

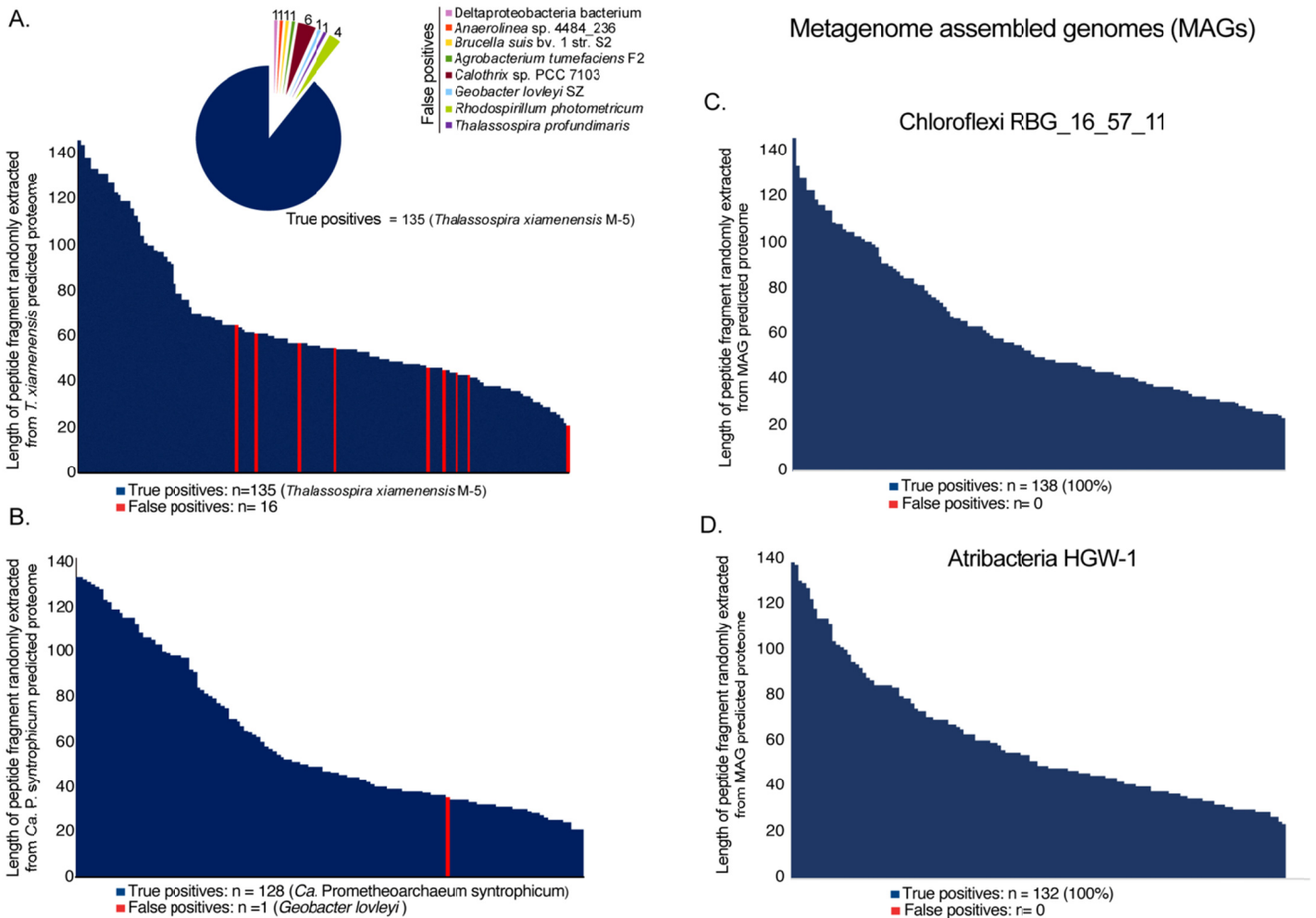
## *Supporting Information*

### **Exploring the abundance, metabolic potential, and gene expression of subseafloor *Chloroflexi* in million-year-old oxic and anoxic abyssal clay**

