

RESEARCH NOTE

Open Access



# Complete genome sequence of the *Sulfodiicoccus acidiphilus* strain HS-1<sup>T</sup>, the first crenarchaeon that lacks *polB3*, isolated from an acidic hot spring in Ohwaku-dani, Hakone, Japan

Hiroyuki D. Sakai<sup>1,2</sup> and Norio Kurosawa<sup>1\*</sup>

## Abstract

**Objective:** *Sulfodiicoccus acidiphilus* HS-1<sup>T</sup> is the type species of the genus *Sulfodiicoccus*, a thermoacidophilic archaeon belonging to the order *Sulfolobales* (class *Thermoprotei*; phylum *Crenarchaeota*). While *S. acidiphilus* HS-1<sup>T</sup> shares many common physiological and phenotypic features with other *Sulfolobales* species, the similarities in their 16S rRNA gene sequences are less than 89%. In order to know the genomic features of *S. acidiphilus* HS-1<sup>T</sup> in the order *Sulfolobales*, we determined and characterized the genome of this strain.

**Results:** The circular genome of *S. acidiphilus* HS-1<sup>T</sup> is comprised of 2353,189 bp with a G+C content of 51.15 mol%. A total of 2459 genes were predicted, including 2411 protein coding and 48 RNA genes. The notable genomic features of *S. acidiphilus* HS-1<sup>T</sup> in *Sulfolobales* species are the absence of genes for *polB3* and the autotrophic carbon fixation pathway, and the distribution pattern of essential genes and sequences related to genomic replication initiation. These insights contribute to an understanding of archaeal genomic diversity and evolution.

**Keywords:** *Crenarchaeota*, *Sulfolobales*, *Sulfodiicoccus*, *polB3*, *cdc6*, 3-Hydroxypropionate/4-hydroxybutyrate, Dicarboxylate/4-hydroxybutyrate

## Introduction

*Sulfodiicoccus acidiphilus* HS-1<sup>T</sup>, represented a novel genus, was recently isolated in our laboratory and validly described [1]. The genus belongs to the order *Sulfolobales*, a well-known taxon of the phylum *Crenarchaeota*, widely inhabits hot acidic environments all over the world [2–5]. The 16S rRNA gene sequence similarities between *S. acidiphilus* and other species in the order *Sulfolobales* were less than 89%. Given the low 16S rRNA gene similarities with other *Sulfolobales* species, we hypothesized that the strain also harbored distinct

genomic features in *Sulfolobales* species. Therefore, we determined the complete genome of *S. acidiphilus* HS-1<sup>T</sup> and compared it with other genomes in *Sulfolobales* species. Genomic analysis revealed that *S. acidiphilus* HS-1<sup>T</sup> has several distinguishing genomic features.

## Main text

### Methods

#### Organism information

The isolation and characterization of *S. acidiphilus* HS-1<sup>T</sup> representing a novel genus *Sulfodiicoccus* was reported previously [1, 6]. The phylogenetic position of *S. acidiphilus* based on 16S rRNA gene sequences is shown in Additional file 1: Figure S1. The general features of *S. acidiphilus* HS-1<sup>T</sup> are shown in Additional file 1: Table S1. HS-1<sup>T</sup> is an irregular cocci, non-motile,

\*Correspondence: kurosawa@soka.ac.jp

<sup>1</sup> Department of Science and Engineering for Sustainable Innovation, Faculty of Science and Engineering, Soka University, 1-236 Tangi-machi, Hachioji, Tokyo 192-8577, Japan

Full list of author information is available at the end of the article



thermoacidophilic archaeon. Optimal growth occurs at 65–70 °C and pH 3–3.5. The strain is obligately aerobic and can utilize the following organics as a sole carbon source: yeast extract, beef extract, casamino acids, peptone, tryptone, xylose, galactose, glucose, maltose, sucrose, raffinose, lactose, aspartic acid and glutamic acid. Chemolithotrophic growth does not occur when  $S^0$ ,  $FeS_2$ ,  $K_2S_4O_6$ ,  $Na_2S_2O_3$  or  $FeSO_4$  acts as an electron donor under aerobic conditions ( $O_2$  as an electron acceptor). The cells are regular to irregular cocci with a diameter of 0.8–1.5  $\mu m$  (Additional file 1: Figure S2).

