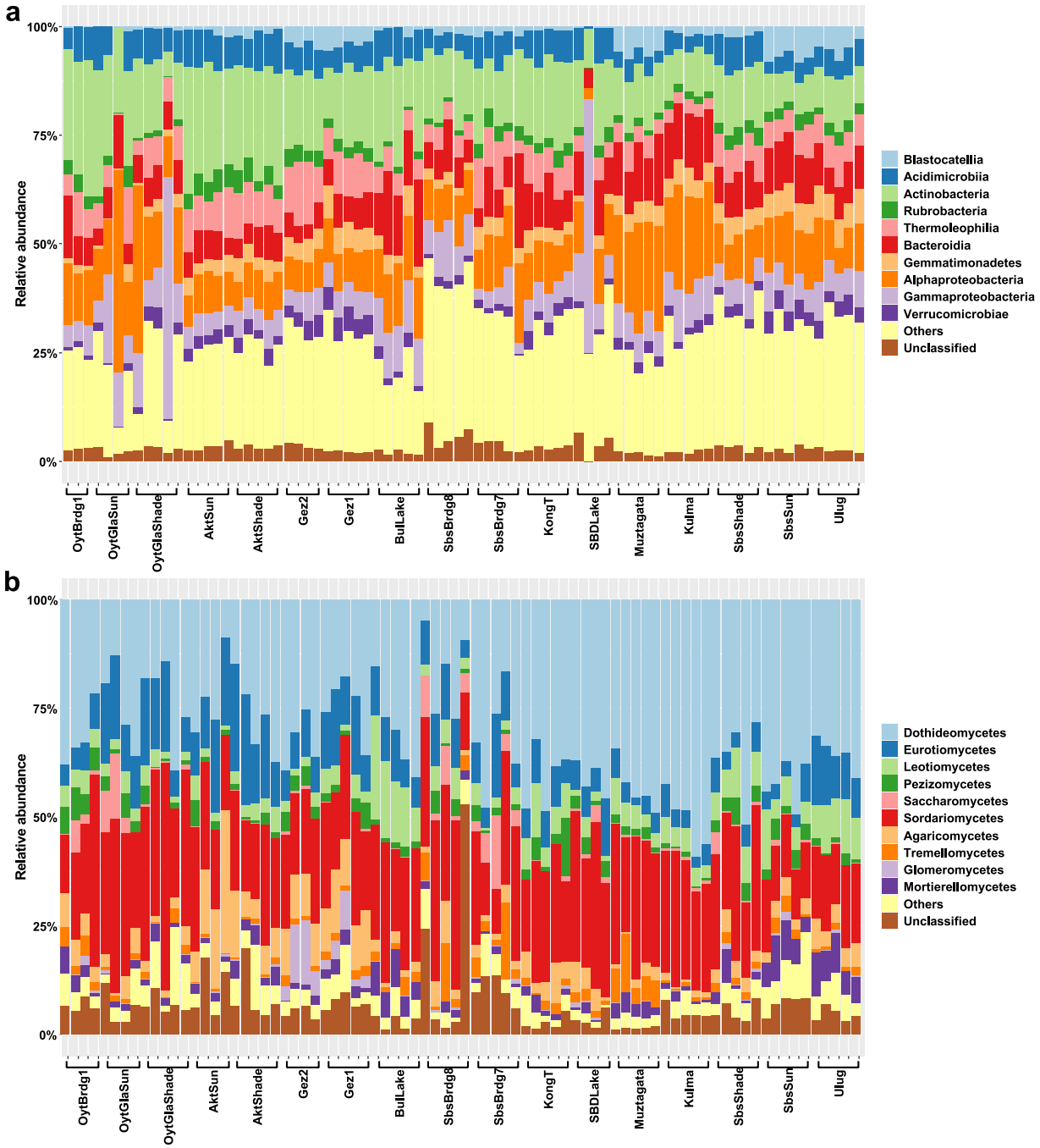

Fig. 3 Soil microbial community structures in Pamir. The bar plots show the taxonomic distribution of the bacterial community (**a**) and fungal community (**b**) at the class level.

Methods

Sampling. In July 2020, 85 soil samples were collected from 17 sites along the eastern slope of the Eastern Pamir Plateau in Xinjiang, China (Fig. 1 and Table 1). At each site, five soil samples were taken as replicates. These replicates were collected from 5 m × 5 m plots using a five-point sampling strategy. The 2 cm of topsoil was removed to discard litters, and approximately 1 kg of soil was collected from depths of 2–20 cm and sieved to remove rocks and debris. The soil samples were immediately transported to the laboratory in an ice box. For subsequent DNA extraction, 50 g of soil from each sample was stored at −80 °C, and the remainder was stored under cool, dry conditions for geochemical analysis. The contour map of sampling sites was drawn using ArcGIS mapping software (Arc Geographic Information System, Environmental Systems Research Institute, Inc. USA) and downloading 30 m resolution DEM data from the Geospatial Data Cloud website (https://www.gscloud.cn/), contour data is generated through the splicing, cropping, and contour tools in the ArcGIS toolbox. Smooth the generated contour data and add a grid, legend, compass, scale bar, and drawing name.

