Genome sequencing and analysis of the genome of *Acidithiobacillus ferrooxidans* strain DLC-5, a heavy metal resistant strain from acid mine drainage in northeast China

Peng Chen[1], Ruixiang Xu[1], Zheng Yan[1], Lei Yan[3], Zhengrong Wu[1], Yan Wei[2], Wenbin Zhao[1], Hongyu Li[1, 2, *]

[1] School of Pharmacy, Lanzhou University, Donggang West Road No. 199, Lanzhou, 730020, PR China

[2] Gansu Key Laboratory of Biomonitoring and Bioremediation for Environmental Pollution, Institute of Microbiology, School of Life Sciences, Lanzhou University, Tianshui Road No. 222, Lanzhou, 730000, PR China

[3] College of Life Science and Technology, Heilongjiang Bayi Agricultural University, Daqing, 163319, PR China

* Corresponding author. Tel. & fax: +86 931 8915686.

E-mail address: chenpeng@lzu.edu.cn (P. Chen) or lihy@lzu.edu.cn (H. Li).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:

ABSTRACT

Acidithiobacillus ferrooxidans (*A. ferrooxidans*) is a gram-negative, extremely acidophilic, mesophilic, chemolithotrophic bacterium and the most well-studied acidophilic organism which is usually found in acid environments such as acid mine drainage. The draft genome sequence of *A. ferrooxidans* ATCC 23270 was first reported in 2000, fourteen years ago. Here we describe the features of this organism, together with the draft genome sequence, and annotation. This is the draft genome sequence from the *A. ferrooxidans*, and the 3,142,890 bp long single replicon genome with its 32,719 protein-coding and 64 RNA genes is a part of the Genomic Encyclopedia of Bacteria and Bacterial project.

Keywords: Acidophilic bacteria, *Acidithiobacillus ferrooxidans*, Bioleaching, Extremophiles, Genome.

1. Introduction

One of the most studied microorganisms involved in biohydrometallurgy is *Acidithiobacillus ferrooxidans* (*A. ferrooxidans*) (Beijerinck 1904; Kelly & Wood 2000; Skerman et al. 1980). Due to its bioleaching capabilities, it is an important member of microbial consortia involved in the industrial recovery of metal under mesophilic conditions (bioleaching or biomining). Recently, *A. ferrooxidans* has played important roles in bioleaching and harnesse environmental contamination (Chen et al. 2011; Yan et al. 2010; Zhao et al. 2015). Like in other acidophilic iron-oxidizing bacterium, it grows optimally at about 35°C in 9K inorganic medium at extremely low pH (pH 1.0–2.0) and fixes both carbon and nitrogen from the atmosphere (Quatrini et al. 2005). *A. ferrooxidans* derives energy from oxidizing reduced sulfur compounds and Fe^{2+} ions to form sulfate and Fe^{3+}, respectively (Chen et al. 2011). To date, two genome sequences of *A. ferrooxidans* strains ATCC 23270 and ATCC 53993 are available in the public databases (Orellana & Jerez 2011; Valdes et al. 2008). These genomic data are useful for the experimental identification of unique proteins or estimation of the phylogenetic relationship among the related strains. Strain DLC-5 (CCTCC-M 2014362) is the type strain of *A. ferrooxidans*, isolated from Wudalianchi in Heihe of Heilongjiang Province, and the type species of the genus *Acidithiobacillus*, which currently contains five species. Here we present a summary classification and a set of features for *A. ferrooxidans* DLC-5, together with the description of the draft genomic sequencing and annotation.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:

Organism Information

Classification and features

A representative genomic 16S rRNA sequence of *A. ferrooxidans* DLC-5 was compared to the 16S rRNA sequences of known *Acidithiobacillus* genus type strains. The 16S rRNA gene sequence identities between *A. ferrooxidans* DLC-5 and all other type strains of species *A. ferrooxidans* were 97.0-99.0%. *A. ferrooxidans* species exhibiting the highest sequence identities to DLC-5 were *A. ferrooxidans* ATCC 23270 and A. *ferrooxidans* ATCC 53993. Figure 1 shows the phylogenetic relationships of *A. ferrooxidans* DLC-5 to other *A. ferrooxidans* species in a 16S rRNA based tree. All the type strains and six strains of *A. ferrooxidans* including DLC-5 were used for the analysis. All six *A. ferrooxidans* strains are closely related to each other, and the 16S rRNA sequences have 100% identities. *A. ferrooxidans* DLC-5 is a extremely acidophilic (pH 1.0–2.0), mesophilic (temperature optimum 30-35 ^{o}C) microorganisms. The image of *A. ferrooxidans* DLC-5 cells grown on 9K medium with ferrous sulfate (44.69 g/L) are shown in Figure 2. The characteristic features are shown in Table 1.

Genome sequencing information

Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position, and is part of the G*enomic Encyclopedia of Bacteria and rchaea* project. The genome project is deposited in the Genomes OnLine Database (Pagani et al. 2012) and the draft genome sequence in GenBank (JNNH00000000.1). Sequencing, finishing and annotation were performed by Shanghai Majorbio Bio-pharm Technology Co. , Ltd. (Majorbio). A summary of the project information is shown in Table 2.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:

Growth conditions and genomic DNA preparation

A. ferrooxidans strain, DLC-5 was grown in 9K medium at 35°C. DNA was isolated from 1.0-1.5 g of cell paste using Qiagen Genomic 500 DNA Kit (Qiagen, Hil-den, Germany) with a modified protocol, st/FT, for cell lysis, as described in Valdes *et al*. (Valdes et al. 2009).

Genome sequencing and assembly

Draft genome sequence of *A. ferrooxidans* type strain DLC-5 was obtained in Illumina Hiseq2000 sequencing technology by Shanghai Majorbio Bio-pharm Technology Co. , Ltd. (Shanghai, China), using the Short Oligonucle-otides Alignment Program (SOAP) denovo alignment tool (http://soap.genomics.org.cn/) processes reads assemble. A library containing 300-bp inserts was constructed. Altogether, 6,372,268 paired reads; 398,580 single reads; total 1,079,535,272 bp bases with average coverage of 221.1 ×. Reads were filtered to remove adapter sequences, low-quality bases (Phred score, ＜ 20), removing the 5'end that contains the bases of it is not A, G, C, T before shearing, remove reads with the containing 10% of N, giving up adapter and small fragments of length less than 25 bp after qualitative pruning. The reads were assembled into 881 contigs (＞ 1,000bp; Contig N50, 102 bp; Contig N90, 569 bp) and 573 scaffolds (＞ 1,000bp; Scaffold N50, 71 bp; Scaffold N90, 333 bp).

Genome annotation

Using the Glimmer 3.0 (http://www.cbcb.umd.edu/software/glimmer/) processes gene prediction. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, Nr, string, GO, COG, KEGG and go databases.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:

Genome Properties

The genome includes two plasmids, for a total size of 3,142,890 bp, with one circular chromosome of 1,832,305 bp (58.3% GC content). For the main chromosome, 4,299 bp genes were predicted, 4,131 bp of which are protein-coding genes. 3,250 bp of protein coding genes were assigned to a putative function with the remaining annotated as hypothetical proteins. The properties and the statistics of the genome are summarized in Tables 3. The distribution of genes into COGs functional categories is presented in Tables 4, and a cellular overview diagram is presented in Figure 4.

Conclusions

Extremely acidophilic bacteria and archaea with special emphasis on bioleaching microorganisms are widely distributed in the extreme acidic environment. In this study we analyzed the genome sequence of *A. ferrooxidans* DLC-5, which was isolated from acid mine drainage in Northeast China. Genome analysis of this strain revealed the presence of key functional characteristics. It may contribute to further studies on important process for bioleaching and acid mine drainage production, such as biofilm formation, energy resources utilization and quorum sensing that could play a role in a possible interrelationship of bioleaching heaps and other acidic environments. In addition, combining with genomes of other members in *Acidithiobacillus*, will make an important advance in understanding of the ecological roles that *Acidithiobacillus* species play in those acidic environments and their relationships with other extremely acidophilic microorganisms.