#### Genomic DNA preparation, genome sequencing and assembly

HS-1<sup>T</sup> was cultivated in a 5 L glass bottle using ~4 L modified Brock's basal salt (MBS) medium [7], supplemented with yeast extract (1 g/L) and glucose (1 g/L) under aerobic conditions (65 °C, pH 3). Approximately 1 L of the culture in early exponential phase was centrifuged ( $OD_{600} \sim 0.1$ , 8000 $\times g$ , 4 °C, 10 min), and the supernatant was removed. DNA was extracted from the resulting cell pellets using the Genomic DNA Buffer Set (QIAGEN) and the Genomic-Tip 500/G (QIAGEN), according to the manufacturer's protocols. The quantity and purity of the extracted DNA was checked spectrophotometrically and through agarose gel electrophoresis. The genome sequencing of the *S. acidiphilus* strain HS-1<sup>T</sup> was performed by Macrogen Inc. (South Korea) using a PacBio RS II sequencer (Pacific Biosciences, Menlo Park, CA, US). De novo assembly was conducted using the Hierarchical Genome Assembly Process v.3.0 (<https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/HGAP>).

#### Genome annotation

Annotation of the protein coding genes and the COG (cluster of orthologous groups) assignments were performed using the on-line annotation server DFAST [8]. The tRNA and rRNA genes were identified using tRNAscan-SE [9, 10] and DFAST, respectively. Pseudogenes were identified using LAST [11] implemented in DFAST. Protein coding genes with Pfam domains [12] were searched using a CD-search program [13] with an e-value threshold of less than  $1e-2$  (database: Pfam v.30.0). Signal peptides were predicted using PRED-SIGNAL [14]. Transmembrane helices were predicted using TMHMM [15]. CRISPR repeats were detected using the CRISPR recognition tool CRT [16]. Genes in internal clusters were predicted using the CD-HIT Suite (sequence identity cut-off: 0.3, minimal alignment coverage for longer and shorter sequences: 0.7, other parameters: default) [17].

#### Search of replication origin

The replication origin (*oriC*) and origin recognition box (ORB) in the chromosome were predicted by Ori-finder 2 [18]. Another replication origin in the chromosome was manually searched. Repeat sequences in the predicted *oriC* region were searched by the REPuter program [19].

#### Reexamination of chemolithotrophic growth on hydrogen

The autotrophic growth of HS-1<sup>T</sup> on hydrogen was reexamined. MBS medium (10 mL, pH 3) was added to a glass test tube and the headspace was filled with  $H_2/CO_2$ /air (80:20:10, 120 kPa). A 50  $\mu L$  of active culture was inoculated into the test tube and incubation occurred at 65 °C for 2 weeks.

#### Genome project history

The complete genome sequence of HS-1<sup>T</sup> (=JCM 31740<sup>T</sup>=InaCC Ar79<sup>T</sup>) was deposited in GenBank under accession number AP018553. The raw data (PacBio reads) used for the assembly was deposited in the DNA Data Bank of Japan under accession number DRA008516. The Bioproject accession number is PRJDB6753. A summary of the genome project is provided in Additional file 1: Table S2.

#### Results and discussion

##### Genome properties

A complete circular genome sequence (2,353,189 bp) was successfully obtained from a total of 225,345 subreads (a total of 1,538,043,255 bp), and a plasmid was absent. The G+C content is 51.15 mol%, which is identical to the reported value of 52.0 mol% estimated by HPLC method [1]. The genome was predicted to contain a total of 2459 genes, of which 2411 code for proteins and 48 code for RNAs (rRNA: 3, tRNA: 45). The genome harbors each one copy of 5S, 16S and 23S rRNA genes. Genes of 16S and 23S rRNA are encoded in a gene cluster with a 191 bp spacer region, while 5S rRNA gene is found in different location. Among the 2411 protein coding genes, 1219 were assigned putative functions. A total of 837 and 244 genes could be assigned COG functional categories and pseudogenes, respectively (Additional file 1: Tables S3, S4, Figure S3). Three CRISPR repeat regions were detected (positions: 1664810–1679609, 1679646–1680862, and 1688612–1701897). No plasmids were detected. Other genomic statistics such as predicted Pfam domains, signal peptides, and transmembrane helices are summarized in Additional file 1: Table S3.