Aurèle Vuillemin\*, Zak Kerrigan, Steven D'Hondt, and William D. Orsi

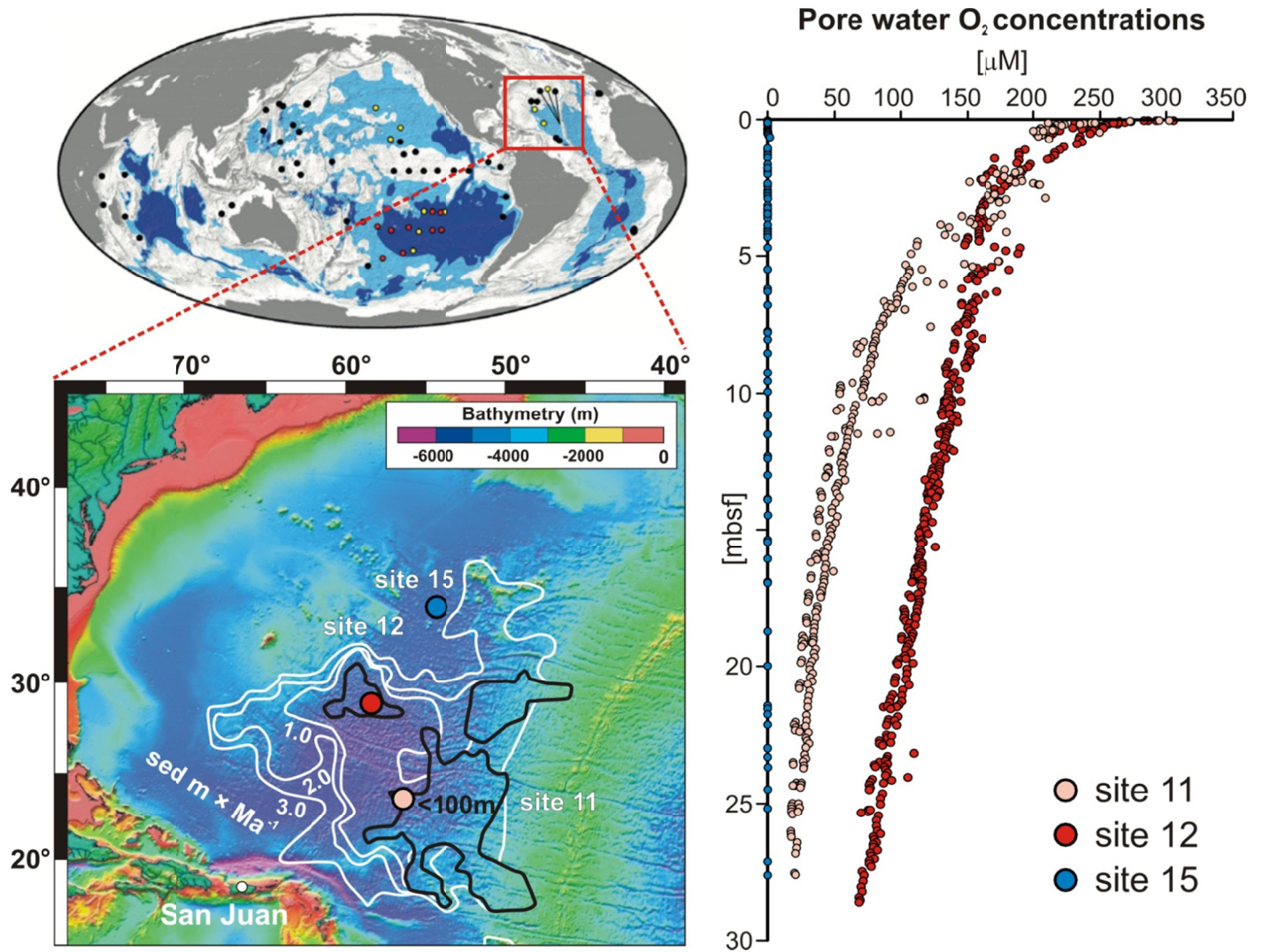
The Supporting Information includes:

- 5 Supplementary Figures
- 1 Supplementary Table
- Supplementary Data (provided separately)