Conflict of interest

There is no conflict of authorship and both the authors approve the final version of the manuscript.

Acknowledgements

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:

This work was supported by Gansu Province Science Foundation for Distinguished Young Scholars (Grant No. 1308RJDA014), Technology Program of Gansu Province (Grant No. 1604FKCA110), Technology Program of Lanzhou City (Grant No. 2015-3-142, Grant No. 2015-3-97 and Grant No. 2015-3-93).

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

We have submitted the whole genome sequence of *A. ferrooxidans* DLC-5 to NCBI:

NCBI Reference Sequence: NZ_ JNNH00000000.1

GenBank: JNNH00000000.1

GI: 666873905.

References

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, and Eppig JT. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet* 25:25-29.

Beijerinck MW. 1904. Phénomènes de réduction produits par les microbes. *Arch Neerl Sci Ser* 9:131-157.

Chen P, Yan L, Leng FF, Nan WB, Yue XX, Zheng YN, Feng N, and Li HY. 2011. Bioleaching of realgar by *Acidithiobacillus ferrooxidans* using ferrous iron and elemental sulfur as the sole and mixed energy sources. *Bioresour Technol* 102:3260-3267. http://dx.doi.org/10.1016/j.biortech.2010.11.059

Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, and Angiuoli SV. 2008. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 26:541-547.

Garrity A. 2005. Validation of publication of new names and new combinations previously effectively published outside the IJSEM. *Int J Syst Evol Microbiol* 55:2235-2238.

Garrity GM, Bell JA, and Lilburn T. 2005a. Class III. *Gammaproteobacteria* class. nov. . In: Garrity GM, Brenner DJ, Krieg NR, and Staley JT, eds. *Bergey's Manual of Systematic Bacteriology*. New York: Springer, 1.

Garrity GM, Bell JA, and Lilburn T. 2005b. Family I. *Acidithiobacillaceae* fam. nov. . In: Brenner DJ, Krieg NR, Staley JT, and Garrity GM, eds. *Bergey's Manual of Systematic Bacteriology*. New York: Springer.

Garrity GM, Bell JA, and Lilburn T. 2005c. Order II. *Acidithiobacillales* ord. nov. In: Garrity GM, Brenner DJ, Krieg NR, and Staley JT, eds. *Bergey's Manual of Systematic Bacteriology*: Springer, 60.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:

Garrity GM, Bell JA, and Lilburn T. 2005d. Phylum XIV. *Proteobacteria phyl*. nov. In: Garrity GM, Brenner DJ, Krieg NR, and Staley JT, eds. *Bergey's Manual of Systematic Bacteriology*. New York: Springer, 1.

Kelly DP, and Wood AP. 2000. Reclassification of some species of *Thiobacillus* to the newly designated genera *Acidithiobacillus* gen. nov., *Halothiobacillus* gen. nov. and *Thermithiobacillus* gen. nov. *Int J Syst Evol Microbiol* 50:511-516.

Li YQ, Wan DS, Huang SS, Leng FF, Yan L, Ni YQ, and Li HY. 2010. Type IV pili of *Acidithiobacillus ferrooxidans* are necessary for sliding, twitching motility, and adherence. *Curr Microbiol* 60:17-24. 10.1007/s00284-009-9494-8

Orellana LH, and Jerez CA. 2011. A genomic island provides *Acidithiobacillus ferrooxidans* ATCC 53993 additional copper resistance: a possible competitive advantage. *Appl Microbiol Biotechnol* 92:761-767. http://dx.doi.org/10.1007/s00253-011-3494-x

Pagani I, Liolios K, Jansson J, Chen I-MA, Smirnova T, Nosrat B, Markowitz VM, and Kyrpides NC. 2012. The Genomes OnLine Database (GOLD) v. 4: status of genomic and metagenomic projects and their associated metadata. *Nucleic acids research* 40:D571-D579.

Quatrini R, Jedlicki E, and Holmes DS. 2005. Genomic insights into the iron uptake mechanisms of the biomining microorganism *Acidithiobacillus ferrooxidans*. *J Ind Microbiol Biotechnol* 32:606-614. http://dx.doi.org/10.1007/s10295-005-0233-2

Skerman VBD, McGowan V, and Sneath PHA. 1980. Approved lists of bacterial names. *Int J Syst Bacteriol* 30:225-230.