##### Insights from the genome sequence

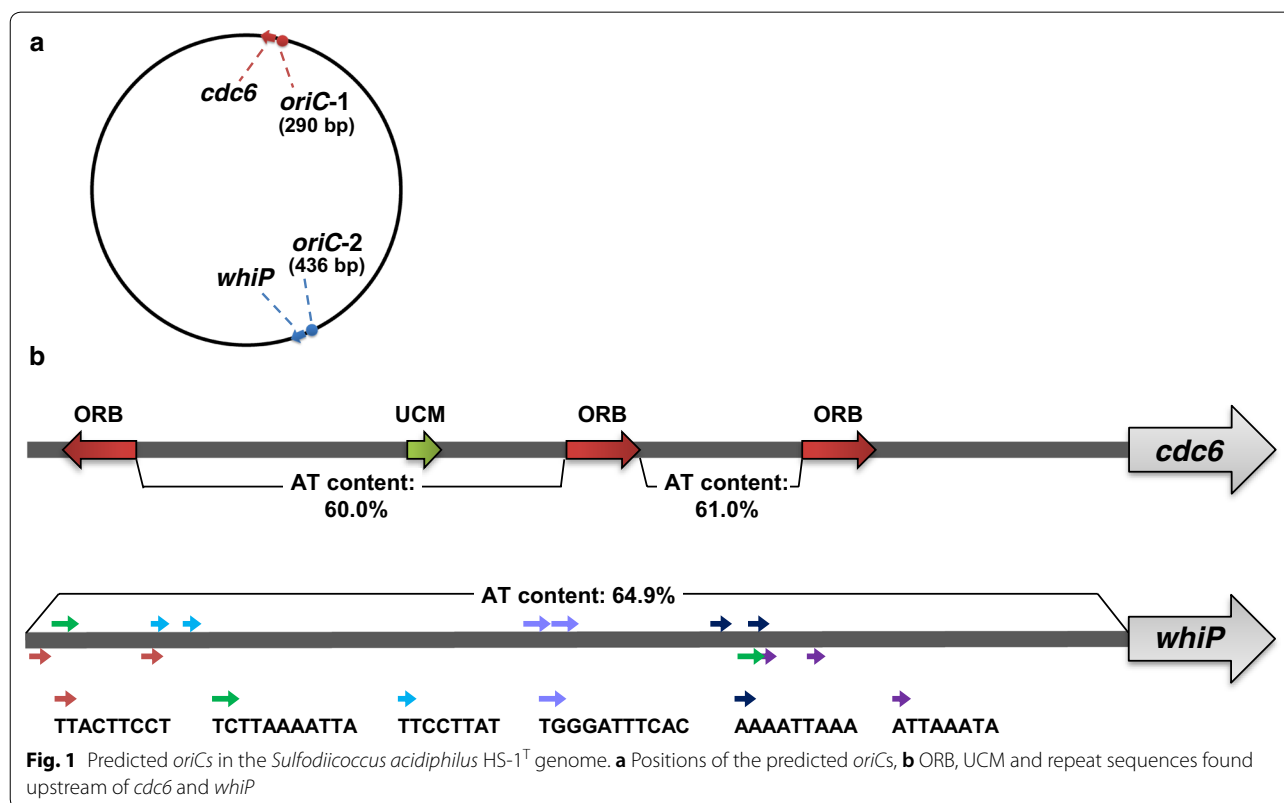
*Replication initiation genes and oriC* All previously identified species of the order *Sulfolobales* (whose com-