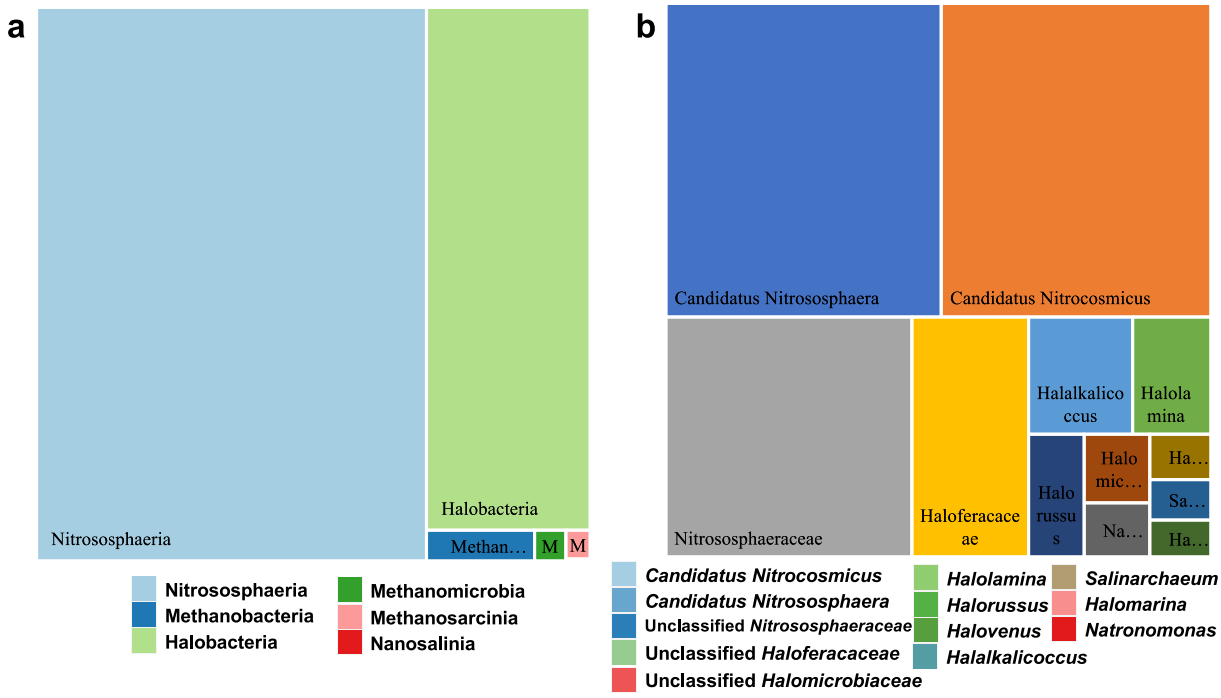

Fig. 4 Average relative abundance of the predominant archaea at the class level (**a**) and genus level (**b**) in Pamir.

DNA extraction and PCR amplification. Total genomic DNA was extracted from 5 g of each soil sample using CTAB/SDS method. DNA concentration and quality was assessed on 1% agarose gels. Due to the failure to extract DNA from 5 soil samples, data from 80 soil samples were finally obtained. DNA was then diluted to 1 ng/μL using sterile deionized water. The specific amplification process employed barcode-attached universal primers targeting the 16S rRNA V4-V5 region for archaea (Arch519F/Arch915R)[27], the 16S rRNA V4 region for bacteria (515 F/806 R)[28], and the ITS1 region for fungi (ITS1-1F-F/ITS1-1F-R)[29]. All PCR reactions were carried out with Phusion® High-Fidelity PCR Master Mix (New England Biolabs). All PCR reactions were performed using 15 μL systems, composing of Phusion® High-Fidelity PCR Master Mix (New England Biolabs), 0.2 μM of both forward and reverse primers, and about 10 ng DNA templates. The thermal cycling consisted of an initial denaturation at 98 °C for 1 min, followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 50 °C for 30 s, and elongation at 72 °C for 30 s, with a final elongation 72 °C for 5 min. PCR products were mixed with 1X loading buffer (contained SYB green) and subjected to electrophoresis on a 2% agarose gel for quality assessment and purified with GeneJETTM Gel Extraction Kit (Thermo Scientific).