Valdes J, Pedroso I, Quatrini R, Dodson RJ, Tettelin H, Blake R, II, Eisen JA, and Holmes DS. 2008. *Acidithiobacillus ferrooxidans* metabolism: from genome

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:

sequence to industrial applications. *BMC Genomics* 9:597. http://dx.doi.org/10.1186/1471-2164-9-597

Valdes J, Quatrini R, Hallberg K, Dopson M, Valenzuela PDT, and Holmes DS. 2009. Draft genome sequence of the extremely acidophilic bacterium *Acidithiobacillus caldus* ATCC 51756 reveals metabolic versatility in the genus *Acidithiobacillus*. *J Bacteriol* 191:5877.

Woese CR, Kandler O, and Wheelis ML. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 87:4576-4579.

Yan L, Yin H, Zhang S, Duan JG, Li Y, Chen P, and Li HY. 2010. Organoarsenic resistance and bioremoval of *Acidithiobacillus ferrooxidans*. *Bioresour Technol* 101:6572-6575. http://dx.doi.org/10.1016/j.biortech.2010.03.065

Zhao Y, Chen P, Nan W, Zhi D, Liu R, and Li H. 2015. The use of (5Z)-4-bromo-5-(bromomethylene)-2(5H)-furanone for controlling acid mine drainage through the inhibition of *Acidithiobacillus ferrooxidans* biofilm formation. *Bioresour Technol* 186:52-57. http://dx.doi.org/10.1016/j.biortech.2015.02.017

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:

Table 1 Classification and general features of *Genusspecies* strain designation[T] (Field et al. 2008)

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain *Bacteria*	TAS (Skerman et al. 1980; Woese et al. 1990)
		Phylum *Proteobacteria*	TAS (Garrity et al. 2005d)
		Class *Gammaproteobacteria*	TAS (Garrity 2005; Garrity et al. 2005a)
		Order *Acidithiobacillales*	TAS (Garrity 2005; Garrity et al. 2005c)
		Family *Acidithiobacillaceae*	TAS (Garrity 2005; Garrity et al. 2005b)
		Genus *Acidithiobacillus*	TAS (Kelly & Wood 2000)
		Species *Acidithiobacillus ferrooxidans*	TAS (Kelly & Wood 2000)
		Strain: DLC-5 (CCTCC-M 2014362)	IDA
	Gram stain	*Negative*	NAS
	Cell shape	*Rod*	NAS
	Motility	*Motile*	TAS (Li et al. 2010)
	Sporulation	*Not reported*	NAS
	Temperature range	*25-35°C*	IDA
	Optimum temperature	*30°C*	IDA
	pH range; Optimum	*1.5–3.5; 2.0*	IDA
	Carbon source	*Atmosphere*	IDA
MIGS-6	Habitat	*Extremely acidophilic*	IDA
MIGS-6.3	Salinity	*0.5% NaCl (w/v)*	IDA
MIGS-22	Oxygen requirement	*Aerobic*	NAS
MIGS-15	Biotic relationship	*free-living*	IDA
MIGS-14	Pathogenicity	*Non-pathogen*	NAS
MIGS-4	Geographic location	*China/ Heilongjiang*	IDA
MIGS-5	Sample collection	*2013*	IDA
MIGS-4.1	Latitude	*48.52*	IDA
MIGS-4.2	Longitude	*126.2*	IDA
MIGS-4.4	Altitude	*Not reported*	NAS

[a] Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project (Ashburner et al. 2000)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:

Table 2 Project information.