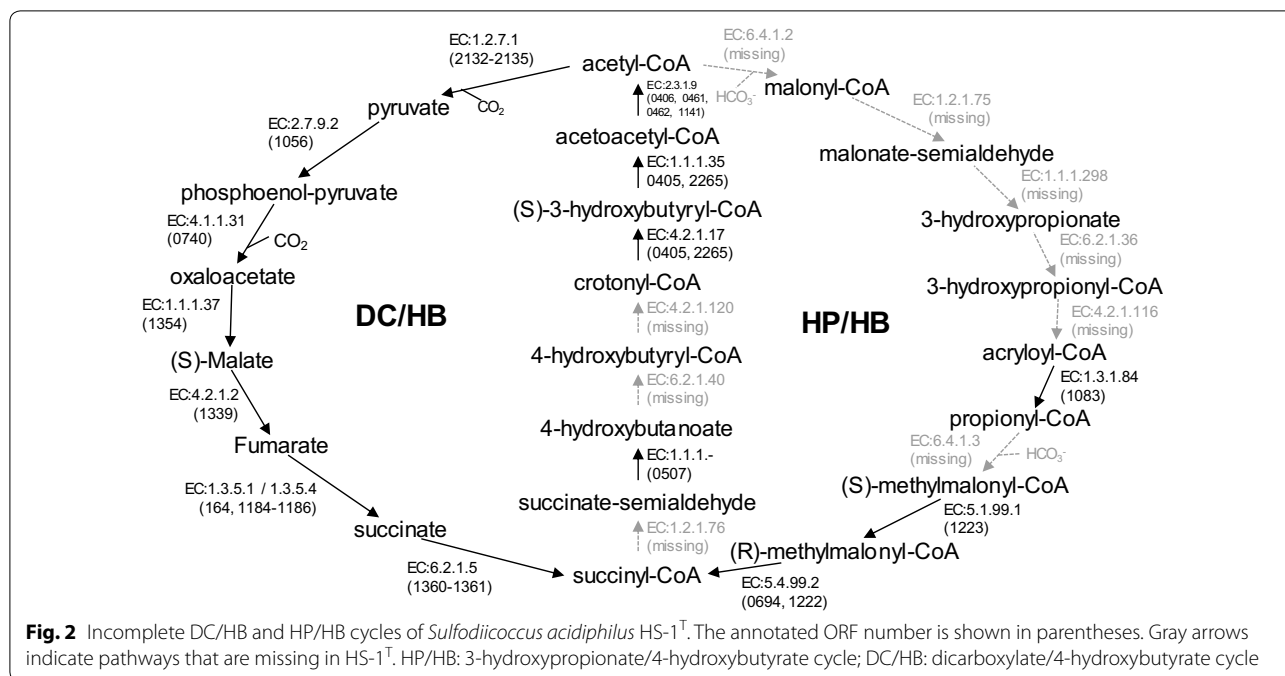
plete genomic sequences are available) have three copies of the *cdc6* gene (a cell division control gene, Additional file 1: Table S5). In contrast, only one copy of the *cdc6* gene (HS1genome0091) was found in the HS-1<sup>T</sup> genome (Fig. 1a). One replication origin (named *oriC-1*, position: 87,046–87,335) located directly upstream of the *cdc6* gene was predicted by the Ori-finder 2. The *oriC-1* had a relatively high content of adenine and thymine residues (A/T rich) (56.55%) and contained three ORBs as typical archaeal genomes with the following sequence: ACCCCCT CTGTTTCCACTGGA [18, 20–22] (Fig. 1b). A previously reported uncharacterized motif (UCM) that exists in the *oriC* regions of several *Crenarchaea* [23] has also been located in HS-1<sup>T</sup> *oriC-1*. These facts indicate that *oriC-1* is a DNA replication origin. Another replication initiation-related gene, *whiP* (HS1genome1070) [23], was found on the opposite side of the *cdc6* (Fig. 1a), however, neither *oriC* nor ORB sequences were found around *whiP* by Ori-finder 2. We then manually searched for putative *oriC* regions, and an A/T rich intergenic region (64.91%, position: 1,016,172–1,016,607, named as *oriC-2*) was found directly upstream of *whiP*. Further attempts to search for repeat sequences in *oriC-2* using the REPuter program [19] confirmed the presence of 12 direct repeats (8–11 bp, Fig. 1b). The role of the direct repeats is unclear, although they may be involved in the recognition of the replication

origin for the DNA initiation protein WhiP. Perhaps the direct repeats alternatively function as an ORB. Further experiments such as DNaseI footprint analyses [23, 24] are required to know the function.

#### Pathways involved in autotrophic growth

Two CO<sub>2</sub> fixation pathways have been reported in the thermophilic autotroph *Crenarchaeota*, namely the 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle and the dicarboxylate/4-hydroxybutyrate (DC/HB) cycle [25]. Most species in the order *Sulfolobales* are thought to possess the HP/HB cycle but not the DC/HB cycle [25–27]. Both cycles are absent in HS-1<sup>T</sup> (Fig. 2). The other known autotrophic pathways (Calvin–Bassham–Benson cycle, reductive citric acid cycle (Amon–Buchanan cycle), reductive acetyl–CoA pathway (Wood–Ljungdahl pathway) and 3-hydroxypropionate cycle) [28] have not been identified in HS-1<sup>T</sup>. These observations are consistent with the incapability of chemolithoautotrophic growth of HS-1<sup>T</sup> [1]. In the previous paper, we mentioned the capability of chemolithoautotrophic growth of HS-1<sup>T</sup> using hydrogen as an electron donor. Based on the genomic information, we carefully reexamined the capability of chemolithoautotrophic growth with serial inoculation using the chemolithoautotrophic medium, and no growth occurred after 2nd inoculation. Thus, we

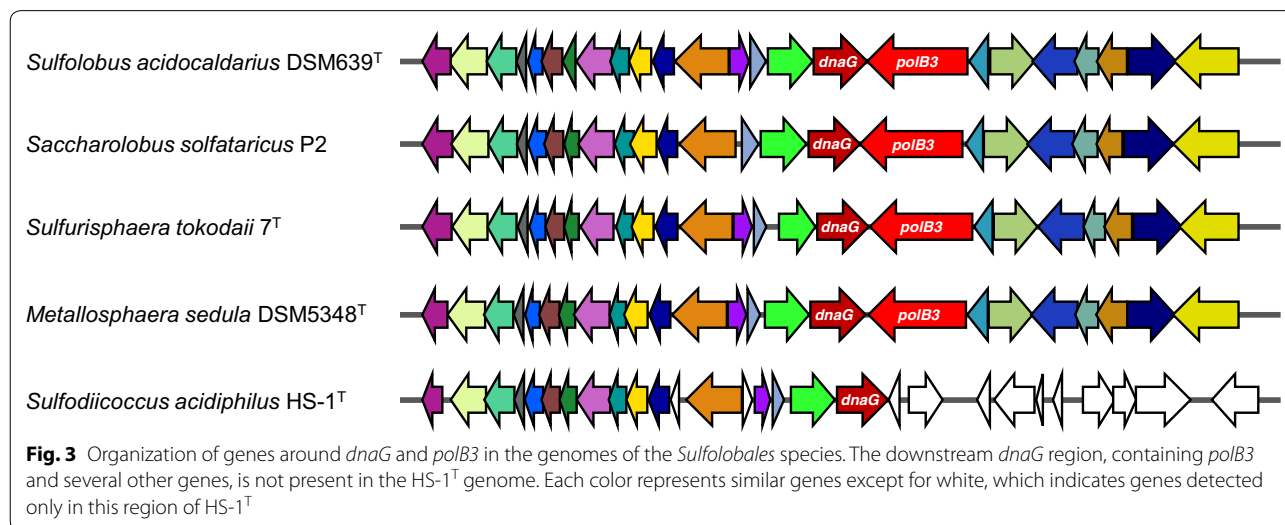




revise our previous description regarding the autotrophic growth of HS-1<sup>T</sup> on hydrogen, to “HS-1<sup>T</sup> does not grow on hydrogen autotrophically.”

**Lack of polB3 in the genome of HS-1<sup>T</sup>** Three groups of family B DNA polymerases (PolB1, PolB2, and PolB3) have been associated with archaea: PolB1 is only distributed in the superphylum that includes *Thaumarchaeota*, *Aigarchaeota*, *Crenarchaeota* and *Korarchaeota* (also known as TACK superphylum); PolB2 is patchily distributed in most of the archaeal lineages; and PolB3 is dis-

tributed in all archaea except for *Thaumarchaeota* [29]. Surprisingly, the HS-1<sup>T</sup> genome lacks *polB3*, although all the crenarchaeal genomes reported before harbor the gene. In the order *Sulfolobales*, *polB3* is located on the downstream region of *dnaG*, with a high genomic synteny around the *dnaG* sequence among *Sulfolobales* species, while the genome structure of the downstream region of *dnaG* of HS-1<sup>T</sup> is different from those in the genomes of other *Sulfolobales* species (Fig. 3). Since all the other species in the order *Sulfolobales* have a *polB3* downstream of a *dnaG*, HS-1<sup>T</sup> may have lost *polB3* and the downstream



