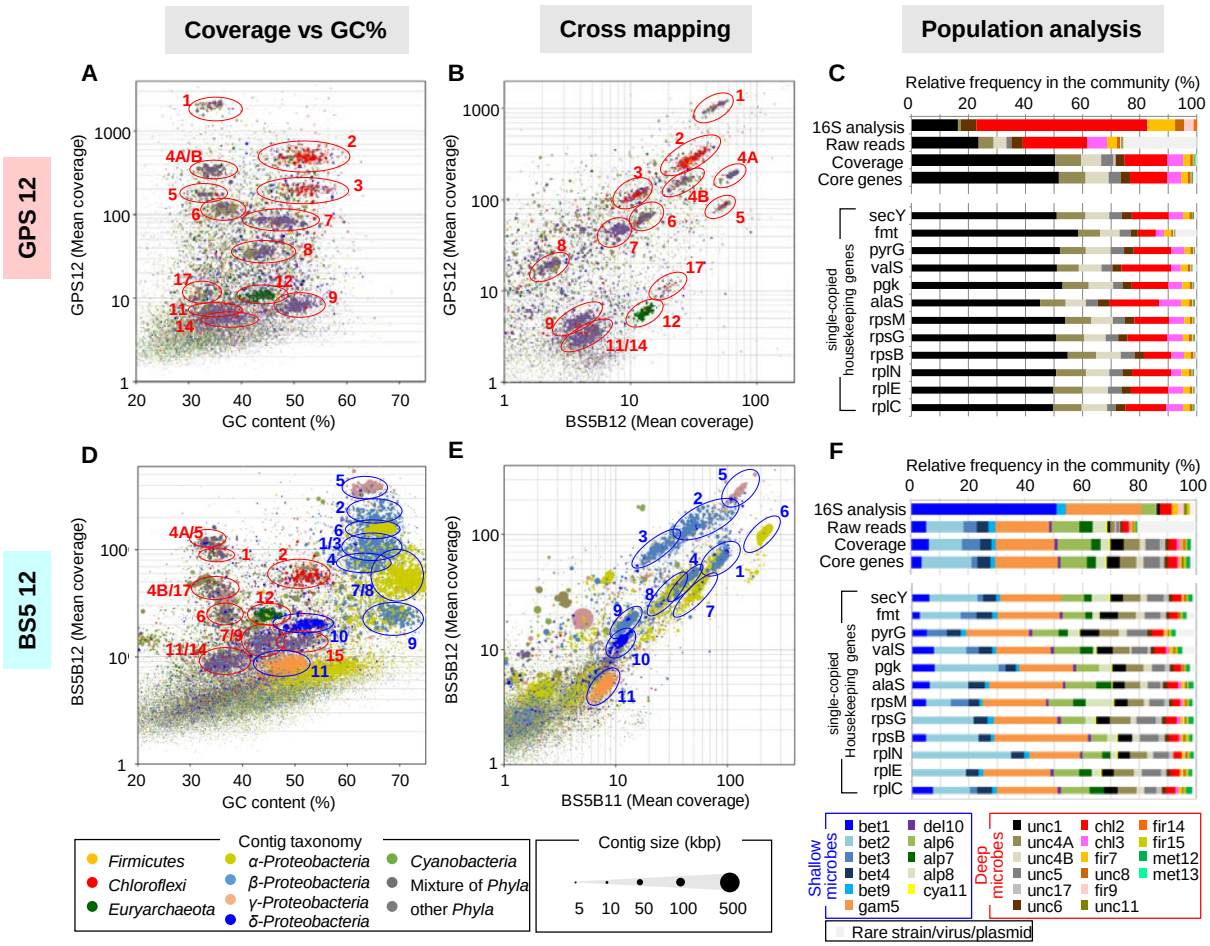
Unusual metabolic diversity of hyperalkaliphilic microbial communities associated with subterranean serpentinization at The Cedars

Shino Suzuki*, Shun'ichi Ishii*, Tatsuhiko Hoshino, Amanda Rietze, Aaron Tenney, Penny L. Morrill, Fumio Inagaki, J. Gijs Kuenen and Kenneth H. Nealson

* Corresponding authors
SS, E-mail: sisuzuki@jamstec.go.jp, Tel: +81-88-878-2224.
SI, E-mail: sishii@jamstec.go.jp, Tel: +81-88-878-2279.