Library preparation and sequencing. The sequencing libraries were generated using TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina, USA) following the manufacturer's recommendations. The library quality was assessed on the Qubit@ 2.0 Fluorometer (Thermo Scientific) and Agilent Bioanalyzer 2100 system. The qualified libraries were mixed in equal amounts and then sequenced on an Illumina NovaSeq platform with a 250 bp paired-end model. The DNA extraction, Amplicon library preparation, and sequencing services were provided by Novogene Co., Ltd. (Beijing, China).

Bioinformatics analysis. The forward and reverse amplicon reads acquired through the Illumina NovaSeq sequencer were demultiplexed utilizing the QIIME2 plugins (version 2018.11). Primer trimming was performed with Cutadapt tool[30]. The subsequent quality filtering, denoising, paired-end merging, and dereplication were conducted via the DADA2 workflow[31]. Chimeric sequences were identified and removed using the *removeBimeradenovo* function (Fig. 6).

Taxonomic assignment for archaeal and bacterial ASVs was conducted using the Naïve Bayes approach (with a minimum of 75 bootstrap cutoff value) following the DADA2 workflow[31], in reference to the SILVA training set (version 138.1)[32]. Fungal ASVs were classified against the UNITE Fungi database (version 10.05.2021_dev)[33]. The taxonomically annotated ASVs were subsequently agglomerated at the species level using the tax_glom function in the "phyloseq" package[31,34]. ASVs not assignable at the species level were clustered into operational taxonomic units (OTUs) based on 97% similarity using the OTU function in the "kmer" package[35]. Representative sequences for these OTUs were classified with the Naïve Bayes approach. In each microbial dataset, singletons, doubletons, and samples with fewer than 1,000 sequences were excluded. Taxonomic assignment reads among the samples were Hellinger transformed using the decostand function in the R "vegan" package[36], and these transformed values were subsequently considered as abundance measures for statistical analyses. Figure 6 depicts the overall course of the production of all datasets.