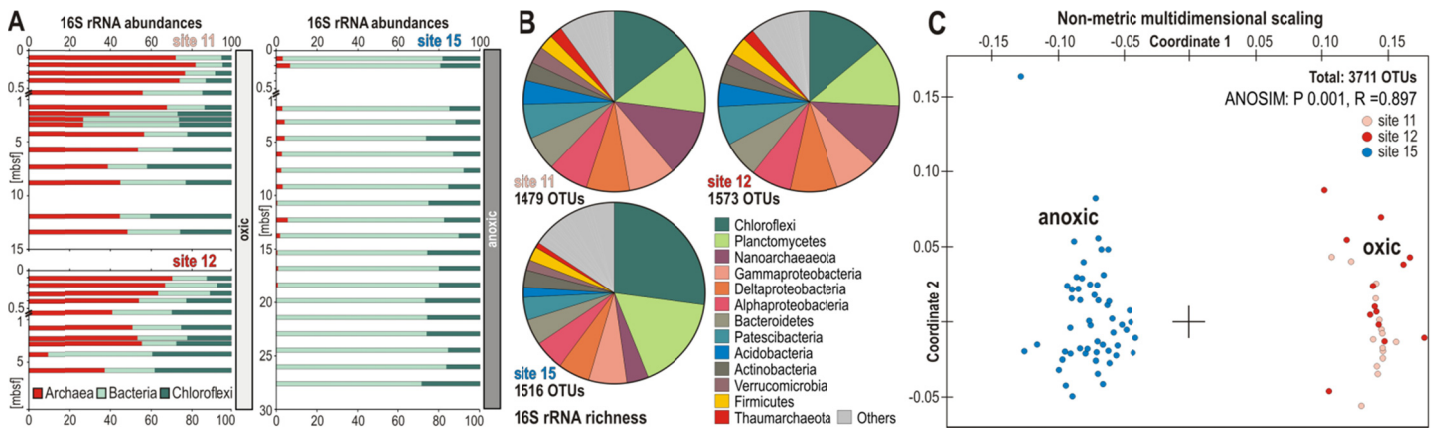


**Supplementary Figure 1 | An *in silico* experiment testing the rate of true positives and false negative annotations of 151 simulated ORFs (randomly selected peptide fragments) from the predicted proteomes of a cultivated bacterium and three metagenomic-assembled genomes. (A)** 89% of simulated ORFs (n=151 randomized peptide fragments) are correctly annotated to the species level (“*T. xiamenensis*”), and 93% of simulated ORFs are correctly annotated to the genus level when comparing against MetaProt. **(B)** The same experiment applied to the predicted proteomes of the archaeon “*Ca. Prometheoarchaeum syntrophicum*”, **(C)** Chloroflexi and **(D)** “*Ca. Atribacteria*” MAGs shows that the false negative rate is much lower than compared to the bacterial genome: only one out of 150 peptide fragments were incorrectly assigned.

The red and blue colors in the histogram plots represent those random peptide fragments that were assigned as false negatives and true positives, respectively: e.g., those that had highest similarity (best BLASTp hit) to an organism other than “*T. xiamenensis*”, “*Ca. Prometheoarchaeum syntrophicum*”, Chloroflexi and “*Ca. Atribacteria*” in the MetaProt database (e.g., the individual thin slices in the pie chart figure above).