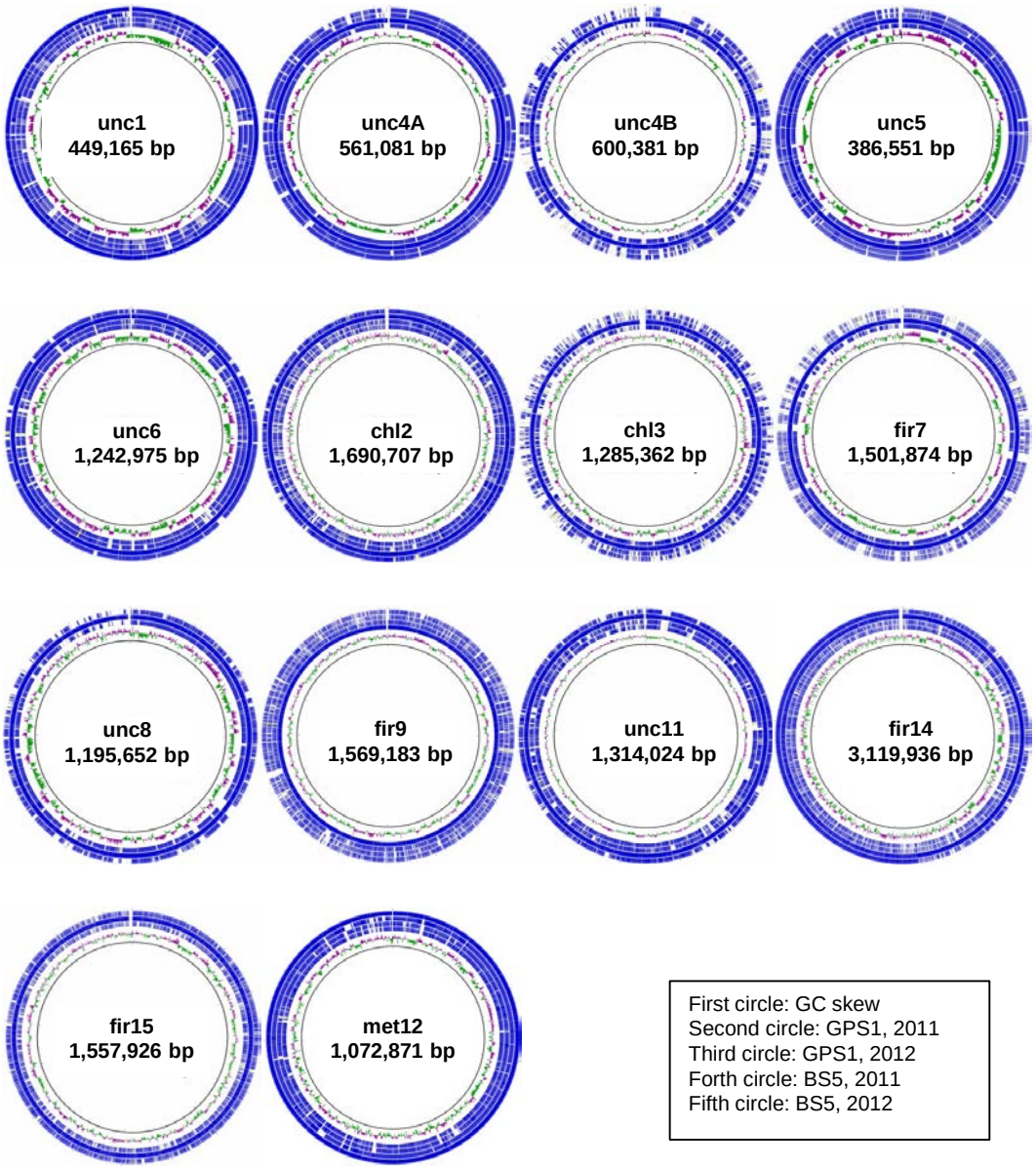
SUPPLEMENTARY INFORMATION

Table of Contents



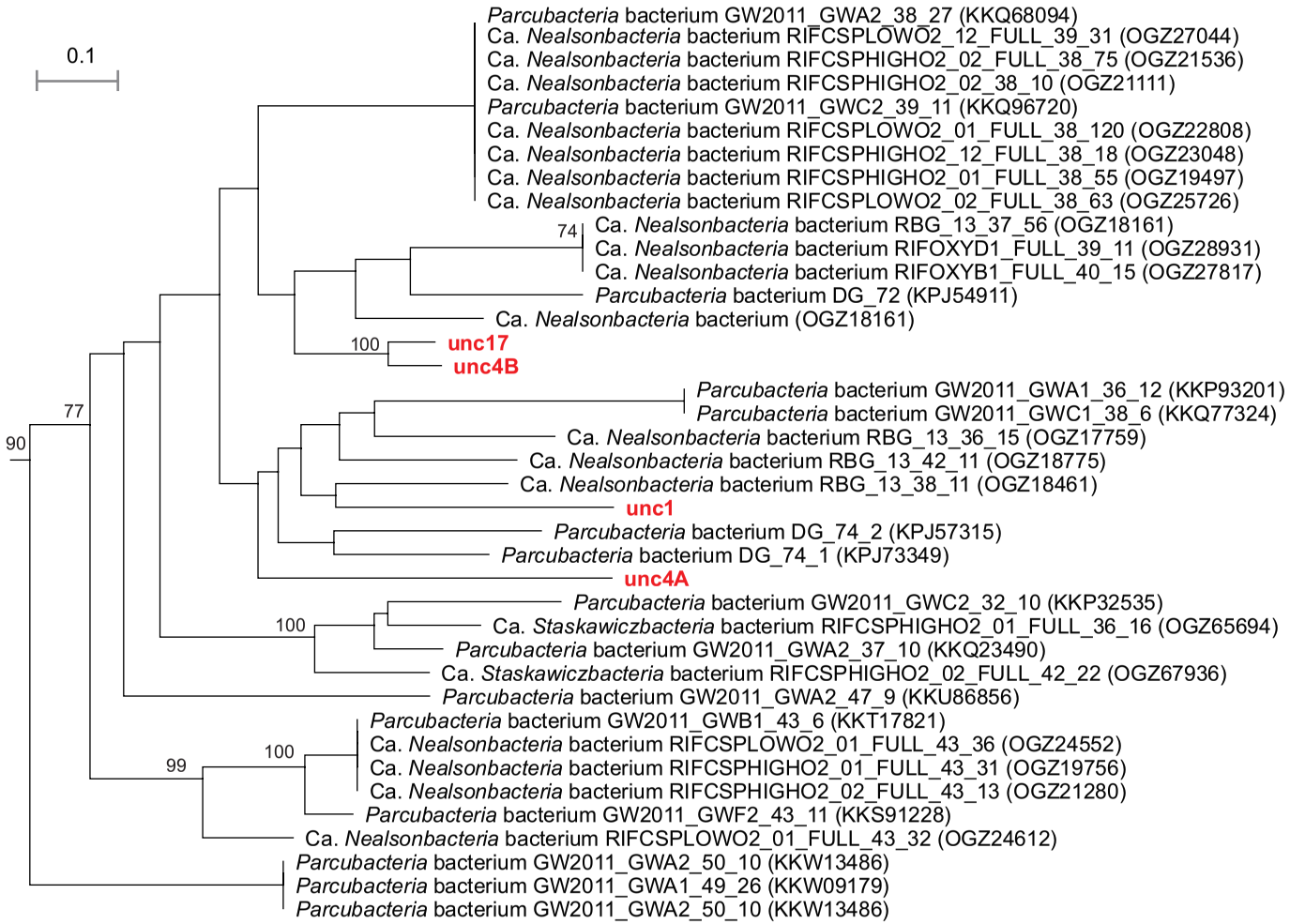
Supplementary Figure 1 | Genome clustering from the assembled contigs of metagenomic sequences of The Cedars springs GPS1 and BS5 in 2012.

Samples collected from GPS1 and BS5 in 2012 are shown as GPS12 or BS5B12 respectively. Colored circles indicate clusters of individual bin-genomes. Bin-genomes were identified using the estimated taxonomic classification (color of dots), length (size of dots), GC content and mean coverage of contigs. Colors for the circles denote groundwater sources where the bin-genomes are associated, i.e. red circles show the deep members, while blue circles show the shallow member. Numbers/letters shown near the circles correspond to the bin-genome IDs in Fig. 3A. (A, D) Clustering based on the mean coverage vs. GC% plot. (B) Clustering for the members of deep groundwater based on coverage-differences between the two springs. (E) Clustering for the members of shallow groundwater based on the coveragedifferences of two years' metagenomes. (C, F) Relative frequencies of each strain within GPS12 (C) and BS5B12 (F) estimated from a relative abundance of 16S rRNA gene (16S analysis), percentage of mapped raw reads to each bin-genome (Raw reads), an average value of reads per kilobase per million mapped reads (RPKMs) of each bin-genome (Coverage), an average RPKM value of 12 single-copied housekeeping genes (Core genes), and a RPKM value of twelve individual single-copied housekeeping genes.