	Archaea				Bacteria				Fungi			
Group	Shannon	Richness	Fisher's α	Simpson	Shannon	Richness	Fisher's α	Simpson	Shannon	Richness	Fisher's α	Simpson
OytBrdg1	3.47 ± 0.57b	55.60 ± 32.75bc	8.93 ± 5.67b	0.95 ± 0.02ab	6.32 ± 0.07a	638.67 ± 40.07abcd	123.06 ± 8.11abcd	1.00 ± 0.00a	4.98 ± 0.20ab	204.75 ± 26.29abc	36.52 ± 5.19abc	0.99 ± 0.01a
OytGlaSun	3.45 ± 0.75b	58.67 ± 26.69b	9.44 ± 4.46b	0.93 ± 0.06ab	5.34 ± 1.04a	361.50 ± 261.47e	68.20 ± 51.06e	0.99 ± 0.01a	4.92 ± 0.38ab	210.75 ± 66.79abc	38.68 ± 14.16abc	0.99 ± 0.01a
OytGlaShade	2.50 ± 0.29c	23.00 ± 3.81d	3.40 ± 0.57c	0.88 ± 0.04bc	5.45 ± 1.66a	519.60 ± 358.72bcde	101.18 ± 70.25bcde	0.96 ± 0.07a	4.44 ± 0.19b	148.40 ± 28.98c	26.69 ± 5.90c	0.98 ± 0.01a
AktSun	2.15 ± 0.12c	19.40 ± 2.30d	2.93 ± 0.43c	0.80 ± 0.03c	6.37 ± 0.05a	667.80 ± 32.34abc	128.94 ± 7.06abc	1.00 ± 0.00a	4.56 ± 0.40ab	180.50 ± 38.42abc	33.49 ± 7.22abc	0.98 ± 0.01a
AktShade	2.25 ± 0.19c	17.80 ± 2.59d	2.57 ± 0.40c	0.85 ± 0.02bc	6.50 ± 0.10a	755.00 ± 66.79ab	147.57 ± 14.12ab	1.00 ± 0.00a	4.92 ± 0.30ab	237.80 ± 41.94abc	45.33 ± 7.27abc	0.98 ± 0.01a
Gez2	2.50 ± 0.22c	20.00 ± 5.94d	2.91 ± 0.98c	0.88 ± 0.02bc	6.62 ± 0.05a	832.25 ± 41.65a	163.88 ± 9.02a	1.00 ± 0.00a	5.22 ± 0.15a	285.50 ± 27.29a	53.84 ± 5.52a	0.99 ± 0.01a
Gez1	2.72 ± 0.12c	22.40 ± 3.78d	3.21 ± 0.61c	0.92 ± 0.01ab	6.58 ± 0.04a	802.20 ± 29.65ab	157.32 ± 6.40ab	1.00 ± 0.00a	5.09 ± 0.34ab	260.60 ± 68.81abc	48.86 ± 12.92ab	0.99 ± 0.01a
BulLake	2.51 ± 0.29c	24.60 ± 7.50d	3.73 ± 1.30c	0.87 ± 0.04bc	6.45 ± 0.09a	728.80 ± 46.24abc	142.39 ± 9.07ab	1.00 ± 0.00a	4.84 ± 0.20ab	217.80 ± 34.45abc	41.19 ± 6.52abc	0.98 ± 0.01a
SbsBrdg8	4.40 ± 0.26a	109.20 ± 18.82a	18.12 ± 3.19a	0.98 ± 0.01a	5.79 ± 0.22a	416.00 ± 79.65de	78.40 ± 16.27de	1.00 ± 0.00a	4.65 ± 0.54ab	167.20 ± 73.27bc	29.63 ± 13.56bc	0.98 ± 0.01a
SbsBrdg7	2.65 ± 0.32c	33.80 ± 11.17 cd	5.49 ± 2.08bc	0.86 ± 0.03bc	6.38 ± 0.25a	693.20 ± 127.87abc	135.06 ± 26.42abc	1.00 ± 0.00a	4.56 ± 0.56ab	171.00 ± 82.69bc	31.22 ± 16.20bc	0.98 ± 0.01a
KongT	2.33 ± 0.45c	25.40 ± 5.64d	3.94 ± 0.87c	0.80 ± 0.10c	6.52 ± 0.12a	767.60 ± 79.83ab	150.19 ± 17.11ab	1.00 ± 0.00a	4.85 ± 0.27ab	227.40 ± 51.83abc	43.68 ± 11.29abc	0.98 ± 0.00a
SBDLake	2.81 ± 0.96c	45.00 ± 25.53bcd	7.53 ± 4.61bc	0.85 ± 0.14bc	5.51 ± 1.39a	453.25 ± 265.91cde	86.55 ± 51.90cde	0.98 ± 0.03a	4.74 ± 0.66ab	210.25 ± 87.45abc	39.17 ± 16.22abc	0.98 ± 0.02a
Muztagata	2.91 ± 0.32bc	40.00 ± 13.64bcd	6.43 ± 2.50bc	0.91 ± 0.03ab	6.56 ± 0.05a	779.40 ± 37.38ab	151.99 ± 8.17ab	1.00 ± 0.00a	4.96 ± 0.15ab	230.20 ± 29.38abc	42.85 ± 6.16abc	0.99 ± 0.00a
Kulma	2.49 ± 0.32c	25.40 ± 6.50d	3.84 ± 1.03c	0.87 ± 0.05bc	6.47 ± 0.04a	726.80 ± 16.99abc	141.45 ± 3.00ab	1.00 ± 0.00a	4.44 ± 0.06b	163.00 ± 9.80bc	30.17 ± 2.12bc	0.98 ± 0.00a
SbsShade	2.41 ± 0.36c	22.00 ± 11.45d	3.27 ± 1.93c	0.87 ± 0.04bc	6.39 ± 0.07a	678.80 ± 43.85abc	131.21 ± 9.27abc	1.00 ± 0.00a	4.59 ± 0.18ab	160.20 ± 31.73bc	28.78 ± 6.83bc	0.98 ± 0.01a
SbsSun	2.71 ± 0.15c	23.40 ± 2.97d	3.39 ± 0.94c	0.91 ± 0.01ab	6.59 ± 0.07a	811.80 ± 53.88	159.53 ± 11.77ab	1.00 ± 0.00a	4.77 ± 0.17ab	192.40 ± 25.46abc	35.28 ± 4.82abc	0.98 ± 0.01a
Ulug	2.78 ± 0.15c	27.00 ± 6.04	4.02 ± 1.02c	0.91 ± 0.01ab	6.51 ± 0.07a	751.60 ± 48.84ab	146.27 ± 10.44ab	1.00 ± 0.00a	5.19 ± 0.08ab	268.40 ± 19.15ab	49.90 ± 3.97ab	0.99 ± 0.00a

Table 3. Soil microbial alpha diversity indices at sampling sites on the Pamir Plateau. The above data represent the mean ± standard deviation ($n \geq 3$). Different letters (a, b, c, d) indicate significant differences between the data (Duncan test, $p < 0.05$).

	Archaea	*p*	Bacteria	*p*	Fungi	*p*
Distances among sites	−0.1719	**	−0.2672	***	−0.1311	***
Moisture	−0.0053	ns	−0.1217	*	−0.1043	***
Total organic matter	−0.0280	ns	0.0828	ns	0.103	*
Total nitrogen	−0.1973	**	−0.1969	ns	−0.122	**
Available nitrogen	0.0580	*	0.0566	ns	0.0506	**
Soluble salt	−0.4375	***	−0.1004	ns	−0.0528	*
pH	0.0765	*	0.0006	ns	−0.0089	ns
Precipitation seasonality	0.0673	*	0.1508	**	0.0179	ns

Table 4. The key environmental factors influential to microbial communities revealed with MRM. The star symbol indicated significant correlations between the data. ns represented $p > 0.05$, *represented $0.01 < p \leq 0.05$, **represented $0.001 < p \leq 0.01$, ***represented $p \leq 0.001$.