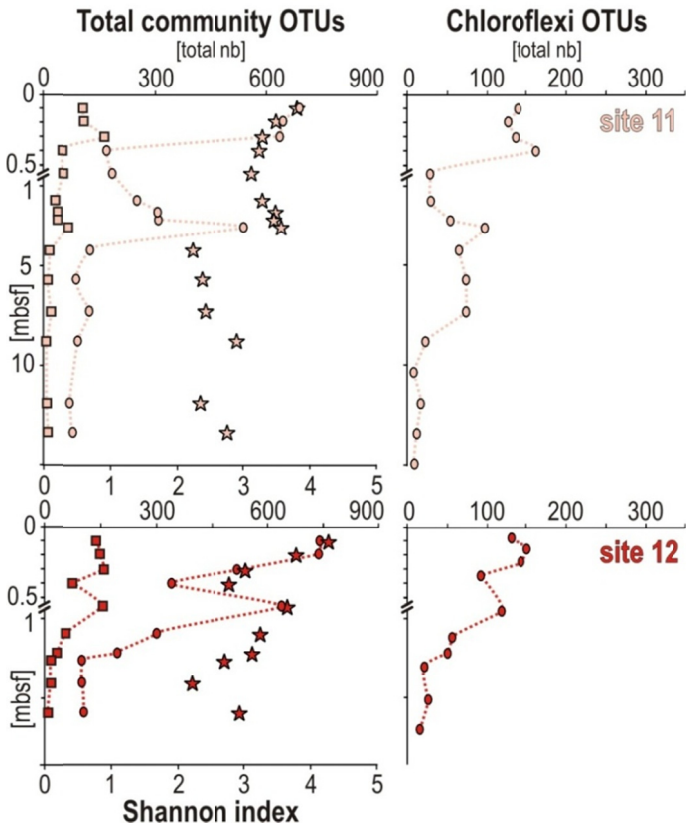


**Supplementary Figure 2 | Map of sampling site 11, 12 and 15 with bathymetry, sedimentation rates and pore water O<sub>2</sub> profiles.** (Left) Global map of subseafloor sedimentation rates and corresponding O<sub>2</sub> penetration is modified from D'Hondt *et al.* (2015). (Right) Note that at site 11 and 12, there is O<sub>2</sub> penetrating to 30 mbsf, whereas at site 15 O<sub>2</sub> is consumed immediately at the seafloor surface and the entire sediment sequence is anoxic.

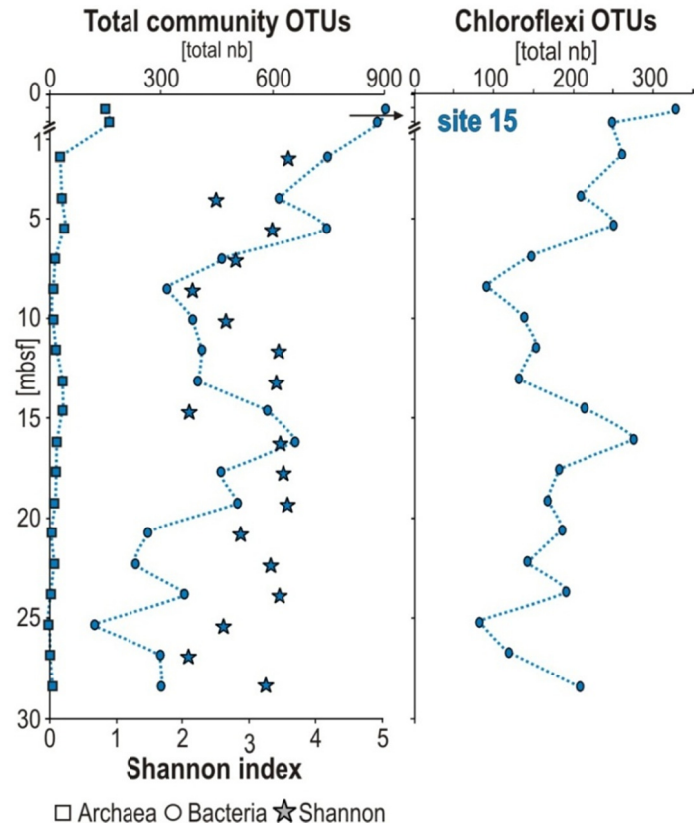


**Supplementary Figure 3 | Microbial abundance, richness beta diversity based on 16S rRNA genes.** (A) Downcore profiles for the relative abundances of Archaea (red), Chloroflexi (dark green) and all other Bacteria (light green). (B) Richness of 16S rRNA genes based on the total number of OTUs. (C) Non-metrical dimensional scaling (NMDS) plot for all three sites based on all OTUs obtained from 16S rRNA gene amplicons, with significant separation between oxic and anoxic sediment.