MIGS ID	Property	Term
MIGS 31	Finishing quality	Finished
MIGS-28	Libraries used	300bp-500 bp Illumina PE library
MIGS 29	Sequencing platforms	Illumina
MIGS 31.2	Fold coverage	221.1 ×
MIGS 30	Assemblers	SOAPdenovo V1.05
MIGS 32	Gene calling method	Glimmer 3.0
	Locus Tag	27 OCT, 2014
	Genbank ID	JNNH00000000.1
	GenBank Date of Release	27 OCT, 2014
	GOLD ID	Gi0074489
	BIOPROJECT	CCTCC-M 2014362
MIGS 13	Source Material Identifier	Biohydrometallurgy, Environmental
	Project relevance	Finished

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:

Table 3 Genome statistics.

Attribute	Value	% of Total
Genome size (bp)	3,142,890	100.0
DNA coding (bp)	2,816,029	89.6
DNA G+C (bp)	1,832,305	57.63
DNA scaffolds	1,333	31.0
Total genes	4,299	100.0
Protein coding genes	4,131	96.1
RNA genes	168	3.9
Pseudo genes	0	
Genes in internal clusters	0	
Genes with function prediction	3,312	77.0
Genes assigned to COGs	3,250	75.6
Genes with Pfam domains	3,486	81.1
Genes with signal peptides	315	7.3
Genes with transmembrane helices	826	19.2
CRISPR repeats	0	

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:

Table 4 Number of genes associated with general COG functional categories.

Code	Value	%age	Description
J	118	5.6	Translation, ribosomal structure and biogenesis
A	1	0.0	RNA processing and modification
K	112	5.4	Transcription
L	157	7.5	Replication, recombination and repair
B	1	0.0	Chromatin structure and dynamics
D	29	1.4	Cell cycle control, Cell division, chromosome partitioning
V	51	2.4	Defense mechanisms
T	59	2.8	Signal transduction mechanisms
M	141	6.7	Cell wall/membrane biogenesis
N	30	1.4	Cell motility
U	76	3.6	Intracellular trafficking and secretion
O	96	4.6	Posttranslational modification, protein turnover, chaperones
C	159	7.6	Energy production and conversion
G	93	4.4	Carbohydrate transport and metabolism
E	146	7.0	Amino acid transport and metabolism
F	41	2.0	Nucleotide transport and metabolism
H	89	4.3	Coenzyme transport and metabolism
I	59	2.8	Lipid transport and metabolism
P	125	6.0	Inorganic ion transport and metabolism
Q	43	2.1	Secondary metabolites biosynthesis, transport and catabolism
R	212	10.1	General function prediction only
S	134	6.4	Function unknown
-	119	5.7	Not in COGs

The total is based on the total number of protein coding genes in the genome.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:

Figure legends

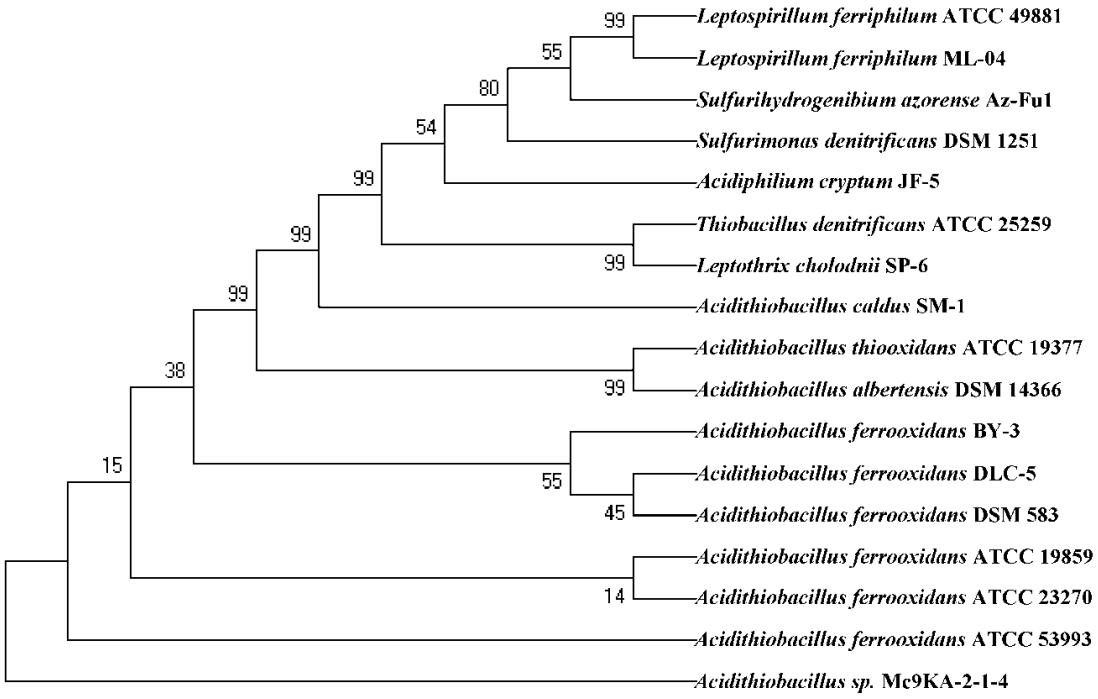
Figure 1. Phylogenetic tree highlighting the position of *A. ferrooxidans* DLC-5 relative to selected *A. ferrooxidans* species. To construct the phylogenetic tree, these sequences were collected and nucleotide sequence alignment was carried out using CLUSTALW. We used the MEGA 5.0 package to generate phylogenetic trees based on 16S rRNA genes with the neighbor-joining (NJ) approach.