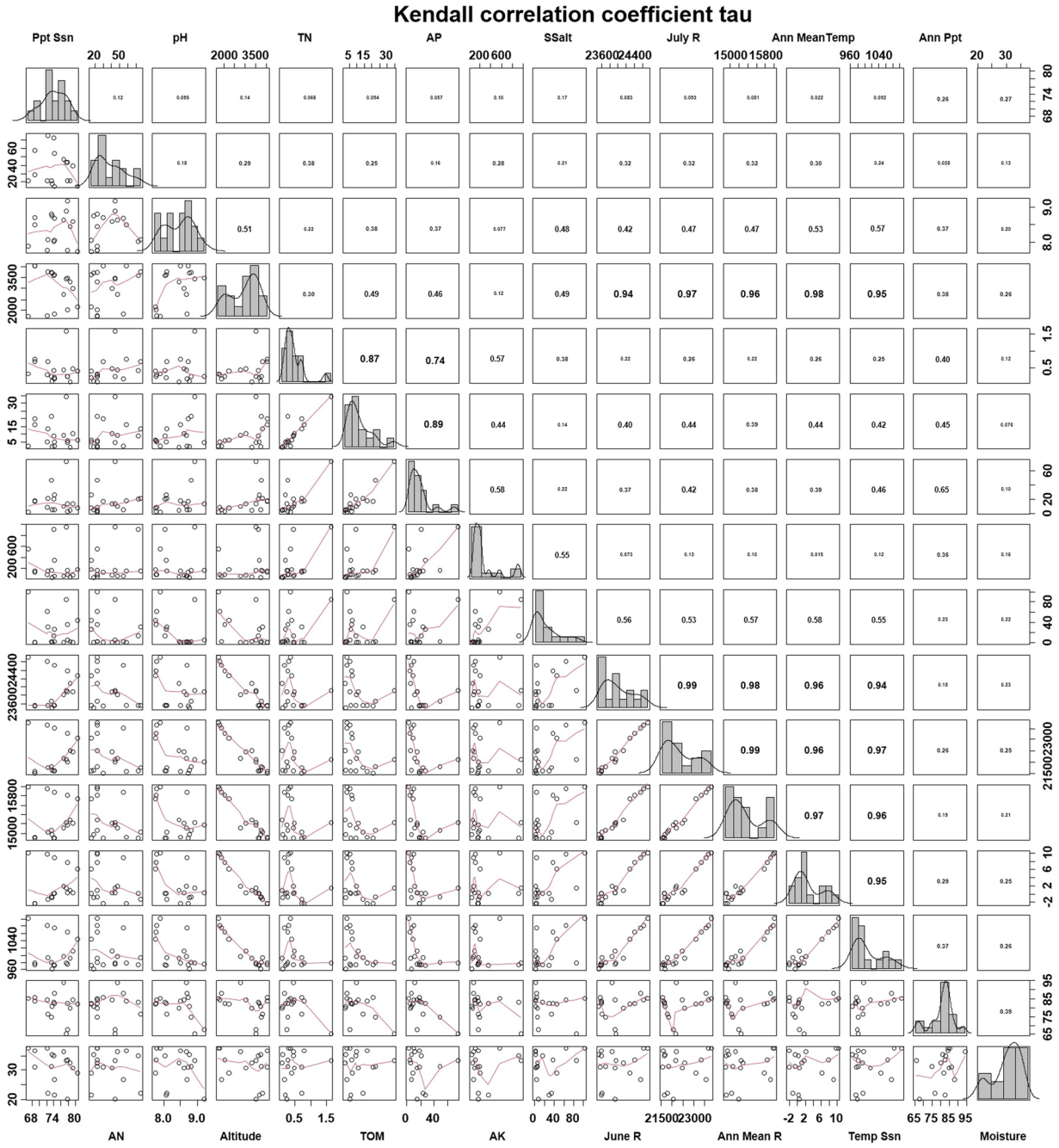

Fig. 5 Kendall correlation indicated self-correlated environmental factors in the present study.