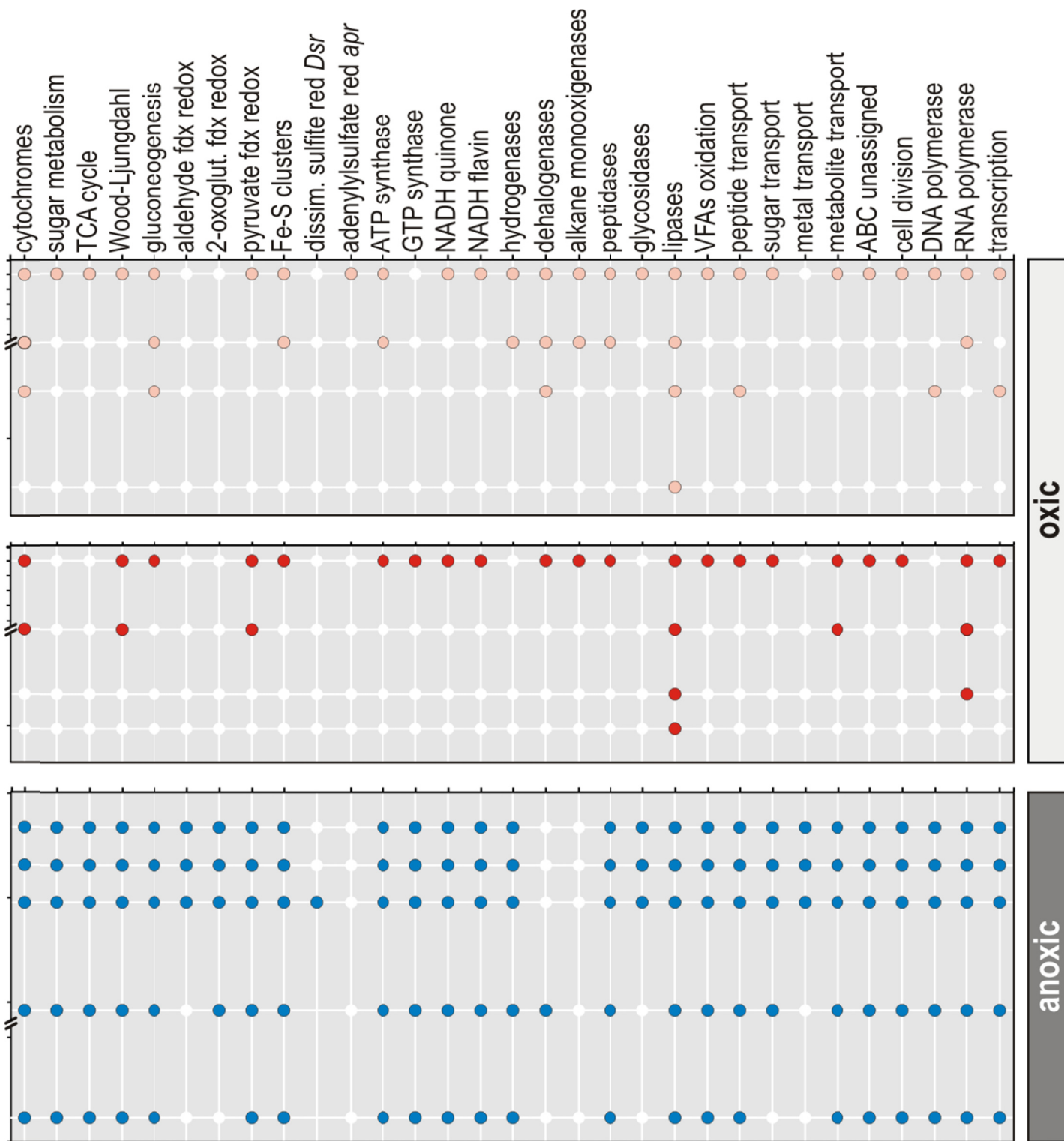
## Oxic sites



## Anoxic site



**Supplementary Figure 4 | Alpha diversity for whole communities and Chloroflexi at the three sites. (Left)** Total number of OTUs assigned to Archaea (squares), and Bacteria (circles) with Shannon indices (stars) calculated for the whole populations at the two oxic sites; number of OTUs assigned to Chloroflexi at the two oxic sites. **(Right)** Same indices calculated for populations from the anoxic site.



**Supplementary Figure 5 | Metabolic potential detected in metagenomic libraries.** Metabolic functions and pathways detected in metagenomes at the three sites. For a detailed list of the genes identified, see Supplementary Data.