Figure 2. Cell morphology of *A. ferrooxidans* DLC-5

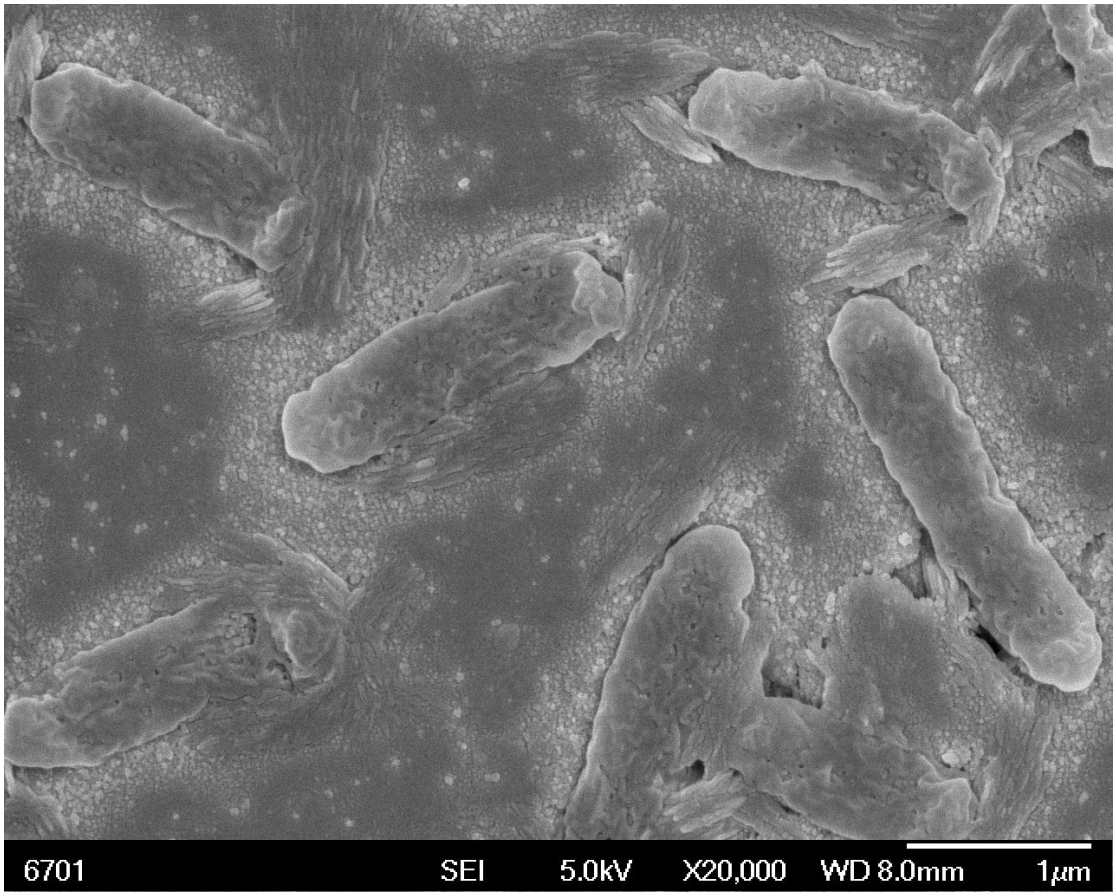
Figure 3. Graphical circular map of the chromosome of *A. ferrooxidans* DLC-5