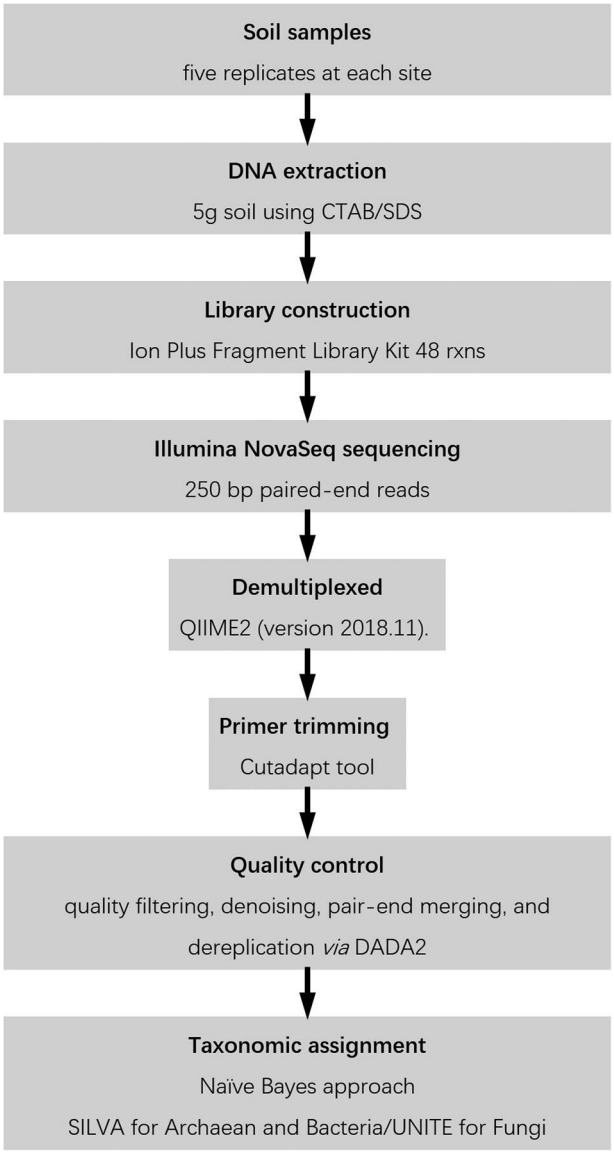

Fig. 6 The Schematic overview of sampling and details of the workflow and tools used for data processing and analysis performed in this study.

Environmental factors analysis. A standard soil test series (NY/T 1121) was conducted. Organic matter (OM) was determined using the $K_2Cr_2O_7$ oxidation method. The total nitrogen (TN) was measured using the Kjeldahl method. Available nitrogen (AN) was determined using the sulfate extraction method. Available phosphorus (AP) was detected using the hydrochloric acid–ammonium fluoride extraction–molybdenum antimony colorimetric method. Available potassium (AK) was detected using the ammonium acetate extraction–flame photometric method. Soluble salt (Salt) was detected using the mass method. The pH was determined using a potentiometric method[37]. Kendall's τ statistic was employed to estimate a rank-based measure of associations between environmental factors, utilizing pairs function within the "graphics" package and panel-related functions in the "MESS" package[38]. Multiple Regression on Distance Matrices (MRM) analysis was conducted to test the principal environmental factors shaping microbial communities, using the *MRM* function in the "ecodist" package[39]. MRM analysis employed the 1-distance$_{\text{Bray-Curtis}}$ similarity measure to represent the microbial community composition.

Data Records

The processed data along with metadata have been deposited in the Sequence Read Archive (SRA) database of the National Center for Biotechnology Information (NCBI) under the BioProject IDs PRJNA1032247 (https://identifiers.org/ncbi/insdc.sra:SRP468803)[40]. The ASV tables, fasta sequences, and taxonomy data for archaea, bacteria, and fungi were respectively uploaded to Figshare repository, resulting in distinct links for each dataset (https://doi.org/10.6084/m9.figshare.26087533[41]; https://doi.org/10.6084/m9.figshare.26087554[42]; https://doi.org/10.6084/m9.figshare.26087566[43]).

Technical Validation

Sampling procedure. Strict aseptic procedures were implemented during soil sample collection to prevent contamination from the human body or between samples. The shoves and sieves were sprayed and wiped with 75% ethanol before and after every sampling, and latex gloves worn by collectors were changed for every sampling. Plastic bags and containers were newly opened or sterilized with ethanol before taking soil samples.

Qualification strategy. Eighty soil samples for each microbial dataset were successfully sequenced, yielding 98,765.09, 89,198.56, and 95,629.65 raw reads per sample in average in archaeal, bacterial, and fungal communities, respectively. The observed and estimated error rates were evaluated after error learning, to determine the optimal quality control strategy. In our research, sequencing quality thresholds for archaeal and bacterial raw reads were expected errors lower than 2 for both pair ends, trimming forward reads to 180 bp and reverse reads to 160 bp, truncating at the first instance of a quality score less than or equal to 2, and removing the reads of final lengths shorter than 100 bp; quality thresholds for fungal raw reads were expected errors at 2 and lengths trimmed to 210 bp for both pair ends, with the same quality truncating and final lengths criteria. The quality filtration, error denoising, pair-ends merging, and chimera removal yielded 56,988.48, 65,828.96, and 71,913.94 clean reads per sample on average in archaeal, bacterial, and fungal communities, respectively, which were subjected to subsequent analysis.