Metagenomes	depth [m]	# sequences	# contigs	# mapped reads	# annotated ORFs	Accession no.	Run Accession no.
Site 11	0.1	15,314,358	104,349	10,153,456	42,254	SAMN10924458	SRX5372537
	0.9	6,479,556	8.857	5,103,508	5,137	SAMN10924458	SRX5372538
	2.8	5,593,704	8.952	5,126,686	3,871	SAMN10924458	SRX5372539
	7.4	10,074,790	602	8,235,415	276	SAMN10924458	SRX5372540
	15	8,920,356	144	7,503,283	18	SAMN10924458	SRX5372541
Site 12	0.1	5,298,416	17.05	4,181,095	9,041	SAMN10924459	SRX5372542
	0.9	3,961,194	3.464	3,741,957	2,151	SAMN10924459	SRX5372543
	3.5	6,528,088	1.407	6,200,598	522	SAMN10924459	SRX5372544
	5.1	5,353,236	32.673	7,503,283	67	SAMN10924459	SRX5372545
Site 15	2.1	9,294,114	29.185	8,723,374	13,134	SAMN13317858	SRR10481892
	3.5	9,226,836	15.319	8,517,416	22,876	SAMN13317859	SRR10481891
	5.2	4,158,110	12.734	3,488,721	9,137	SAMN13317860	SRR10481887
	15	8,321,518	10.194	7,435,919	8,568	SAMN13317861	SRR10481886
	29	10,746,984	2.725	10,039,210	2,894	SAMN13317862	SRR10481885
Transcriptomes	depth [m]	# sequences	# contigs	# mapped reads	# annotated ORFs	Accession no.	Run Accession no.
Site 15	0.1 (A)	4,025,956	1'795	3'104'523	3,343	SAMN13317863	SRR10481884
	0.1 (B)	3,532,876	2'044	2'770'901	98	SAMN13317863	SRR10481884
	0.1 (C)	4,960,078	3'174	4'055'833	305	SAMN13317863	SRR10481884
	2.1 (A)	5,193,626	2'382	3'243'072	329	SAMN13317864	SRR10481883
	2.1 (B)	6,188,052	1'782	5'327'917	188	SAMN13317864	SRR10481883
	2.1 (C)	21,222,363	31'254	19'007'366	276	SAMN13317864	SRR10481883
	3.5 (A)	3,640,918	861	3'121'393	152	SAMN13317865	SRR10481882
	3.5 (B)	2,017,858	546	1'612'249	135	SAMN13317865	SRR10481882
	3.5 (C)	3'419'882	1'124	2'343'420	2,963	SAMN13317865	SRR10481882
	3.5 (D)	3'054'074	746	2'420'948	1,827	SAMN13317865	SRR10481882
	3.5 (E)	1'949'964	550	1'428'941	1,304	SAMN13317865	SRR10481882
	3.5 (F)	2'742'258	656	2'298'551	1,545	SAMN13317865	SRR10481882
	3.5 (G)	3'093'748	1030	2'410'463	258	SAMN13317865	SRR10481882
	5.2 (A)	5'347'824	10'204	2'001'032	1,262	SAMN13317866	SRR10481881
	5.2 (B)	3'673'384	6'075	1'142'750	1,098	SAMN13317866	SRR10481881
	6.7 (A)	4'672'286	5'583	1'712'037	1,748	SAMN13317867	SRR10481880
	6.7 (B)	7'253'508	6'654	2'396'752	1,493	SAMN13317867	SRR10481880
	8.2 (A)	5'311'371	3'378	3'834'338	2,730	SAMN13317868	SRR10481890
	8.2 (B)	6'140'536	2'055	2'666'190	285	SAMN13317868	SRR10481890
	8.2 (C)	3'405'518	1'312	2'424'526	306	SAMN13317868	SRR10481890
9.7 (A)	6'159'540	3'346	3'256'870	623	SAMN13317869	SRR10481889	
9.7 (B)	4'494'022	2'571	1'519'639	560	SAMN13317869	SRR10481889	
15.9 (A)	1'091'556	1'192	849'855	72	SAMN13317870	SRR10481888	
15.9 (B)	3'214'672	835	2'790'696	132	SAMN13317870	SRR10481888	

**Supplementary Table 1 | Sequencing and assembly results for the subsurface metagenomes at site 11, 12 and 15, and transcriptomes at site 15 with their corresponding accession and run accession numbers.**