Supplement of Biogeosciences, 22, 6275–6289, 2025
https://doi.org/10.5194/bg-22-6275-2025-supplement

Biogeosciences
Open Access

Supplement of

On-site microbiome study of silica structures in a subterranean Mars analog environment

Martina Cappelletti et al.

Correspondence to: Martina Cappelletti (martina.cappelletti2@unibo.it)

Supplementary File

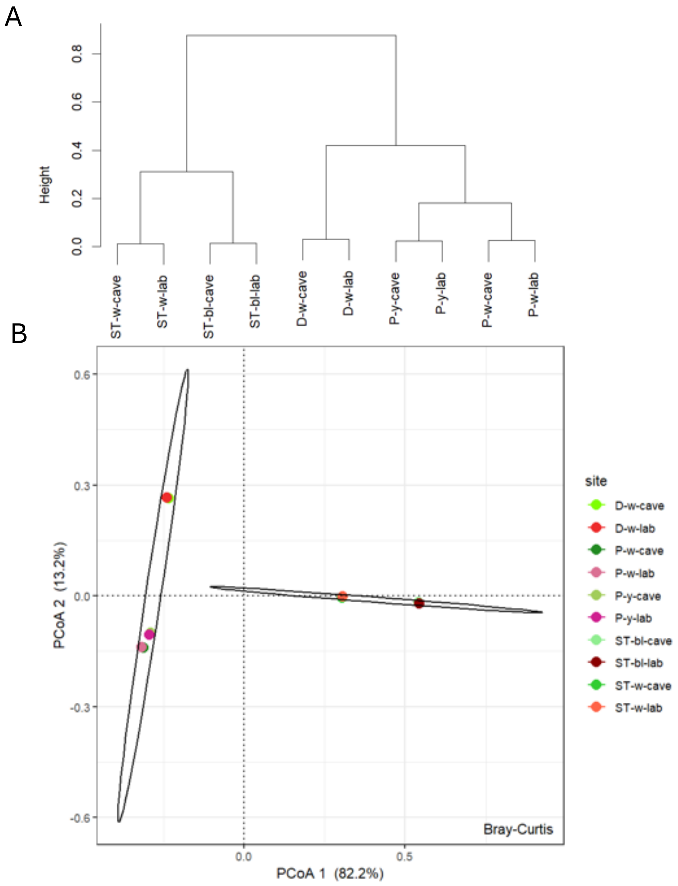

Figure S1: Comparison between the results obtained using the Cave_ and Lab_pipeline in terms of beta-diversity. The legend acronyms refer to Table S1. Cluster dendrogram (A) and Principal coordinates analysis (B) of Bray-Curtis dissimilarities of all the samples analysed using cave_pipeline (names with "-cave") and lab_pipeline (names with "-lab).

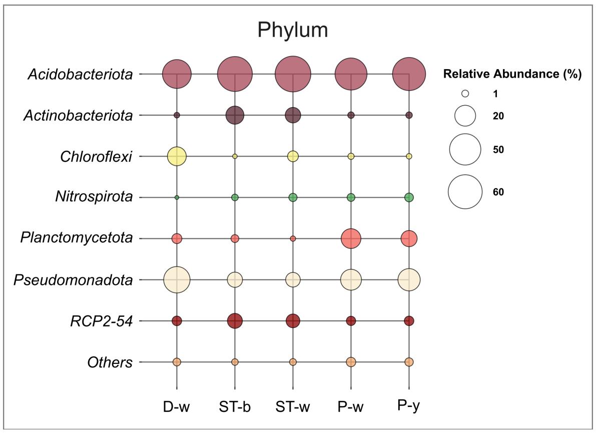

Figure S2: Bubble plot showing the most abundant microbial taxa in each sample at the phylum level.

Table S1: **Description of the samples collected in Imawarì Yeuta cave, including samples number, barcode association and sampling site characteristics**

Sample code	Description
D-w	White dots with dendritic aspect, similar to Qz samples that were previously described in Ghezzi et al. (2021, 2022)
P-y	Yellow patina/biofilm on the quartzite floor of Imawarì Yeutà
P-w	White patina/biofilm on the quartzite floor of Imawarì Yeutà immediately around the yellow patina
ST-w	White colonies/patina at the interface of the silica stromatolite with the quartzite rock
ST-b	Blackish/greyish patina covering the top part of the silica stromatolite

Table S2: Description of the main characteristics of the two pipelines that were applied to analyse the ONT sequencing data including rate of subsampling, basecalling applied, internet need, and time to process 100k row reads

	Cave_procedure	Lab_procedure
Subsampling	25%	100%
Basecalling mode	Fast[a]	Accurate
Internet connection	Not needed	Preferable[b]
Taxonomy classification	EMU classifier (Silva database)	EMU classifier (Silva database)
Time to process 100k raw reads	Ca. 10 min[c]	Ca. 45 min - 1 h

[a] Basecalling was directly performed during the sequencing run by the MinK1C Minion

[b] Internet connection is needed with computers/laptops with standard computational capacity

[c] Carrying out the subsampling

Table S3: Ecoplate carbon sources classification

Biochemical classification	**Compounds**
Carbohydrates	B-Methyl-D-Glucoside
	D-Xylose
	i-Erythritol
	D-Mannitol
	N-Acetyl-D-Glucosamine
	D-Cellobiose
	Glucose-1-Phosphate
	α-D-Lactose
	D,L-α-Glycerol Phosphate

Carboxylic acid	D- Galactonic Acid γ-Lactone
	Pyruvic Acid Methyl Ester
	D-Galacturonic Acid
	2-Hydroxy Benzoic Acid
	4-Hydroxy Benzoic Acid
	γ-Amino Butyric Acid
	D-Glucosaminic Acid
	Itaconic Acid
	α-Keto Butyric Acid
	D-Malic Acid
Complex carbon sources	Tween 40
	Tween 80
	α-Cyclodextrin
	Glycogen
Amino acids	L-Arginine
	L-Asparagine
	L-Phenylalanine
	L-Serine
	L-Threonine
	Glycyl-L-Glutamic Acid
Amines	Phenylethyl-amine
	Putrescine

Table S4: Description of the raw sequencing data obtained by processing ONT sequencing reads obtained in the cave using the Cave_pipeline and Lab_pipeline

Sample	Cave_pipeline			Lab_pipeline		
	# basecalled reads	Filtered reads	# reads mapped by EMU	# basecalled reads	Filtered reads	# reads mapped by EMU
P-y	117509	27233	27231	105820	97993	97980
P-w	102785	24244	24240	94623	88661	88649
ST-w	132029	24476	24463	113894	86442	86397
ST-bl	63345	13473	13466	56284	48034	48023
D-w	135197	32719	32715	124739	119934	119917

Table S5: Statistical analysis of the correlation between the cave and lab datasets for each sample considering the whole dataset using Chi-square goodness of fit test

Sample/Barcode	p-value[a]				
	Taxonomy level				
	Phylum	Class	Order	Family	Genus
P-y	>0.99*	>0.99*	>0.99*	>0.99*	>0.99*
P-w	>0.99*	>0.99*	>0.99*	>0.99*	>0.99*
ST-w	>0.99*	>0.99*	>0.99*	>0.99*	>0.99*
ST-b	>0.99*	>0.99*	>0.99*	>0.99*	>0.99*
D-w	>0.99*	>0.99*	>0.99*	>0.99*	>0.99*

[a] Calculated through Chi-square goodness of fit test setting the null hypothesis (H0) = the two samples are NOT significantly different

[*] Values for which H0 is NOT rejected, the two datasets are not statistically different

Table S6:Shannon index values calculated using the results obtained from the cave_pipeline and the lab_pipeline

Sample/Barcode	Shannon index									
	Phylum		Class		Order		Family		Genus	
	CAVE	LAB	CAVE	LAB	CAVE	LAB	CAVE	LAB	CAVE	LAB
P-y	1.31	1.33	1.46	1.48	2.51	2.55	2.68	2.72	2.86	2.91
P-w	1.41	1.42	1.57	1.62	2.58	2.64	2.78	2.84	2.96	3.01
ST-w	1.17	1.19	1.25	1.27	1.83	1.86	1.84	1.88	1.92	1.96
ST-bl	1.18	1.19	1.29	1.28	1.51	1.52	1.52	1.53	1.57	1.58
D-w	1.36	1.37	1.58	1.60	2.49	2.52	2.67	2.70	3.04	3.03

Table S7: Percentage of Amplicon sequence variants (ASV) unclassified at all taxonomy levels in each sample

Sample/Barcode	Unclassified ASVs %				
	Taxonomy level				
	Phylum	Class	Order	Family	Genus
P-y	<0.001	3.30	5.64	59.91	79.99
P-w	<0.001	3.32	5.91	54.12	78.67
ST-w	<0.001	8.07	10.08	79.25	86.67
ST-b	<0.001	9.67	15.71	84.26	90.18
D-w	<0.001	3.54	4.92	45.12	66.97