Taxonomy annotation. In order to take into account both ASV-based high taxonomy resolution and diversity evaluation at species level, this study employed a two-step approach for taxonomic annotation. ASVs assigned with species names were agglomerated and thus were not analogous to remaining ASVs at the taxonomic hierarchy. Consequently, the remaining ASVs that were unable to be annotated at species level, were clustered into OTUs based on 97% similarity, and annotated at genus level or above. The microbial diversity and community composition were assessed at species level with species agglomerated from ASVs and OTUs clustered from ASVs. The threshold of bootstrap value was set to 75 rather than the default 50.

Code availability

R codes applied in the present study are available at: https://github.com/xnus/PamirSoilMicrobes.

Received: 30 May 2024; Accepted: 29 July 2024;
Published online: 09 August 2024

References

1. Johnson, D., Campbell, C. D., Lee, J. A., Callaghan, T. V. & Gwynn-Jones, D. Arctic microorganisms respond more to elevated UV-B radiation than CO_2. *Nature.* **416**, 82–83 (2002).
2. Kong, W. D. A review of microbial diversity in polar terrestrial environments. *Biodiversity Science.* **21**, 456–467 (2013).
3. Yang, S. P., Jiang, J. & Yan, P. Eco-geographic distribution of rare and endangered plants and endemism in the Pamirs Region of China. *Journal of Arid Land Resources and Environment.* **32**, 115–120 (2018).
4. Mętrak, M., Sulwiński, M., Chachulski, Ł., Wilk, M.B., Suskamalawska, M. Creeping environmental problems in the pamir mountains: Landscape conditions, climate change, wise use and threats. In Climate Change Impacts on High-Altitude Ecosystems. *Springer Science and Business Media LLC: Berlin, Germany*. 665–694 (2015).
5. Yao, T. *et al*. Different glacier status with atmospheric circulations in Tibetan Plateau and surroundings. *Nat. Clim. Change.* **2**(9), 663–667 (2012).
6. Mętrak, M. *et al*. Nature's Patchwork: How water sources and soil salinity determine the distribution and structure of halophytic plant communities in arid environments of the Eastern Pamir. *PLoS ONE* **12**, e0174496 (2017).
7. Xiang, S. R., Yao, T. D., An, L. Z., Xu, B. L. & Wang, J. X. 16S rRNA Sequences and Differences in Bacteria Isolated from the Muztag Ata Glacier at Increasing Depths. *Applied and Environmental Microbiology.* **71**, 4619–4627 (2005).
8. Singh, V.P., Singh, P., Haritashya, U.K. (Eds.). Pamirs. *Encyclopedia of Snow, Ice and Glaciers*. 13–815 Springer, Netherlands (2011).
9. Mölg, N., Bolch, T., Rastner, P., Strozzi, T. & Paul, F. A consistent glacier inventory for Karakoram and Pamir derived from Landsat data: distribution of debris cover and mapping challenges. *Earth Syst. Sci. Data* **10**, 1807–1827 (2018).
10. Barandun, M. *et al*. Hot Spots of glacier mass balance variability in central Asia. *Geophysical Research Letters* **48**, e2020GL092084 (2021).
11. Monika, M. *et al*. Morphology and distribution of biological soil crusts and their potential role in soil-forming processes under dry high-altitude periglacial conditions (Eastern Pamir, Tajikistan). *Geoderma Regional.* **33**, e00636 (2023).
12. Zhang, Q. & Duan, K. Q. Characteristics of precipitation in the Pamirs in 2017 based on WRF simulation. *Arid Land Geography* **44**, 1707–1716 (2021).
13. Veresoglou, S. D., Halley, J. M. & Rillig, M. C. Extinction risk of soil biota. *Nature Communications*. **6**, 8862–8871 (2015).
14. Ma, B., Wang, H., Dsouza, M. & Xu, J. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME Journal.* **10**, 1891–1901 (2016).
15. Li, M. Y., Wang, J. L. & Gulbahar, S. Culturable bacterial diversity in snow, ice and melt water of theYangbark Glacier, Muztag Ata. *Journal of Glaciology and Geocryology.* **37**, 1634–1641 (2015).
16. Fan, S. Y. *et al*. Variations in Soil Enzyme Activities and Microbial Communities along an Altitudinal Gradient on the Eastern Qinghai–Tibetan Plateau. *Forests* **12**, 681–681 (2021).
17. Li, Y. M. *et al*. Function and distribution of nitrogen-cycling microbial communities in the Napahai plateau wetland. *Archives of Microbiology* **205**, 357–357 (2023).
18. Xu, M. J. *et al*. Sensitivity of soil nitrifying and denitrifying microorganisms to nitrogen deposition on the Qinghai-Tibetan plateau. *Annals of Microbiology volume* **71**, 6 (2021).
19. Xu, Y. T., Sun, R., Yan, W. M. & Zhong, Y. Q. W. Divergent response of soil microbes to environmental stress change under different plant communities in the Loess Plateau. *Catena* **230**, 107240 (2023).
20. Agre, N. S. & Orleanskii, V. K. Thermophil actinomycetes in Pamir and their antagonistic properties. *Mon Bull Minist Health Public Health Lab Serv.* **31**, 95–102 (1962).
21. KuznetsovM, V. D. Actinomycetes of various Pamir soils and their antagonistic properties. *Mikrobiologiia* **29**, 563–570 (1969).
22. Aksenov, S. I., Babyeva, I. P. & Golubev, V. I. On the mechanism of adaptation of microorganisms to conditions of extreme low humidity. *Life Sci Space Res.* **11**, 55–61 (1973).

23. Szymon, Z., Marcin, N., Janusz, B., Piotr, M. & Arkadiusz, N. Fungal root endophyte associations of plants endemic to the Pamir Alay Mountains of Central Asia. *Symbiosis* **54**, 139–149 (2011).
24. Khomutovska, N. *et al*. Life in extreme habitats: diversity of endolithic microorganisms from cold desert ecosystems of eastern Pamir. *Polish Journal of Ecology* **65**, 303–319 (2017).
25. Khomutovska, N., de los Ríos, A. & Jasser, I. Diversity and Colonization Strategies of Endolithic Cyanobacteria in the Cold Mountain Desert of Pamir. *Microorganisms* **9**, 6 (2021).
26. Bu, X. Y., Ren, M., Wan, C. X. & Zhang, L. L. Diversity of aerobic cold-adapted bacteria and archaea isolated from the Pamir Plateau. *Acta Microbiologica Sinica.* **62**, 2568–2581 (2022).
27. Wei, S. *et al*. Comparative evaluation of three archaeal primer pairs for exploring archaeal communities in deep-sea sediments and permafrost soils. *Extremophiles* **23**, 747–757 (2019).
28. Walters, W. *et al*. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems* **22**, e00009–15 (2015).
29. Usyk, M., Zolnik, C. P., Patel, H., Levi, M. H. & Burk, R. D. Novel ITS1 Fungal Primers for Characterization of the Mycobiome. *mSphere* **13**, e00488–17 (2017).
30. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal.* **17**, 10–2 (2011).
31. Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J. & Holmes, S. P. Bioconductor workflow for microbiome data analysis: From raw reads to community analyses. *F1000Research.* **5**, 1492 (2016).
32. Yilmaz, P. *et al*. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research.* **42**, D643–D8 (2014).
33. Nilsson, R. H. *et al*. K. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* **47**, D259–D64 (2018).
34. McMurdie, P. J. & Holmes, S. phyloseq: An r package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE.* **8**, e61217 (2013).
35. Wilkinson, S. kmer: an R package for fast alignment-free clustering of biological sequences. https://cran.r-project.org/package=kmer Accessed (2018).
36. Oksanen, J. *et al* vegan: Community Ecology Package. *R package version 2.5-6.* (2019).
37. Zhu, J., Wang, H. N., Tang, Q. Y., Gu, M. Y. & Zhang, Z. D. Composition and distribution of bacterial communities and potential radiation-resistant bacteria at different elevations in the eastern Pamirs. *Front. Microbiol.* **15**, 1427806 (2024).
38. Ekstrøm, C. MESS: Miscellaneous Esoteric Statistical Scripts. CRAN.R-project.org/package=MESS. Accessed (2022).
39. Goslee, S. C. & Urban, D. L. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* **22**, 1–19 (2007).
40. *NCBI Sequence Read Archive.* https://identifiers.org/ncbi/insdc.sra:SRP468803 (2023).
41. Zhu, J. Amplicon data of 80 soil archaea from the Pamir Plateau. *figshare* https://doi.org/10.6084/m9.figshare.26087533 (2024).
42. Zhu, J. Amplicon data of 80 soil bacteria from the Pamir Plateau. *figshare* https://doi.org/10.6084/m9.figshare.26087554 (2024).
43. Zhu, J. Amplicon data of 80 soil fungi from the Pamir Plateau. *figshare* https://doi.org/10.6084/m9.figshare.26087566 (2024).

Acknowledgements

This work was supported by the Project of Fund for Stable Support to Agricultural Sci-Tech Renovation (xjnkywdzc-2023005).

Author contributions

J.Z., X.S., Z.D.Z. and J.W.C. designed the study. J.Z., Q.Y.T., M.Y.G. and Z.D.Z. collected the samples. X.S. and J.Z. performed the analysis. J.Z. and X.S. wrote the paper and prepared the figure and tables. All co-authors commented on the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Z.-D.Z. or J.-W.C.