region of *dnaG* during the course of its evolution. The roles of PolB1, PolB2 and PolB3 in *Saccharolobus solfataricus* (synonym: *Sulfolobus solfataricus*), a model organism of the order *Sulfolobales*, were previously investigated in vivo by Choi et al. [30]. The authors showed that PolB1 was catalytically much more efficient and processive than PolB2 and PolB3, suggesting that PolB1 plays a catalytic role as the main replicative DNA polymerase. They also suggested that PolB2 and PolB3 have limited catalytic roles in translesion DNA synthesis and may not be involved in chromosomal DNA replication [30]. The lack of a *polB3* gene in *S. acidiphilus* HS-1<sup>T</sup> indicates that PolB3 is not essential for either DNA replication or translesion DNA synthesis in *Sulfolobales* or the phylum *Crenarchaeota*.

### Limitations

This study focused on noteworthy genomic features of *S. acidiphilus* HS-1<sup>T</sup> that are distinct from other *Sulfolobales* species. Although our genomic analyses revealed some exceptions of genomic features in *Sulfolobales*, further molecular biology assessment and biochemical analyses are needed to resolve the issues raised in this manuscript.

### Additional file

**Additional file 1: Table S1.** Classification and general features of *Sulfolobus acidiphilus* strain HS-1<sup>T</sup>. **Table S2.** Project information. **Table S3.** Genome statistics. **Table S4.** Number of genes associated with general COG functional categories. **Table S5.** Number of *cdc6* and *whiP* genes in the order *Sulfolobales*. **Figure S1.** The maximum likelihood phylogenetic tree of the order *Sulfolobales* based on the 16S rRNA gene. **Figure S2.** Scanning electron micrographs of *Sulfolobus acidiphilus* HS-1<sup>T</sup>. **Figure S3.** A circular map of the *S. acidiphilus* strain HS-1<sup>T</sup> genome.

### Abbreviations

HPLC: high performance liquid chromatography; HP/HB: 3-hydroxypropionate/4-hydroxybutyrate; DC/HB: dicarboxylate/4-hydroxybutyrate; ORB: origin recognition box; UCM: uncharacterized motif.

### Acknowledgements

Not applicable.

### Authors' contributions

HDS and NK conceived and designed this project, also wrote the manuscript. HDS conducted wet-lab work, electron microscopy, and bioinformatics analysis. Both authors read and approved the final manuscript.

### Funding

Not applicable.

### Availability of data and materials

The complete genome sequence of HS-1<sup>T</sup> is available in GenBank repository, <https://www.ncbi.nlm.nih.gov/genbank/>, under accession number AP018553 (the Bioproject accession number is PRJDB6753). The raw data (PacBio reads) used for the assembly is available in the DNA Data Bank of Japan repository, <https://www.ddbj.nig.ac.jp/>, under accession number DRA008516.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup> Department of Science and Engineering for Sustainable Innovation, Faculty of Science and Engineering, Soka University, 1-236 Tangi-machi, Hachioji, Tokyo 192-8577, Japan. <sup>2</sup> Japan Society for the Promotion of Science, Chiyoda-ku, Tokyo 102-8471, Japan.

Received: 21 February 2019 Accepted: 17 July 2019

Published online: 22 July 2019

### References

- Sakai HD, Kurosawa N. *Sulfolobus acidiphilus* gen. nov., sp. nov., a sulfur-inhibited thermoacidophilic archaeon belonging to the order *Sulfolobales* isolated from a terrestrial acidic hot spring. *Int J Syst Evol Microbiol*. 2017;67:1880–6.
- Satoh T, Watanabe K, Yamamoto H, Yamamoto S, Kurosawa N. Archaeal community structures in the solfataric acidic hot springs with different temperatures and elemental compositions. *Archaea*. 2013. <https://doi.org/10.1155/2013/723871>.
- Kato S, Itoh T, Yamagishi A. Archaeal diversity in a terrestrial acidic spring field revealed by a novel PCR primer targeting archaeal 16S rRNA genes. *FEMS Microbiol Lett*. 2011;319:34–43.
- Perevalova AA, Kolganova TV, Birkeland NK, Schleper C, Bonch-Osmolovskaya EA, Lebedinsky AV. Distribution of Crenarchaeota representatives in terrestrial hot springs of Russia and Iceland. *Appl Environ Microbiol*. 2008;74:7620–8.
- Urbietta MS, Toril EG, Giaveno MA, Bazán AA, Donati ER. Archaeal and bacterial diversity in five different hydrothermal ponds in the Copahue region in Argentina. *Syst Appl Microbiol*. 2014;37:429–41.
- Sakai HD, Kurosawa N. Exploration and isolation of novel thermophiles in frozen enrichment cultures derived from a terrestrial acidic hot spring. *Extremophiles*. 2016;20:207–14.
- Kurosawa N, Itoh YH, Iwai T, Sugai A, Uda I, Kimura N, et al. *Sulfurisphaera ohwakuensis* gen. nov., sp. nov., a novel extremely thermophilic acidophile of the order *Sulfolobales*. *Int J Syst Bacteriol*. 1998;48:451–6.
- Tanizawa Y, Fujisawa T, Nakamura Y. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics*. 2017;34:1037–9.
- Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25:955–64.
- Cros MJ, Monte AD, Mariette J, Bardou P, Grenier-Boley B, Gautheret D, et al. RNAspace.org: an integrated environment for the prediction, annotation, and analysis of ncRNA. *RNA*. 2011;17:1947–56.
- Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. *Genome Res*. 2011;21:487–93.
- Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44:D279–85.
- Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a Conserved domain database for the functional annotation of proteins. *Nucleic Acids Res*. 2011;39:225–9.
- Bagos PG, Tsigos KD, Plessas SK, Liakopoulos TD, Hamodrakas SJ. Prediction of signal peptides in archaea. *Protein Eng Des Sel*. 2009;22:27–35.
- Krogh A, Rn Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol*. 2001;305:567–80.
- Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyripides NC, et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinform*. 2007;8:1–8.



17. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*. 2010;26:680–2.
18. Luo H, Zhang CT, Gao F. Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes. *Front Microbiol*. 2014;5:1–6.
19. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 2001;29:4633–42.
20. Ausiannikava D, Allers T. Diversity of DNA replication in the archaea. *Genes*. 2017;8:56.
21. Manzella MP, Holmes DE, Rocheleau JM, Chung A, Reguera G, Kashefi K. The complete genome sequence and emendation of the hyperthermophilic, obligate iron-reducing archaeon “*Geoglobus ahangari*” strain 234<sup>T</sup>. *Std Genomic Sci*. 2015;10:1–19.
22. Wu Z, Liu J, Yang H, Xiang H. DNA replication origins in archaea. *Front Microbiol*. 2014;5:1–7.
23. Robinson NP, Bell SD. Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc Natl Acad Sci*. 2007;104:5806–11.
24. Robinson NP, Dionne I, Lundgren M, Marsh VL, Bernander R, Bell SD. Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell*. 2004;116:25–38.
25. Berg IA, Ramos-Vera WH, Petri A, Huber H, Fuchs G. Study of the distribution of autotrophic CO<sub>2</sub> fixation cycles in Crenarchaeota. *Microbiology*. 2010;156:256–69.
26. Dai X, Wang H, Zhang Z, Li K, Zhang X, Mora-López M, et al. Genome sequencing of *Sulfolobus* sp. A20 from costa rica and comparative analyses of the putative pathways of carbon, nitrogen, and sulfur metabolism in various sulfolobus strains. *Front Microbiol*. 2016;7:1902.
27. Urbietta MS, Rascovan N, Vázquez MP, Donati E. Genome analysis of the thermoacidophilic archaeon *Acidianus copahuensis* focusing on the metabolisms associated to biomining activities. *BMC Genomics*. 2017;18:1–14.
28. Berg IA, Kockelkorn D, Buckel W, Fuchs G. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science*. 2007;318:1782–7.
29. Makarova KS, Krupovic M, Koonin EV. Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. *Front Microbiol*. 2014;5:1–10.
30. Choi JY, Eoff RL, Pence MG, Wang J, Martin MV, Kim EJ, et al. Roles of the four DNA polymerases of the crenarchaeon *Sulfolobus solfataricus* and accessory proteins in DNA replication. *J Biol Chem*. 2011;286:31180–93.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