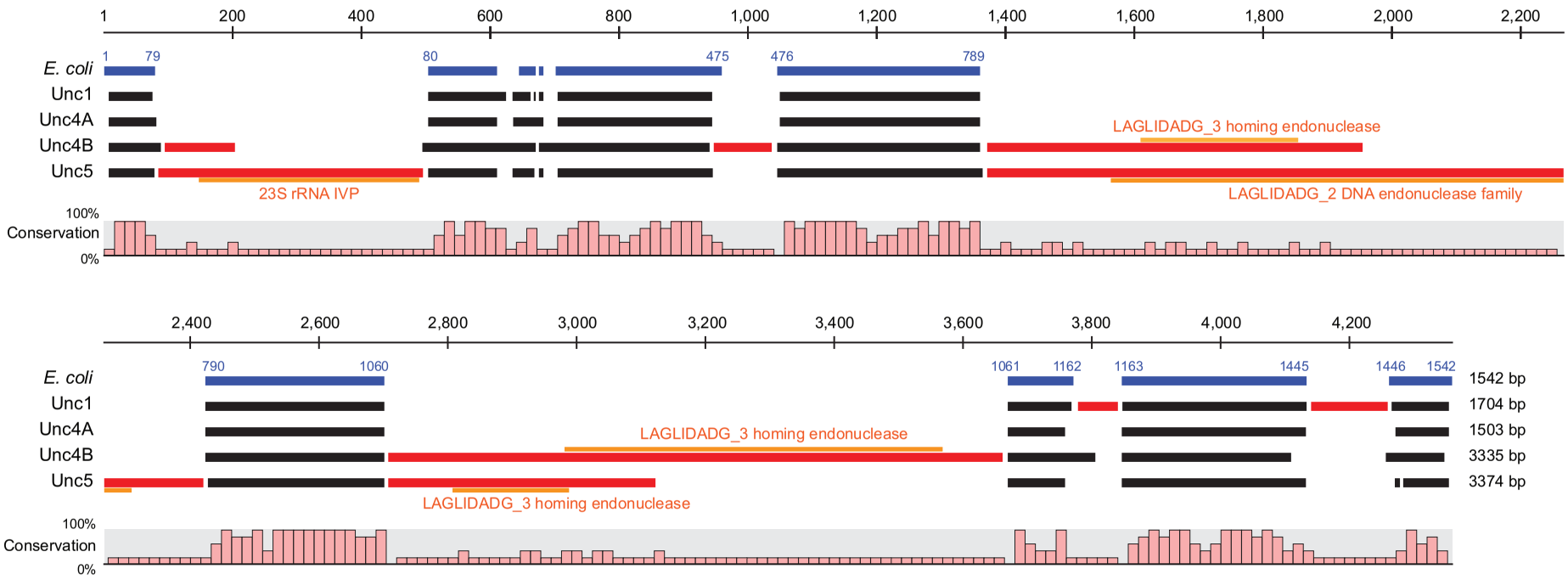
Supplementary Figure 2 | Blastn similarity of bin-genomes of deep members recovered from different metagenomic dataset from different springs (GPS1 and BS5) and years (2011 and 2012).

Due to the low coverage, bin-genomes of unc8 and fir15 were not allowed to recover from metagenomic data set of BS5 (2012) or GPS1 (2012) respectively.



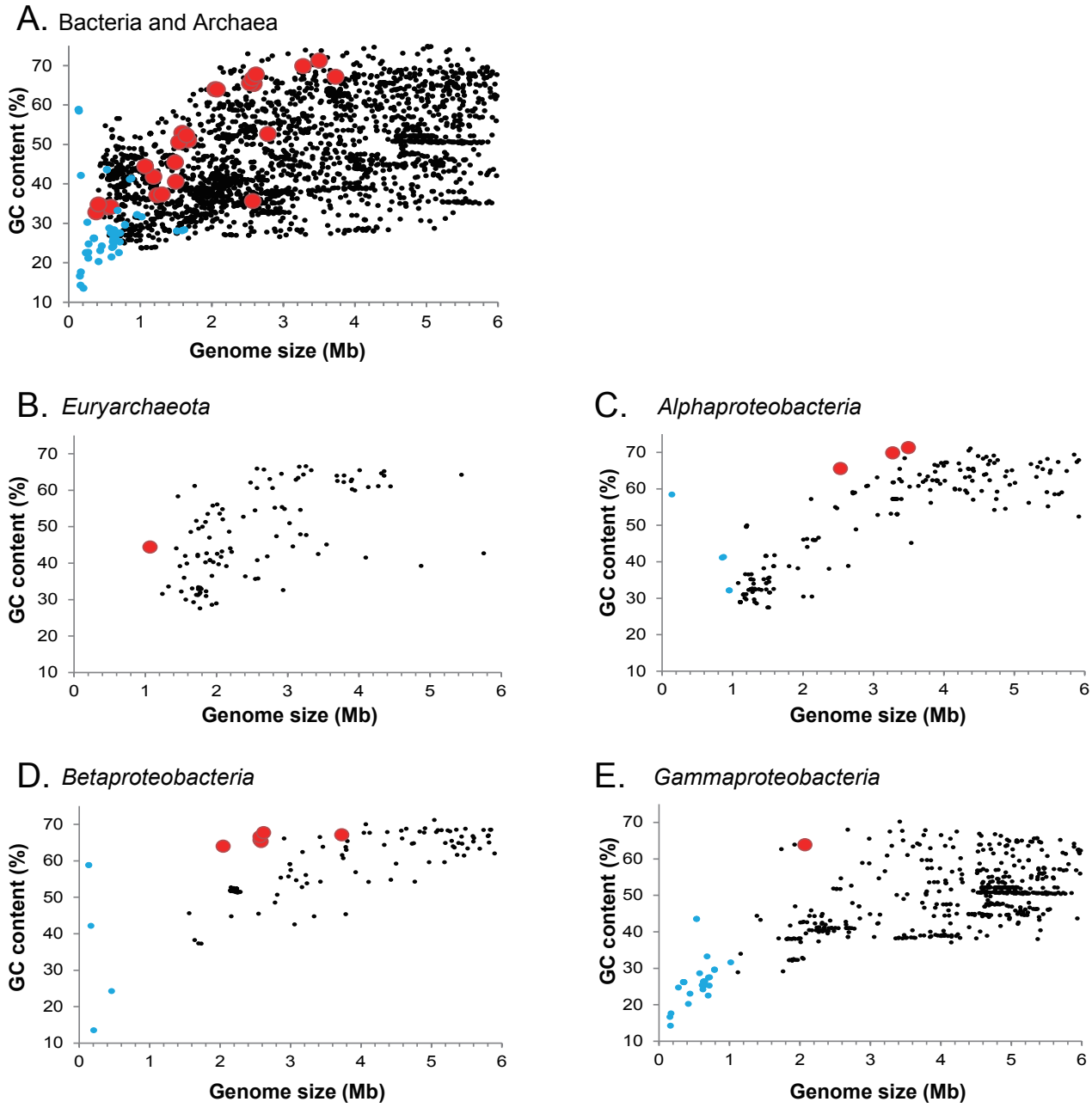
Supplementary Figure 3 | Maximum Likelyhood RPS3 phylogeney of OD1 genomes recovered from The Cedars.

Tree topologies are supported by bootstrap values for 100 replicates and the values greater than 70 were indicated. Red font denotes bin-genome ID from The Cedars serpentinization springs.



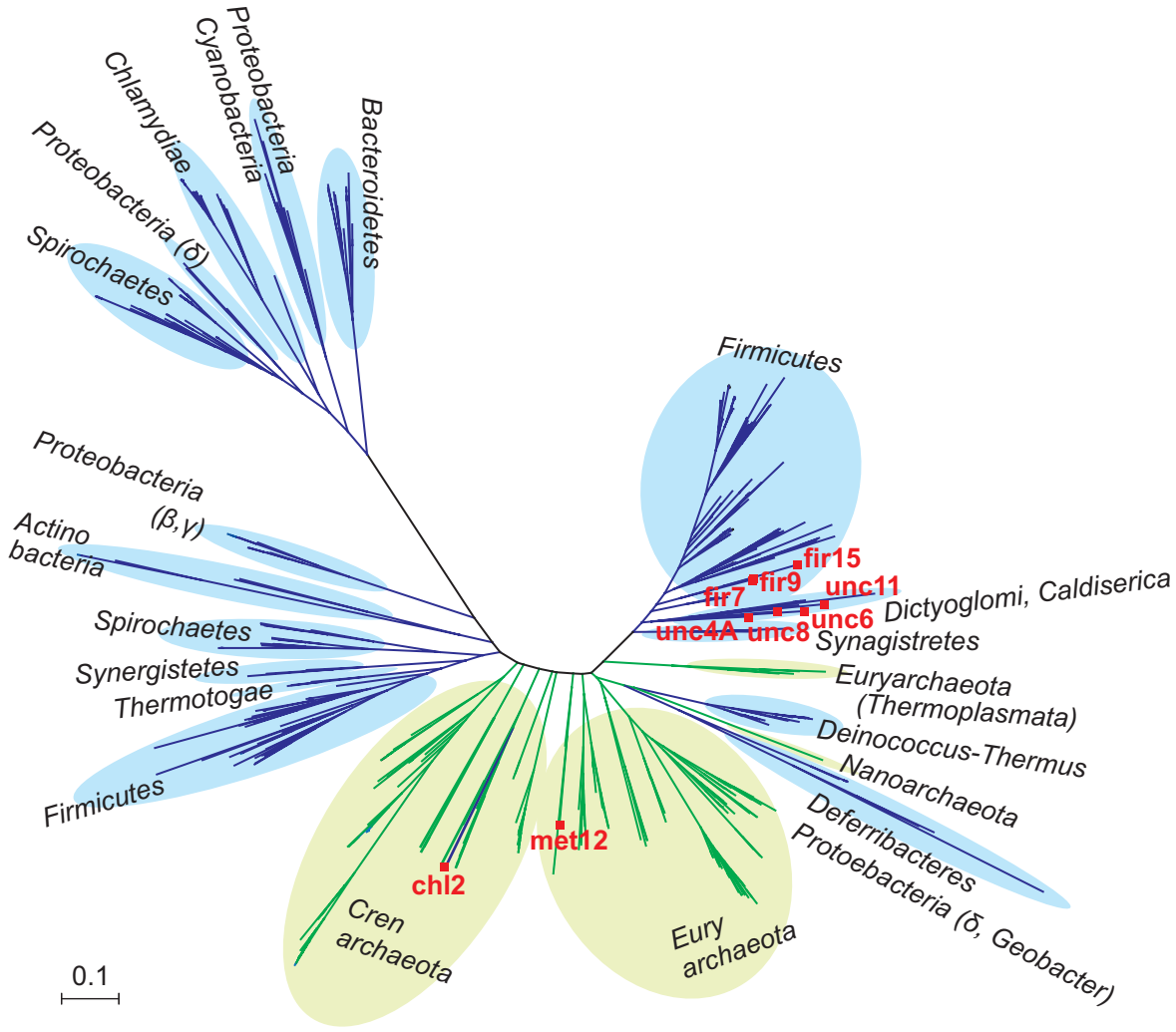
Supplementary Figure 4 | Intron-encoding 16S rRNA gene from the Candidate phylum OD1 bin-genomes.

Alignment of 16S rRNA encoding region corresponding to *Escherichia coli* K12 gene positions (position described above bars by blue numbers). Insertion regions are shown by red bars, while ORFs encoded in the insertion regions were identified by BLASTX against nr (orange bars) and applied to pfam search for functional annotation (orange words).



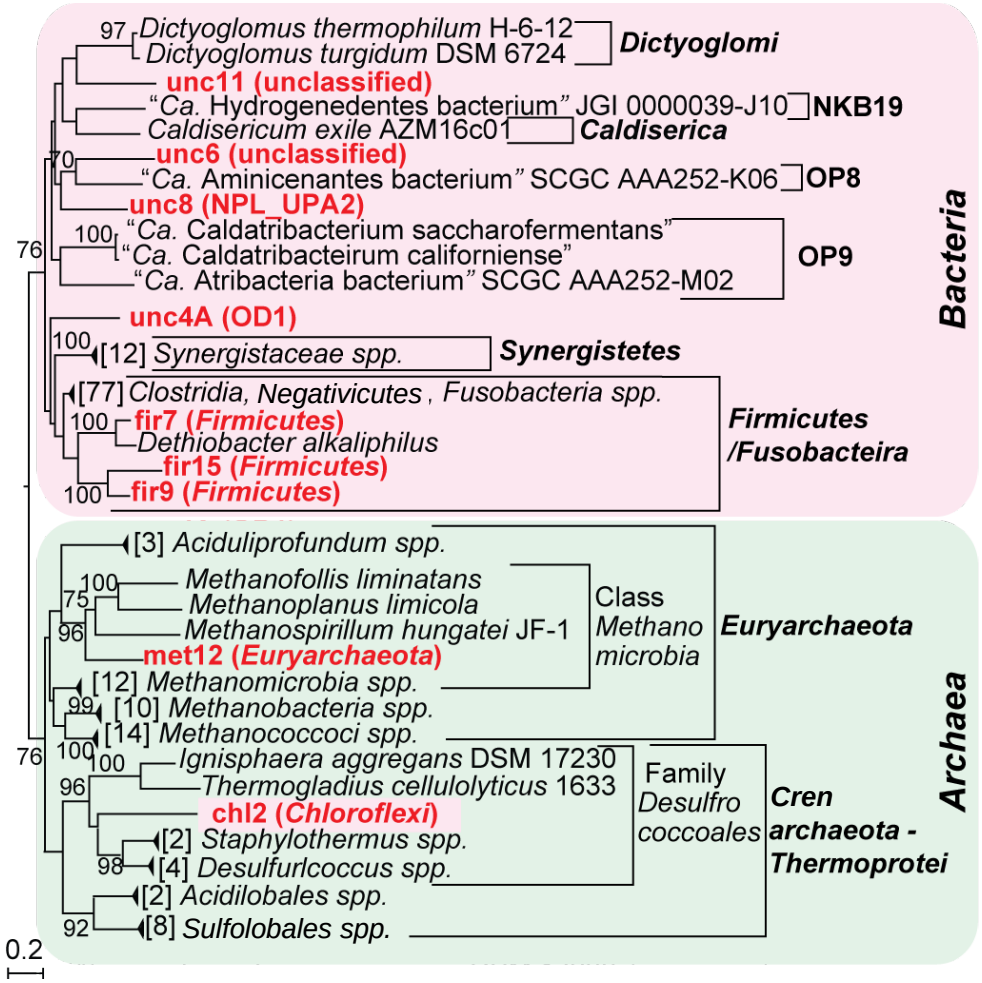
Supplementary Figure 5 | GC content versus genome size for the bin-genomes of The Cedars members with reference genome sequences in each taxonomic group.

Red, blue and black dots denote the bin-genomes from The Cedars organisms, bacterial endosymbionts/obligate parasite of eukaryotes, or genomes from free-living prokaryote respectively. Genome size and GC contents information for the references were obtained from all completed bacterial and archaeal genomes in the NCBI genome database. Total numbers of reference genomes analyzed here are 3170.



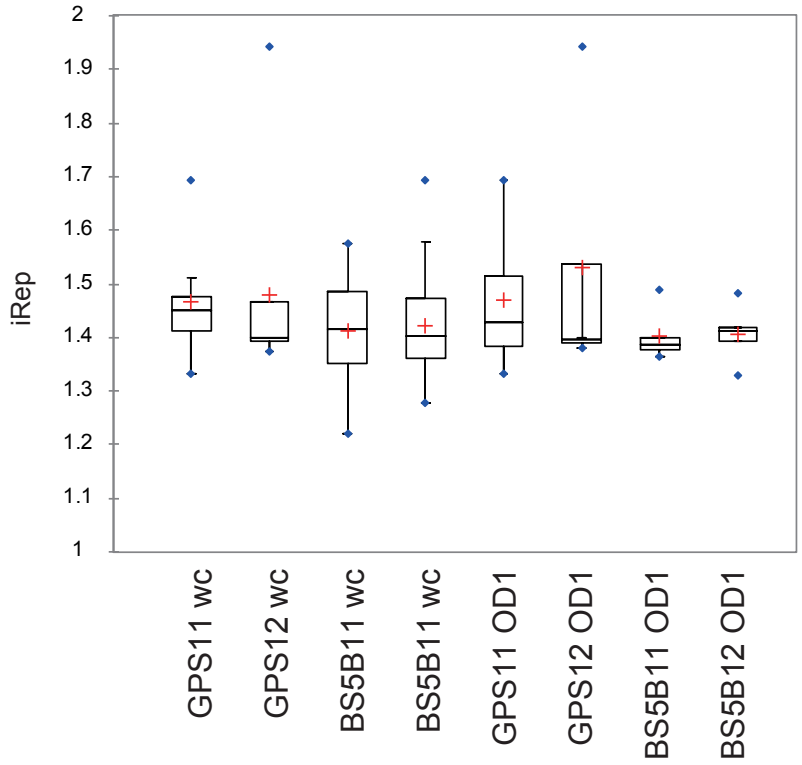
Supplementary Figure 6 | Maximum likelihood phylogenetic tree constructed for the A-type ATPase subunit A (NtpA).

Red letters show the position of NtpA from The Cedars organisms. Blue colored areas and nodes indicate bacterial phyla while green colored areas and nodes indicate the archaeal phyla. Reference sequences were obtained from KEGG database (K02117).

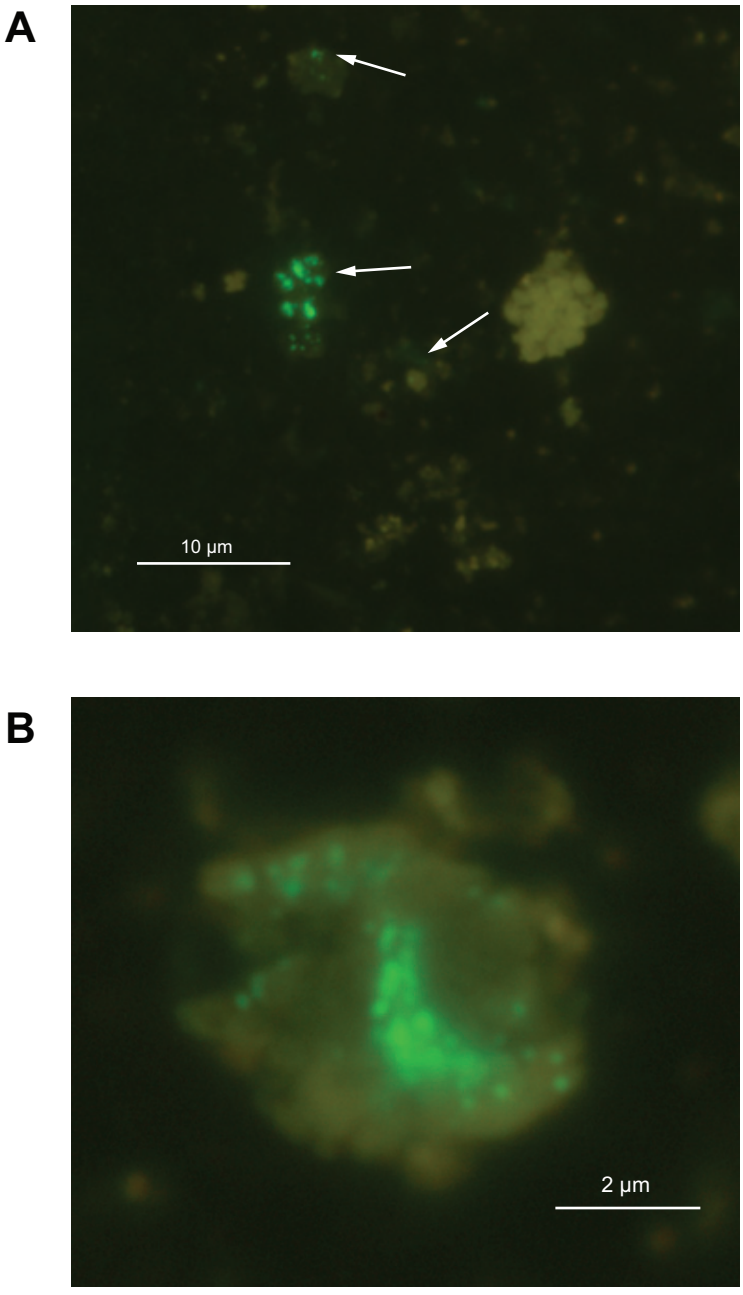


Supplementary Figure 7 | Phylogenetic analysis of NtpA.

Red letters show the position of NtpA from The Cedars organisms. Closest relatives in NCBI database (nr) were employed as the reference sequences.



Supplementary Figure 8 | iRep values were determined for whole community (wc) and OD1 in each metagenomic dataset.



Supplementary Figure 9 | CARD-FISH detection of OD1 organisms (green fluorescent) in GPS1.

A) A whole view of the CARD-FISH detection under fluorescent microscopy. Arrows indicate OD1 organisms. Detected OD1 organisms were associated with mineral particles.
B) A close view of the detection of OD1 organisms. The avarage cell size is around 0.2-0.3 µm.

Supplementary Table 1 Water geochemistry of BS5 and GPS1

	BS5	GPS1
pH	11.6 ± 0.1	11.9 ± 0.1
Temp. (ºC)	17.4 ± 0.5	17.1 ± 0.1
E_h (mV)	- 585±33	-656±9
Cond. (µS/cm)	870 ± 70	3010 ± 10
DO (mM)	<dl	<dl
N_2 (% by vol)	53.6	36.6±0.4
H_2 (% by vol)	34	50.9±1.1
CH_4 (% by vol)	5.3 ± 1.2	7.4±3.1
Na^+ (mM)	1.98±0.06	14.69±0.41
K^+ (mM)	0.03 ± 0.01	0.13 ± 0.01
NH_4^+ (mM)	<dl	<dl
Ca^{2+} (mM)	1.17 ± 0.11	0.94 ± 0.10
Mg^{2+} (mM)	0.036 ± 0.040	0.004 ± 0.004
Cl^- (mM)	1.49 ± 0.11	8.73 ± 0.10
SO_4^{2-} (mM)	0.001± 0.000	<dl
PO_4^{3-} (mM)	<dl	<dl
NO_3^- (mM)	<dl	<dl
TIC (mM)	0.07	0.035
DOC (mM)	0.02	0.17
f_{deep} [a]	0.14 ± 0.02	1
Reference	Morrill et al 2013	

Average data from multiple mesurements is shown.

[a] Average fraction of deep groundwater

Supplementary Table 2. Stats of metagenome assembly.

Sample	Total bases (bp)	Number of contig	N50 Length (bp)	Contigs over 1 kbp	Max Contig length	Original raw reads count	Mapped raw reads count[a]	% Mapped back (%)[a]
GPS1 (2011)	59,351,457	52,742	2,490	12,612	114,683	54,543,392	49,337,795	91.14
GPS1 (2012)	38,098,402	33,758	2,475	8,521	100,250	45,983,339	42,017,066	91.91
BS5B (2011)	88,663,216	66,400	3,969	16,255	179,628	55,675,843	48,255,451	86.58
BS5B (2012)	94,598,029	67,210	3,753	19,135	428,608	50,843,647	46,333,039	91.14

[a] Numbers of reads to perform mapping raw reads back to the assembled contigs (with length cutoff 0.8 and similarity cutoff 0.8).

[b] Numbers of reads to perform mapping raw reads back to the assembled contigs as paired (with length cutoff 0.8 and similarity cutoff 0.8).

Lists of Additional Data

Supplementary Data 1 | Single-copied housekeeping gene list for validation of bin-genome association.

Supplementary Data 2 | Single copied gene list for validation of draft genome assembly within domain Archaea

Supplementary Data 3 | Genomic features of all bin-genomes associated from metagenomes of GPS1 and BS5 springs in 2011 and 2012.

Supplementary Data 4 | Analysis of ANIb and TETRA among the recovered bin-genomes affiliated with the phylum OD1.

Supplementary Data 5 | Analysis of ANIb and TETRA among the recovered bin-genomes affiliated with the class Proteobacteria.

Supplementary Data 6 | Analysis of ANIb and TETRA among the recovered bin-genomes affiliated with the Firmicutes.

Supplementary Data 7 | Analysis of ANIb and TETRA among the recovered bin-genomes affiliated with the phylum Chloroflexi.

Supplementary Data 8 | Analysis of ANIb and TETRA among the recovered bin-genomes.

Supplementary Data 9 | Marker genes related to cell activity in The Cedars microbial communities.

Supplementary Data 10 | Metabolic marker genes related to respiration and fermentation in The Cedars microbial communities.

Supplementary Data 11 | MAPLE pathway analysis of dominant representative bin-genomes.

Supplementary Data 12 | MAPLE complex analysis of dominant representative bin-genomes.

Supplementary Data 13 | MAPLE functional analysis of dominant representative bin-genomes.

Supplementary Data 14 | Percentage of predicted genes per genome.