From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

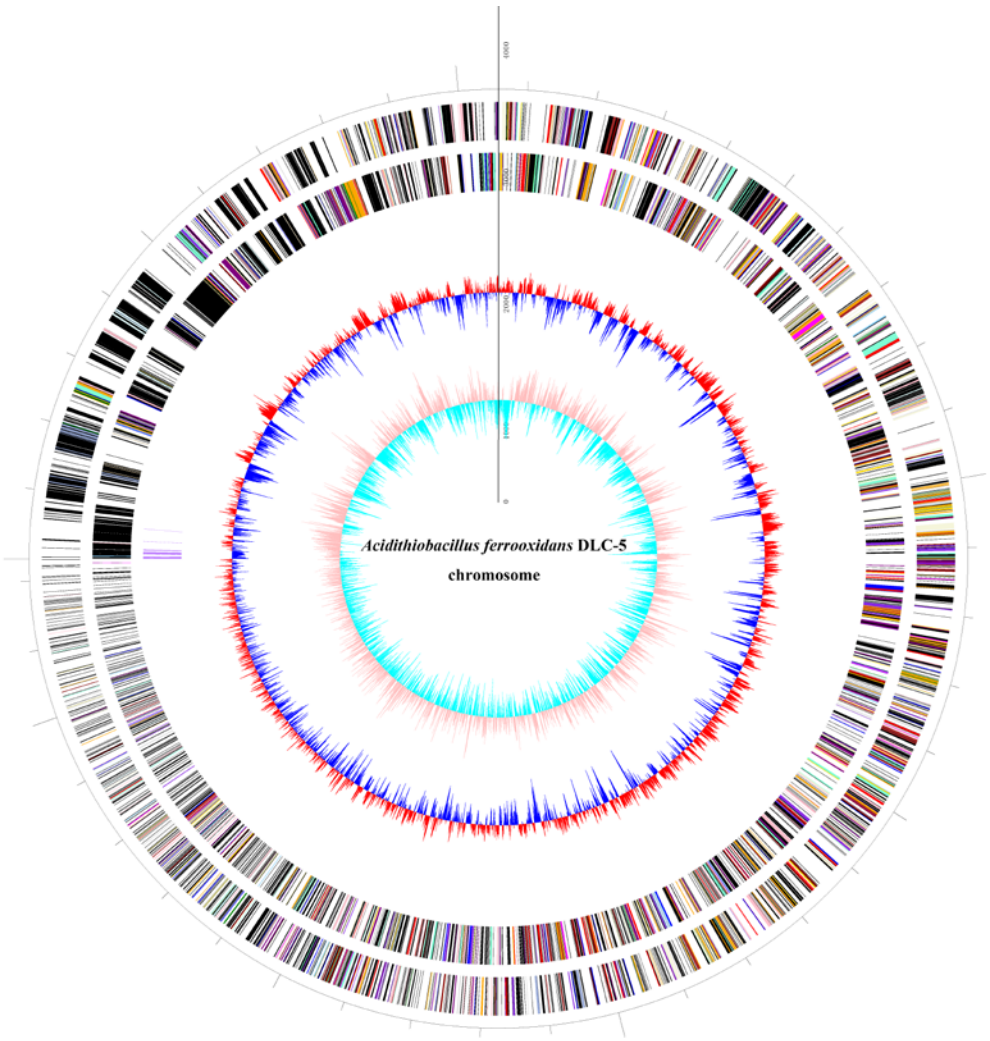
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ:



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2620v1 | CC BY 4.0 Open Access | rec: 6 Dec 2016, publ: