The Journal of Microbiology (2011) Vol. 49, No. 6, pp. 890-901

DOI 10.1007/s12275-011-1099-9

Complete Genome of *Leptospirillum ferriphilum* ML-04 Provides Insight into Its Physiology and Environmental Adaptation[§]

Shuang Mi[1], Jian Song[1,2], Jianqun Lin[1]*, Yuanyuan Che[1], Huajun Zheng[3], and Jianqiang Lin[1]

[1]*State Key Laboratory of Microbial Technology, Shandong University, Jinan 250100, P. R. China*
[2]*Zibo Vocational Institute, Zibo 255314, P. R. China*
[3]*Chinese National Human Genome Center at Shanghai, Shanghai 201203, P. R. China*

(Received March 2, 2011 / Accepted July 27, 2011)

Leptospirillum ferriphilum **has been identified as the dominant, moderately thermophilic, bioleaching microorganism in bioleaching processes. It is an acidic and chemolithoautrophic bacterium that gains electrons from ferrous iron oxidation for energy production and cell growth. Genetic information about this microorganism has been limited until now, which has hindered its further exploration. In this study, the complete genome of *L. ferripilum* ML-04 is sequenced and annotated. The bacterium has a single circular chromosome of 2,406,157 bp containing 2,471 coding sequences (CDS), 2 rRNA operons, 48 tRNA genes, a large number of mobile genetic elements and 2 genomic islands. *In silico* analysis shows *L. ferriphilum* ML-04 fixes carbon through a reductive citric acid (rTCA) cycle, and obtains nitrogen through ammonium assimilation. The genes related to "cell envelope biogenesis, outer membrane" (6.9%) and "DNA replication, recombination and repair" (5.6%) are abundant, and a large number of genes related to heavy metal detoxification, oxidative and acidic stress defense, and signal transduction pathways were detected. The genomic plasticity, plentiful cell envelope components, inorganic element metabolic abilities and stress response mechanisms found the base for this organism's survival in the bioleaching niche.**

Keywords**:** *Leptospirillum ferriphilum*, genome, physiology, bioleaching, adaptation

The use of acidophilic and chemolithotrophic iron- and sulfur-oxidizing microbes in bioleaching processes to recover metals from certain types of copper, uranium, and gold-bearing minerals or mineral concentrates is well established. *Leptospirillum ferriphilum* is a Gram-negative, moderately thermophilic, acidophilic and strictly chemolithoautotrophic bacterium, oxidizing ferrous iron as the sole energy source and fixing carbon dioxide from the atmosphere (Coram and Rawlings, 2002). It is the dominant iron-oxidizing organism in industrial biooxidation tanks, especially at high redox potential (Rawlings *et al.*, 1999; Coram and Rawlings, 2002).

Until now knowledge of *L. ferriphilum* has been limited, because *L. ferriphilum* is difficult to handle and no genetic manipulation method has been constructed. On the other hand, genome analysis is a powerful way to acquire an enormous amount of information, which allows investigators to pay attention to important issues suggested by bioinformatic analysis and predictions. Presently, there is a community genomic and proteomic analysis of *Leptospirillum rubarum* (Group II) and *Leptospirillum ferrodiazotrophum* (Group III) isolated from acid mine drainage biofilms (Goltsman *et al.*, 2009), but no complete genome of *L. ferriphilum* has been announced. In the present study, we report the complete genome sequence of pure cultured *L. ferriphilum* ML-04 and provide analyses of its physiology and environmental adaptation.

* For correspondence. E-mail: jianqunlin@sdu.edu.cn; Tel.: +86-531-8836-4426; Fax: +86-531-8856-5610
§ Supplemental material for this article may be found at http://www.springer.com/content/120956

Materials and Methods

Bacterial growth and DNA extraction

L. ferriphilum ML-04 was isolated from acidic water near a hot spring in Tengchong, Yunnan province of China. Genomic DNA was extracted and then separated into two portions, one for genomic sequencing and the other for further gap closing, stored at -80°C.

Genome sequencing and assembly

The genome was sequenced using a Roche GS FLX system. A total of 224,019 reads, constituting 23.6 fold coverage of the genome, were assembled into a total of 121 contigs. Seventy four of the 121 contigs were longer than 500 bp. The contig relationships were decided by comparing them to reference genomes (two incomplete genome sequences of *Leptospirillum rubarum* and *Leptospirillum* sp. Group II '5-way CG') and by multiplex PCR reactions. Then the gaps were filled using the ABI 3730 sequencer. The final assembly of the sequences was conducted by using the phred, phrap, consed software package (http://www.genome.washington.edu). Low quality regions of the genome were resequenced. The final accuracy of the genome sequence was 99.9962%.

Sequence analysis and annotation

Coding sequences were predicted using a combination of Glimmer (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/glimmer_3.cgi) and Zcurve (http://tubic.tju.edu.cn/Zcurve_B/). The functions of the coding sequences were annotated by searching all the peptides against GenBank's non-redundant (nr) protein database using BLASTP, as well as the data bases of swiss-port (http://www.expasy.ch/sprot/), COG (http:// www.ncbi.nlm.nih.gov/COG/), Pfam (http://pfam.sanger.

Table 1. General features of the *L. ferriphilum* ML-04 genome and its comparison with *L. rubarum* and *L.* sp. Group II '5-way CG'

	L. ferriphilum ML-04	*L. rubarum* draft genome	*Leptospirillum* sp. Group II '5-way CG' draft genome
Total size (bp)	2,406,157	2,636,861	2,724,188
GC content (%)	54.55%	54.62%	54.15%
ORFs	2,471	2,601	2,544
With assigned function	1,534	1,703	1,615
Conserved hypothetical protein	667	559	544
Hypothetical protein	270	339	385
rRNA operons	2	1	1
tRNA	48	48	54

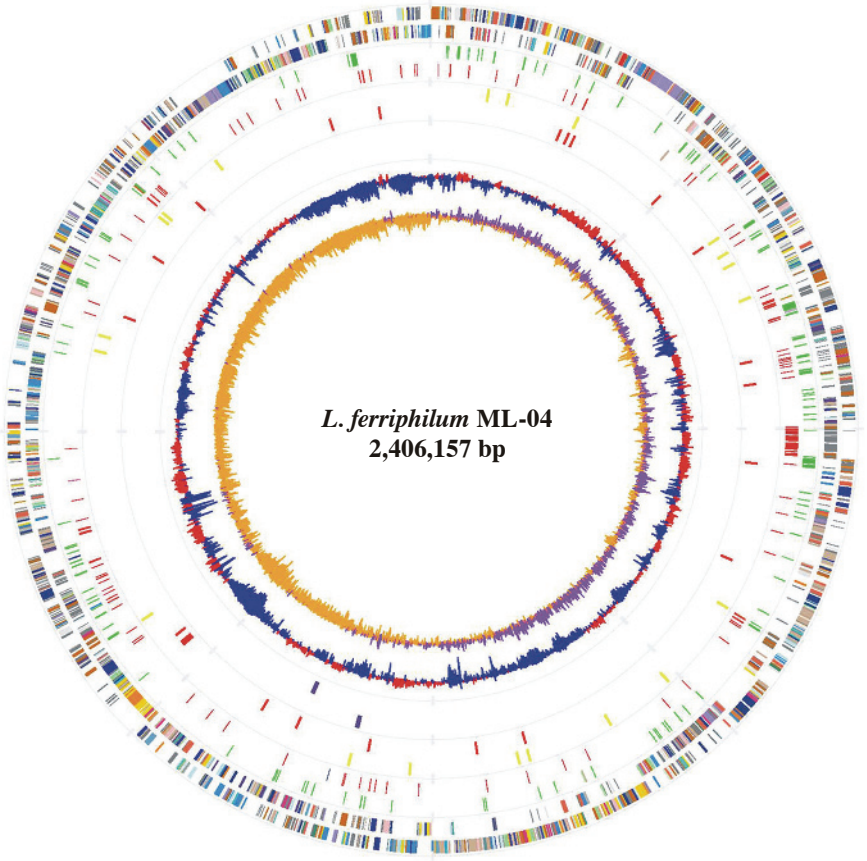

Fig. 1. Circular representation of the *L. ferriphilum* ML-04 genome sequence. The two outer circles represent predicted protein encoding genes on the forward and reverse strands. Functional categories are indicated by color as follows: energy production and conversion (royal blue4), DNA replication, recombination and repair (dark orange3), amino acid transport and metabolism (dodger blue1), transcription (dark orange1), translation, ribosomal structure and biogenesis (gold1), RNA processing and modification (orange3), chromatin structure and dynamics (maroon), cell division and chromosome partitioning (antiquewhite1), nuclear structure (yellow), defense mechanisms (pink), signal transduction mechanisms (tomato1), cell envelope biogenesis, outer membrane (peach puff3), cell motility and secretion (medium purple1), cytoskeleton (red), extracellular structures (green), intracellular trafficking and secretion (deep pink), posttranslational modification, protein turnover, chaperones (pale green1), carbohydrate transport and metabolism (blue1), nucleotide transport and metabolism (sky blue3), coenzyme metabolism (light blue1), lipid metabolism (cyan3), inorganic ion transport and metabolism (medium purple4), secondary metabolite biosynthesis, transport, and catabolism (aquamarine4), general function prediction only (gray90), function unknown (gray70), not in COGs (gray50). The third and fourth circles (forward and reverse strands) indicate the unique proteins in *L. ferriphilum* ML-04 comparing with the proteins of reference sequences (*Leptospirillum rubarum* and *Leptospirillum* sp. Group II '5-way CG', incomplete genome sequences). The fifth and sixth circles (forward and reverse strands) indicate tRNA genes. The seventh and eighth circles (forward and reverse strands) show rRNA. The ninth circle shows genomic G+C percentage content: above median GC content (red), less than or equal to the median (blue). The tenth circle shows GC skew (G-C)/(G+C): values >0 (purple), values <0 (yellow).

ac.uk/), interPro (http://www.ebi.ac.uk/Tools/InterProScan/), and KEGG (http://www.genome.jp/kegg/). When needed, gene function confirmation was performed manually one by one. The transfer RNA genes were predicted using tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/), and the ribosomal RNA genes were found by BLASTN searching against the ML-04 genome using the 16S, 23S, and 5S rRNA genes of *Leptospirillum* sp. Group II '5-way CG'. The insertion sequences (IS) were found by first using the IS Finder database (http://www-is.biotoul.fr/is.html), and then the annotated transposons were searched against NCBI nucleotide sequence databases to find IS elements. The repeat sequences were identified using Tandem Repeats Finder Program (Benson, 1999). Model metabolic pathways were constructed using KEGG (http://www.genome.jp/kegg/). Subcellular localizations of the proteins were predicted by PSORTb (http://www.psort.org/psortb/), SignalP (http://www.cbs.dtu.dk/services/SignalP/) and ConPred II (http://bioinfo.si.hirosaki-u.ac.jp/~ConPred2/). Genomic islands were searched using IslandViewer (http://www.pathogenomics.sfu.ca/islandviewer/genome_ submit.php), and codon usage was analyzed using the software codonW (ftp://molbiol.ox.ac.uk/cu/Win32codonW.zip).

Genome comparison

The genome data of *L. rubarum*, *L.* sp. Group II '5-way CG', and *L. ferrodiazotrophum* were retrieved at http://www.ncbi.nlm.nih.gov/genomes. Orthologs between *L. ferriphilum* ML-04, *L. rubarum*, *L.* sp. Group II '5-way CG', and *L. ferrodiazotrophum* were detected by an all-vs-all reciprocal BLASTP search against their protein sets, respectively. "Significant similarity" was defined as E-value ≤1e-05, the length of the protein alignment ≥60%, and amino acid sequence identity ≥30%, and then the best hit was selected.

Nucleotide sequence accession number

The complete genome sequence of *L. ferriphilum* ML-04 was deposited at GenBank under accession number CP002919.

Results and Discussion

General features of genome

L. ferriphilum ML-04 consists of a single circular chromosome of 2,406,157 bp with a G+C content of 54.55%. Plasmids were not found in this strain. The chromosome encodes 2,471 coding sequences (CDS), which represent 89.96% of the genome, 2 rRNA operons and 48 tRNA genes for all the 20 amino acids. Among the 2,471 CDS, 1,534 were assigned putative functions, 667 were predicted as conserved hypothetical proteins, and 270 were predicted as hypothetical proteins (Table 1). There were 1,645 predicted proteins that belonged to the functional category of Cluster of Orthologous Groups (COG). The origin and terminus of replication, predicted by GC-skew analysis, were positioned diametrically opposite from each other. Directly adjacent to the replication origin, a *dnaA* gene was predicted, therefore, the base numbering start point of the genome was chosen at the first base of the *dnaA* gene (Fig. 1).

Genomic plasticity and codon usage

The genome of *L. ferriphilum* ML-04 contains many mobile genetic elements. 36 transposases of IS21, ISNCY, IS4, ISL3, IS1634, IS1595, IS110, and IS256 families and 17 integrases were found. It indicates that recent gene duplication events have happened because several adjacent ORFs occur as iden-

Table 2. Codon usage in *L. ferriphilum* ML-04

AA[a]	Codon	Number	RSCU[b]
Phe	UUU	18,144	**1.11**
	UUC	14,544	0.89
Leu	UUA	1,685	0.13
	UUG	8,737	0.65
	CUU	20,197	1.50
	CUC	20,186	1.50
	CUA	794	0.06
	CUG	28,990	**2.16**
Ile	AUU	14,246	1.07
	AUC	23,596	**1.77**
	AUA	2,111	0.16
Met	AUG	16,434	1.00
Val	GUU	12,869	1.01
	GUC	22,471	**1.76**
	GUA	1,859	0.15
	GUG	13,736	1.08
Ser	UCU	8,019	0.93
	UCC	20,374	**2.36**
	UCA	3,566	0.41
	UCG	10,160	1.18
	AGU	3,339	0.39
	AGC	6,298	0.73
Pro	CCU	7,358	0.71
	CCC	14,068	1.36
	CCA	3,428	0.33
	CCG	16,473	**1.59**
Thr	ACU	3,472	0.39
	ACC	13,406	**1.49**
	ACA	7,224	0.80
	ACG	11,913	1.32
Ala	GCU	7,357	0.55
	GCC	23,630	**1.78**
	GCA	7,888	0.59
	GCG	14,156	1.07
Tyr	UAU	9,824	**1.14**
	UAC	7,483	0.86
TER	UAA	505	0.63
	UAG	421	0.53
	UGA	1,477	**1.84**
His	CAU	9,273	**1.15**
	CAC	6,787	0.85
Gln	CAA	5,155	0.47
	CAG	16,749	**1.53**
Asn	AAU	8,929	0.86
	AAC	11,933	**1.14**
Lys	AAA	17,855	**1.11**
	AAG	14,261	0.89
Asp	GAU	16,858	0.94
	GAC	18,897	**1.06**
Glu	GAA	29,245	**1.25**
	GAG	17,576	0.75
Cys	UGU	2,474	0.87
	UGC	3,239	**1.13**
Trp	UGG	9,336	1.00
Arg	CGU	7,631	0.87
	CGC	9,233	1.05
	CGA	5,517	0.63
	CGG	19,655	**2.24**
	AGA	5,393	0.62
	AGG	5,122	0.58
Gly	GGU	8,863	0.61
	GGC	11,566	0.79
	GGA	22,338	**1.53**
	GGG	15,823	1.08

[a] AA is the abbreviation of amino acid.
[b] RSCU is the abbreviation of the relative synonymous codon usage The preferentiallly used codons are described in bold.

tical transposases. A region (LFML04_2454-LFML04_2462) related with arsenic resistance may have been gained by horizontal gene transfer because it is flanked by a phage integrase on the 5′ end and by an IS 21 element on the 3′ end. DNA repeats suggest genomic plasticity, and 44 tandem repeats with period sizes ranging from 8 to 126 bp were found (Supplementary data Table 1). Many tandem repeats locate in protein-coding genes, and thus could be the mechanism for generation of targeted gene variation. The existence of a large number of mobile elements and tandem repeats indicates the evolution potential of the strain for adaptation to the extreme environment. Islandviewer identified two genomic islands (GI) in the genome (supplementary 2). GI-I encodes eight ORFs (LFML04_0280-LFML04_0287), including topoisomerase IA and deoxyuridine 5'-triphosphate nucleotidohydrolase both involved in DNA processing and repair, peptidase subunit HslV and heat shock protein HslU forming an ATP-dependent protease complex that may have served as a precursor or ancestral type of proteasome, glucose-inhibited division protein A involved in tRNA processing, phage integrase, acetylglutamate kinase and one conserved hypothetical protein. GI-II encodes eleven ORFs (LFML04_0503-LFML04_0513), including seven proteins of unknown function, two transposases, one IstB-like ATP-binding protein and one endodeoxyribonuclease RusA involved in correction of mistakes in genetic recombination and DNA repair. Codon usage analysis shows that codons ending with C are the most representative in the preferentially used codons (8 out of 18). The value of codon effective number is 49.83 (>40), suggesting codon preference is lowly biased in *L. ferriphilum* ML-04 (Table 2).

Comparative genome

Based on 16S rRNA gene analysis, the phylogenetic tree of *Leptospirillum* spp. was constructed (Fig. 2). *L. ferriphilum* ML-04, *L. rubarum*, and *L.* sp. Group II '5-way CG' are close phylogenetic relatives with 16S rRNA gene identity of 99%. Ortholog analysis among the three strains shows that 1,921 genes are shared and constitute the core genome. *L. ferriphilum* ML-04 encoded a total of 460 strain-specific proteins. Strain-specific proteins of known function mostly belonged to the COGs categories of "Cell envelope biogenesis, outer

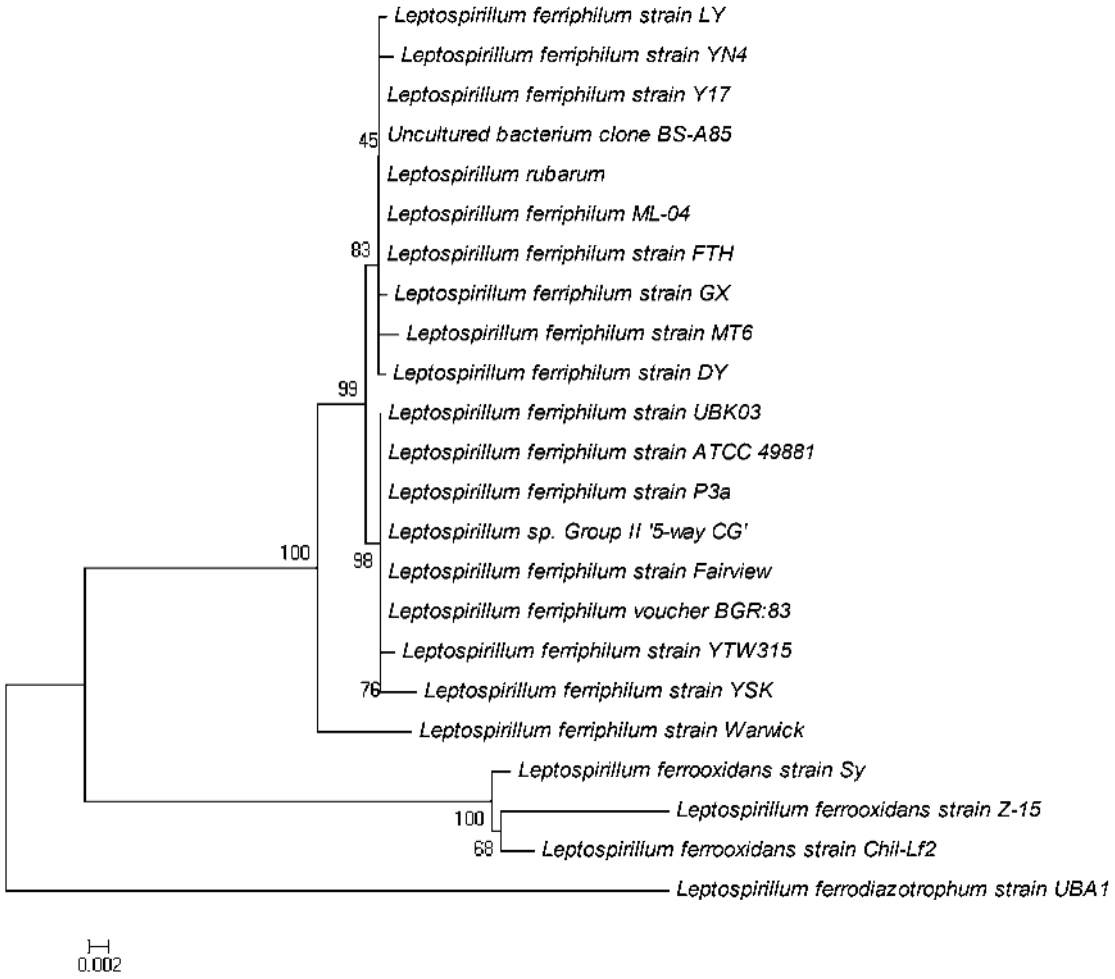

Fig. 2. Phylogenetic tree based on 16S rRNA genes of *Leptospirillum* spp. (Neighbor-Joining method). Bootstrap values are labeled at the nodes.

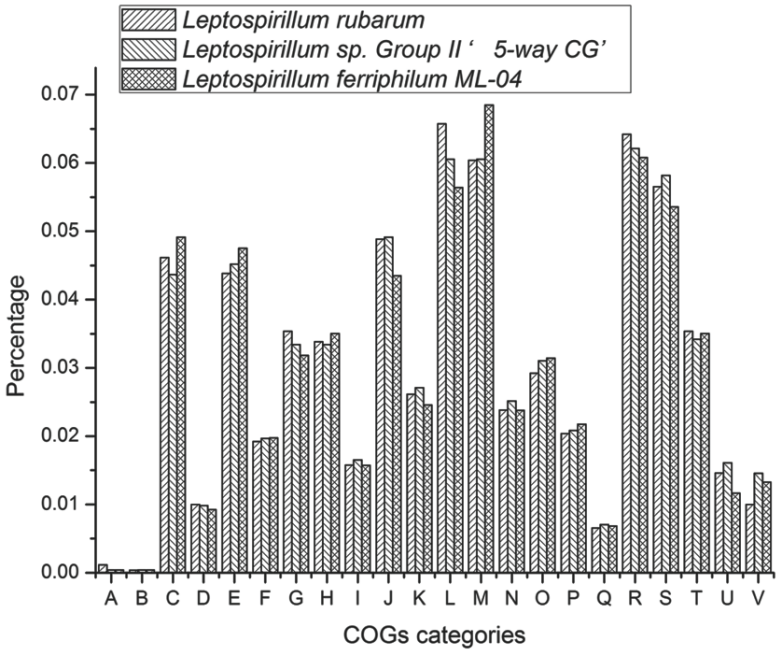

Fig. 3. Comparison of proportions of functional categories in the genomes of *Leptospirillum rubarum*, *Leptospirillum* sp. Group II '5-way CG' and *Leptospirillum ferriphilum* ML-04. A, RNA processing and modification; B, Chromatin structure and dynamics; C, Energy production and conversion; D, Cell division and chromosome partitioning; E, Amino acid transport and metabolism; F, Nucleotide transport and metabolism; G, Carbohydrate transport and metabolism; H, Coenzyme metabolism; I, Lipid metabolism; J, Translation, ribosomal structure and biogenesis; K, Transcription; L, DNA replication, recombination and repair; M, Cell envelope biogenesis, outer membrane; N, Cell motility and secretion; O, Posttranslational modification, protein turnover, chaperones; P, Inorganic ion transport and metabolism; Q, Secondary metabolite biosynthesis, transport and catabolism; R, General function prediction only; S, Function unknown; T, Signal transduction mechanisms; U, Intracellular trafficking and secretion; V, Defense mechanisms

membrane, and DNA replication, recombination and repair", which indicates that the defense mechanisms of *L. ferriphilum* ML-04 to the external environment may be unique, compared with the other two strains. The arsenic resistance gene cluster (LFML04_2453-LFML04_2463) containing *arsRCDA*, *arsB*, and *pstS* exists in *L. ferrriphilum* ML-04 only. By proportion analysis of functional categories, it shows the most representative functional categories of the three strains are "DNA replication, recombination and repair" and "cell envelope biogenesis, outer membrane" (Fig. 3). But in the genome of *Escherichia coli* ATCC 8739, the largest percentages are "amino acid transport and metabolism" (8.67%), and "carbohydrate transport and metabolism" (8.16%). In *L. ferriphilum* ML-04, the category of "DNA replication, recombination, and repair" contains large numbers of transposases, integrases, and DNA repair related enzymes, and the category of "cell envelope biogenesis, outer membrane" contains enzymes involved in the synthesis of the cell membrane, extracellular polymeric substances and membrane channel proteins. These are characteristic for *L. ferriphilum* ML-04 living in a condition of extreme acidity and an abundance of metal ions. DNA repair proteins play roles when cell damage occurs, and a transposable element is a medium through which a bacterium could acquire new genes for resistance to environmental stress, thus, the above two categories may have come to constitute the largest percentages in *L. ferriphilum* as the result of long-term adaptation. *L. ferriphilum* ML-04 and *L. ferrodiazotrophum* share 92% identity of the 16S rRNA gene. Ortholog analysis between the two strains indicates 1,741 proteins are shared, while 742 proteins are unique in *L. ferriphilum* ML-04. Among the unique proteins in *L. ferriphilum*, the proteins of transposable elements account for a large proportion, except for the proteins of unknown functions. Furthermore, the following genes of known function are unique in *L. ferriphilum* ML-04: the genes *ectABCD* (LFML04_0405-LFML04_0402) involved in the syntheses of ectoine and hydroxyectoine which are compatible solvents resistant to osmotic and thermal stresses, cellulose synthase (LFML04_0982-LFML04_0985, LFML04_2273), mercuric transport protein MerT (LFML04_0653), and fatty acid desaturase (LFML04_0736).

Chemolithoautotrophy

L. ferriphilum has well-developed abilities for inorganic carbon and nitrogen fixation, biomacromolecule syntheses, and ferrous iron oxidation for energy production when living in an inorganic environment.

Carbon metabolism for carbon dioxide fixation and the TCA cycle

L. ferriphilum ML-04 has a complete reductive citric acid (rTCA) cycle, which is used for CO_2 fixation (Fig. 4, Supplementary data Table 3). Four key enzymes of the rTCA cycle and pyruvate formation were detected in *L. ferriphilum* ML-04, namely, fumarate reductase (FRD) (LFML04_0823, LFML04_

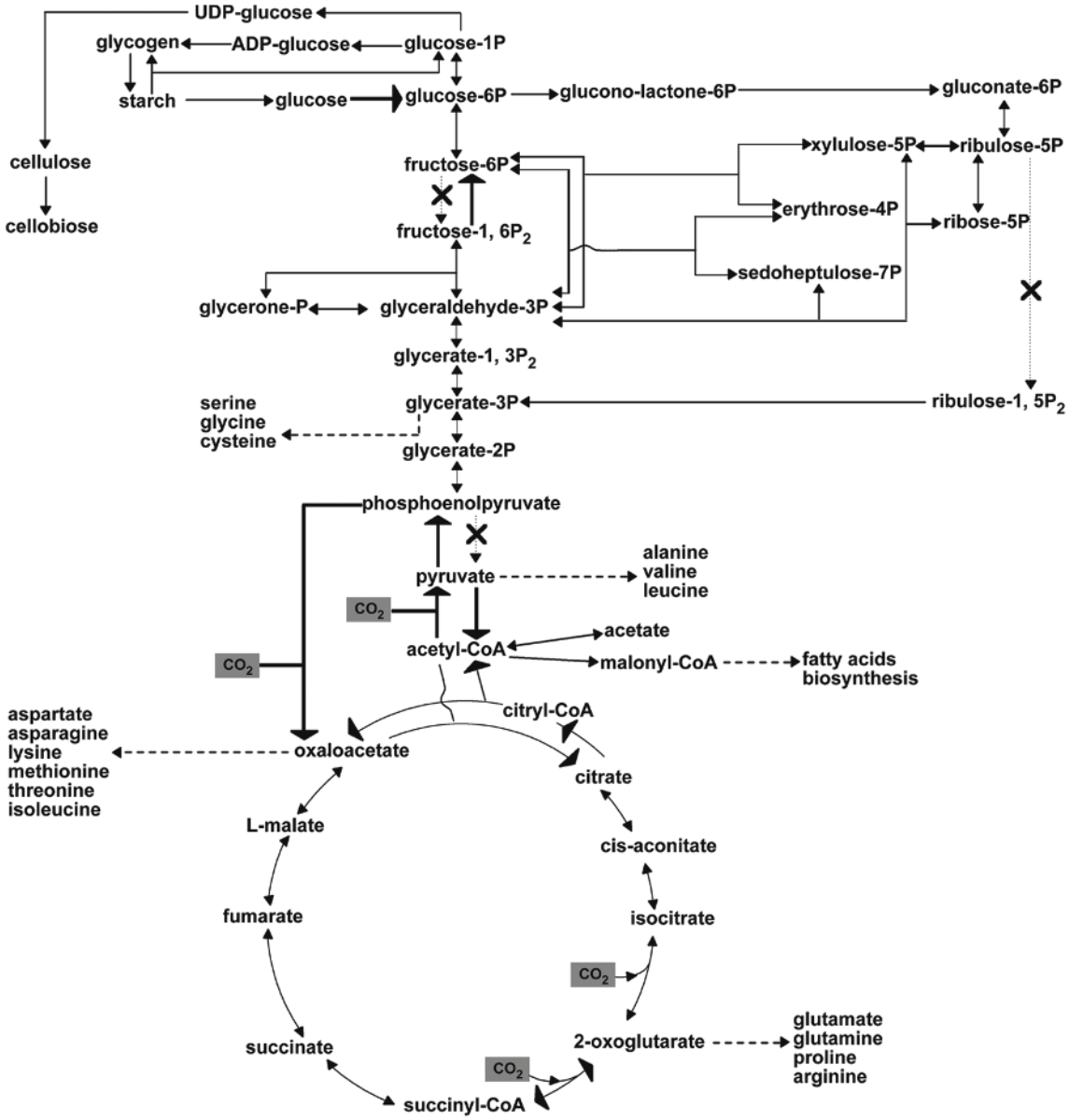

Fig. 4. Predicted central pathways of carbon metabolism for *L. ferriphilum* ML-04.

0824), 2-oxoglutarate ferredoxin oxidoreductase (OGOR) (LFML04_2421-LFML04_2424), pyruvate ferredoxin oxidoreductase (PFOR) (LFML04_2415-LFML04_2418), and phosphoenolpyruvate carboxylase (PEPC) (LFML04_2477). Although one of the key enzymes, ATP citrate lyase (ACL), was not annotated, its function could be replaced by succinyl-CoA synthetase (LFML04_0818, LFML04_0819, LFML04_0825, LFML04_0826) and citrate synthase (LFML04_0821) (Levican *et al*., 2008; Goltsman *et al.*, 2009).

The other three pathways of CO_2 fixation, the Calvin-Bassham-Benson (CBB) cycle, the reductive acetyl-coenzyme A pathway, and the 3-hydroxpropionate cycle, do not exist in *L. ferriphilum* ML-04. Rubisco is one of the key enzymes of the CBB cycle. Two gene copies of a non-canonical form IV Rubisco-like protein (RLP) (LFML04_2084, LFML04_2516) were identified, but they were not demonstrated to be involved in CO_2 fixation (Hanson and Tabita, 2001).

L. ferriphilum ML-04 may have a complete TCA cycle (Fig. 4, Supplementary data Table 3). Two copies of the dihydrolipoamide dehydrogenase subunit of the 2-oxoglutarate dehydrogenase complex (LFML04_0159, LFML04_1324) were identified, but the other two subunits were not found. Alternatively, 2-oxoglutarate ferredoxin oxidoreductase (LFML04_2421-LFML04_2424) could serve this function and catalyze 2-oxoglutarate to succinyl-CoA (Kerscher *et al*., 1982; Mai and Adams, 1996; Zhang *et al*., 1996; Yun *et al*., 2002). Besides, two main components of the succinate dehydrogenase/fumarate reductase complex were identified in *L. ferriphi-*

lum ML-04, which are responsible for interconversion of succinate and fumarate (Maklashina *et al*., 1998).

The glyoxylate cycle does not exist in *L. ferriphilum* ML-04 because of the lack of the key enzymes isocitrate lyase and malate synthase. In addition, two important enzymes involved in the replenishment of TCA cycle intermediates - pyruvate carboxylase and malic enzyme - were not found. Only phosphoenolpyruvate carboxylase (PEPC) (LFML04_2477), which is involved in the replenishment of intermediates, was identified.

Carbon metabolisms of carbohydrates and acetate

L. ferriphilum ML-04 has an incomplete glycolysis (Embden-Meyerhoff, EM) pathway as it lacks 6-phosphofructokinase, hexokinase and pyruvate kinase, although glucokinase (LFML04_1634) or other carbohydrate kinase family enzymes could play the role of hexokinase. Similar results were reported in *L. rubarum* (Goltsman *et al.*, 2009). The Entner-Doudoroff (ED) pathway does not exist in *L. ferriphilum* ML-04, as the two key enzymes of the ED pathway, 6-phosphogluconate dehydrase and 2-keto-3-deoxy-6-phosphogluconate aldolase, were not identified.

L. ferriphilum ML-04 has a complete pentose phosphate pathway (Fig. 4, Supplementary data Table 3) to produce the reducing equivalents (in the form of NADPH) and ribose-5-phosphate. Glucose can be oxidized to glyceraldehyde 3-phosphate through the pentose phosphate pathway. Maltose can be split into glucose by 4-alpha-glucanotransferase (LFML04_2392) (Seibold *et al.*, 2009). Fructose and sucrose cannot be utilized by *L. ferriphilum* ML-04 because of the absence of fructokinase. The enzymes for catabolism of mannose, sorbose, fucose, rhamnulose, lactose, and galactose do not exist in *L. ferriphilum* ML-04.

Acetate can be produced by alcohol dehydrogenase (LFML04_0384, LFML04_2026, LFML04_0746, LFML04_0036, LFML04_1645) and aldehyde dehydrogenase (LFML04_ 2247) from alcohol, or by cysteine synthase (LFML04_0887), which utilizes O-acetyl-L-serine and hydrogen sulfide (H_2S) to produce acetate and L-cysteine. Acetate can be converted to acetyl-CoA by acetyl-CoA synthetase (LFML04_0102).

Carbon metabolism for gluconeogenesis, glycogen, starch, and cellulose biosyntheses

L. ferriphilum ML-04 has almost all the enzymes needed for gluconeogenesis (Supplementary data Table 3), which ensures that phosphorylated glucose is generated from non-carbohydrate carbon substrates. The enzymes of pyruvate carboxylase and phosphoenolpyruvate carboxykinase for transforming pyruvate into phosphoenolpyruvate were not found, but this reaction could be catalyzed by phosphoenolpyruvate synthase (LFML04_2110) (Hutchins *et al*., 2001). However, glucose-6-phosphatase as the key enzyme in gluconeogenesis, glucose-1-phosphatase and glucose-1-phosphate phosphodismutase were not identified in this organism, which indicates that phosphorylated glucose could not be converted to glucose. Glycogen and starch could be the major forms of short term energy storage in *L. ferriphilum* ML-04, as the organism has the genes required for synthesis and hydrolysis of glycogen and starch. Glucose-1-phosphate adenylyltransferase (LFML04_2394), and glycogen/starch synthase (ADP-glucose type) (LFML04_1405) are involved in the synthesis of amylose (Preiss, 1984). No glycogen branching enzyme was found. Glycogen phosphorylase (LFML04_1326), glycogen debranching enzyme (LFML04_1274, LFML04_1513, LFML04_2195, LFML04_2196), phosphoglucomutase (LFML04_1327) catalyzing glycogen catabolism, and alpha-amylase (LFML04_0743) hydrolyzing starch and glycogen were found. The UTP- glucose-1-phosphate uridylyltransferase (LFML04_0141, LFML04_0776), and cellulose synthase (UDP-forming) subunit B (LFML04_0983), subunit A-like protein (LFML04_0984, LFML04_0985), and two subunit C-like proteins (LFML04_0982, LFML04_2273) were identified, which catalyze the biosynthesis of cellulose from glucose-1-phosphate (Wong *et al*., 1990). Endoglucanase (LFML04_2272) was found, which hydrolyzes cellulose into cellobiose. Cellulose biosynthesis was reported in *L. rubarum* (Goltsman *et al*., 2009).

Carbon metabolisms of fatty acid biosynthesis

L. ferriphilum ML-04 has all the genes for fatty acid biosynthesis (Supplementary data Table 3) encoding 3-oxoacyl-(acyl-carrier-protein) synthase III (FabH, LFML04_2133, LFML04_0059), 3-oxoacyl-(acyl-carrier-protein) synthase II (FabF, LFML04_2129), 3-oxoacyl-(acyl-carrier-protein) reductase (FabG, LFML04_2131, LFML04_0260), beta-hydroxyacyl- (acyl-carrier-protein) dehydratase (FabZ, LFML04_1219), NADH-dependent enoyl-(acyl-carrier-protein) reductase (FabI, LFML04_1072), malonyl-CoA-(acyl-carrier-protein) transacylase (FabD, LFML04_2132), acyl carrier protein (ACP, LFML04_2130), and acetyl-CoA carboxylase (LFML04_1358, LFML04_1522, LFML04_1523, LFML04_1940). The fatty acid desaturase (LFML04_0736) and acyl-(acyl-carrier-protein) desaturase (LFML04_1474) were found.

L. ferriphilum ML-04 does not have a β-oxidation pathway because it lacks acyl-CoA dehydrogenase and beta-keto-thiolase, even though it has acyl-coenzyme A synthetase (LFML04_0102), enoyl-CoA hydratase (LFML04_0753), and 3-hydroxyacyl-CoA dehydrogenase (LFML04_1540). Therefore, fatty acids cannot be used as energy source in *L. ferriphilum* ML-04.

Carbon metabolism for precursor biosynthesis of extracellular polymeric substances (EPS)

UDP-galactose, UDP-glucose, dTDP-rhamnose are the precursors for EPS biosynthesis, and they are synthesized from glucose-1-phosphate (glucose-1-P) (Barreto *et al*., 2005). The UTP-glucose-1-phosphate uridylyltransferase (GalU, LFML04_0141, LFML04_0776), UDP-glucose 4-epimerase (GalE, LFML04_0810, LFML04_0896), and the RfbABCD complex (dTDP-rhamnose synthase, LFML04_0724-LFML04_0726, LFML04_1453, LFML04_1567) were identified in *L. ferriphilum* ML-04 (Supplementary data Table 3). The *rfbB*, *rfbA*, and *rfbC* were located together (LFML04_0724-LFML04_0726), and two others, *rfb*B (LFML04_1453) and *rfbD* (LFML04_1567), were located separately.

The production and secretion of EPS by metal leaching bacteria is regarded as a key mechanism promoting irreversible cell attachment to the surfaces of ores and the development of biofilms. EPS is useful in providing a micro-environment suitable for efficient extraction of ores, and helps gather the bacterial cells together to communally take up nutrients and defend against their hostile environment.

Nitrogen metabolism

There are three ammonium transporters and three nitrogen regulatory proteins PII in the genome of *L. ferriphilum* ML-04. Among them, two ammonium transporters (LFML04_0365, LFML04_0367) and one nitrogen regulatory protein PII (LFML04_0366) are in a cluster. The third ammonium transporter (LFML04_0946) is located next to the second nitrogen regulatory protein PII (LFML04_0947). The third nitrogen regulatory protein PII (LFML04_2275) is located separately, and upstream of it is a Nif-specific regulatory protein (NifA) (LFML04_2279) involved in nitrogenase operon regulation, even though there is no nitrogenase in *L. ferriphilum* ML-04 (Dixon, 1998). When ammonium is inside the cell, it is incorporated into glutamine by glutamine synthetase (GlnA, LFML04_1897), and subsequently transformed to glutamate by a predicted glutamate synthase (ferredoxin) (LFML04_1478, LFML04_1479) (Muro-Pastor *et al.*, 2005). Ammonium can be incorporated into glycine by aminomethyltransferase (LFML04_2384). The predicted putative class-I glutamine amidotransferase (LFML04_1446) can hydrolyze glutamine and transfer the ammonia to other substrates. AmoA (LFML04_1996), one subunit of ammonia monooxygenase, which can oxidize ammonia to hydroxylamine was identified.

It is not clear whether nitrite can be utilized. The NapC/NirT family cytochrome c (LFML04_2062) which is one subunit (NrfH) of cytochrome c nitrite reductase was found, but the catalytic subunit (NrfA) was not predicted (Simon *et al.*, 2000). Nitrite or sulfite reductase-like protein (LFML04_1409), and nitrite reductase small subunit-like protein (LFML04_2396) were located.

The regulation of nitrogen metabolism is controlled by the above mentioned PII family proteins through detection of the intracellular glutamine and ATP levels (Leigh and Dodsworth, 2007). The co-location of ammonium transporter and protein PII might indicate that protein PII is involved in ammonium uptake regulation. Multiple copies of signal transduction histidine kinases and putative two-component, sigma54 specific transcriptional regulators (Fis family) are present in the genome, and they are homologues to the ntrB/ntrC two-component regulatory system in nitrogen assimilation (Levican *et al.*, 2008). So the regulation of nitrogen metabolism in *L. ferriphilum* ML-04 might be accomplished by PII, NifA, and NtrB/NtrC-like proteins together. Nitrogen metabolism is similar to *L. rubarum* (Goltsman *et al.*, 2009).

Ferrous iron oxidation

Many genes participating in ferrous iron oxidation were identified. One predicted protein (LFML04_0321) shares 61% amino acid sequence identity with cytochrome 572, and the other two predicted proteins (LFML04_2057, LFML04_2239) share 81% and 80% amino acid sequence identities with cytochrome 579. Cytochrome 572 and 579 are important components for ferrous iron oxidation and electron transfer (Jeans *et al.*, 2008; Singer *et al.*, 2008). Two proteins of unknown function (LFML04_0633, LFML04_1787) were found to be similar to PioC, which is related to ferrous iron oxidation in *Rhodopseudomonas palustris* TIE-1. Each of the above two proteins contains a high potential iron-sulfur protein domain (HiPIP), which functions in electron transfer. The ferrous iron oxidation activity decreases by 60% after *pioC* is knocked out (Jiao and Newman, 2007). So these two proteins (LFML04_0633, LFML04_1787) are candidates for an electron transfer chain. Fourteen subunits of the NADH dehydrogenase (ubiquinone) (LFML04_0929-LFML04_0943) and eight subunits of F0F1-type ATP synthase (LFML04_0079-LFML04_0086) were found. The cyd operon was located encoding a cytochrome bd-type quinol oxidase (LFML04_0537-LFML04_0540), which is a terminal oxidase in the aerobic respiratory chain and usually functions in micro-aerobic conditions. Two gene clusters encoding putative bc1 complexes were identified. In the first gene cluster (LFML04_1073-LFML04_1074), cytochrome b and cytochrome c are fused into one protein con-

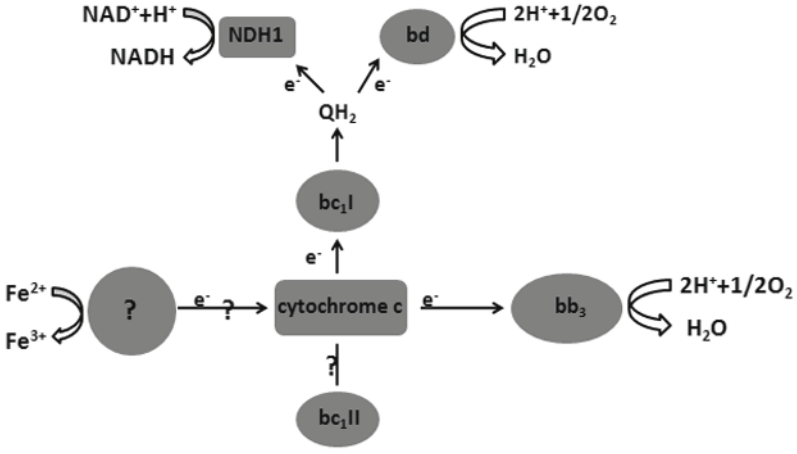

Fig. 5. The deduced Fe(II) oxidation pathway in *L. ferriphilum* ML-04. Cytochrome c is the major electron transfer protein. In the downhill electron pathway, the electrons are deduced to be transferred through cytochrome c to the terminal cbb3 type cytochrome oxidase to drive ATP synthesis. In the uphill pathway, the electrons are deduced to be first transferred through cytochrome c to cytochrome bc1 complex, then to the quinone pool, and finally to the NADH dehydrogenase complex to reduce NAD^+ to NADH. Alternatively, the electrons transfer from the quinone pool to the cytochrome bd-type quinol oxidase to drive ATP synthesis.

taining three domains: the N-terminal domain and C-terminal domain of cytochrome b/b6, and the monohaem domain of cytochrome c. The cytochrome b of the second bc1 complex (LFML04_1934-LFML04_1930) is split into two predicted proteins: N-terminal domain of cytochrome b/b6 (LFML04_1933) and C-terminal domain of cytochrome b/b6 (LFML04_1932). Two subunit I (LFML04_1329, LFML04_1847) and two mono-heme subunits (LFML04_1458, LFML04_1964) of the cbb3 type cytochrome oxidase were noted, and constituted the core functional unit of the enzyme (Ducluzeau *et al*., 2008). Two copies of subunit I of cytochrome c oxidase (Cox1) (LFML04_1956, LFML04_1958) were found. Many cytochrome c and NADH ubiquinone oxidoreductase subunits were found dispersed in the genome, which could be potential components of the electron transfer chain (Supplementary data Table 4). That two electron transfer branches, such as the uphill and downhill branches in *Acidithiobacillus ferrooxidans* (Quatrini *et al*., 2009), could exist in *L. ferriphilum* ML-04 is suggested by the existence of a cbb3 type cytochrome oxidase, two bc1 complexes, and multiple copies of cytochrome c. The electron transfer chain is probably more complex than that in *A. ferrooxidans*, since more components, such as the cytochrome bd-quinol oxidase, subunits of cytochrome c oxidase and two predicted bc1 complexes were found. A model of the electron transfer chain was deduced from the above information (Fig. 5).

Sulfur metabolism

Bacteria take up sulfur in the form of inorganic sulfate and organic sulfur compounds, for example, sulfate esters, sulfonates, and sulfur containing amino acids. In the genome of *L. ferriphilum* ML-04, no genes encoding cysteine and methionine transporters, and no genes encoding the typical sulfate transporters such as the SulT and SulP families and a sulfate ester transporter were found. One predicted periplasmic component of a probable ABC transporter (LFML04_0205) shares 52% amino acid sequence identity with a putative sulfate-binding protein. Another ABC transporter, ATP-binding protein (LFML04_0276), was predicted to be a sulfonate or nitrate ABC transporter by domain analysis, and shares sequence identity of over 30% with the ATP-binding component of a sulfonate transporter (TauB) (Eichhorn, 2000). The downstream gene (LFML04_0277) also encodes a putative ABC transporter ATP-binding protein and has an ABC nitrate/sulphonate/bicarbonate family transporter domain. One unknown function protein (LFML04_1767) has 28% sequence identity with a putative sulfate-binding protein. Almost all enzymes involved in cysteine biosynthesis from sulfate exist (van der Ploeg *et al*., 2001; Valdes *et al.*, 2003), including sulphate adenylyltransferase (LFML04_0491), adenylylsulfate reductase (LFML04_0489, LFML04_0490), hemoprotein beta-component (CysI) of nitrite or sulphite reductase (LFML04_1409), uroporphyrin-III C-methyltransferase/synthase (CysG) (LFML04_1985), serine O-acetyltransferase (CysE) (LFML04_1256), and cysteine synthase (LFML04_0887), but the flavoprotein alpha-component (CysJ) of sulfite reductase (NADPH) was not found.

In the genome of *L. ferriphilum* ML-04, there are two sulfide-quinone reductases (LFML04_0046, LFML04_0541) involved in sulfide oxidization, and one cytochrome bd-type quinol oxidase (LFML04_0537-LFML04_0540) located in a cluster with one of the sulfide-quinone reductases (LFML04_0541). Several predicted proteins potentially participate in the transformation among different kinds of sulfur compounds. The protein (LFML04_0203) with a rhodanese-like domain, which breaks the S-S bond of thiosulfate to generate sulfur and sulfite, was identified. Those predicted proteins (LFML04_0489, LFML04_0490, LFML04_0491) involved in cysteine biosynthesis can also catalyze sulfite transformation to sulfate. Two gene clusters (LFML04_0356-LFML04_0357, LFML04_1201-LFML04_1205) of DsrE family-like proteins, involved in sulfur transfer during the sulfur oxidation process, were found (Dahl *et al.*, 2008).

Stress response

Bioleaching bacteria live in extremely hostile conditions, usually with high concentrations of heavy metal ions, high oxidative stress and extreme acidity. *L. ferriphilum* ML-04 has acquired several kinds of resistance mechanisms during evolution as described in the following sections.

Heavy metal resistance

Two gene clusters and several dispersed genes related to arsenic resistance were detected in *L. ferriphilum* ML-04 (Li *et al*., 2010). Sequence analysis shows there are one phage integrase and one IS 21 element in the upstream and downstream sequences of *LfML04Ars2,* which indicates that the arsenic resistance gene cluster was possibly acquired through horizontal gene transfer. There is also a new discovery of a conserved hypothetical protein (LFML04_1001), with a domain related to heavy metal transport or detoxification, arrayed together with *arsRC* (LFML04_1004) and *arsB* (LFML04_1002). *arsRC* and *arsB* are similar to *Lfars* in *L. ferriphilum* strain Fairview (Tuffin *et al*., 2006).

The genes related to mercury resistance are dispersed in the genome of *L. ferriphilum* ML-04. The mercuric transport protein of MerT (LFML04_0653), mercury (II) reductase of MerA (LFML04_0795), and transcriptional regulatory proteins of the MerR family (LFML04_0796, LFML04_1293) were found.

L. ferriphillum ML-04 is resistant to Cu^{2+} up to 0.9 mol/L (data not shown). One copper translocating P-type ATPase (LFML04_0368) was found. Four complete *czcCBA*-like clusters (LFML04_0190-LFML04_0192, LFML04_0637-LFML04_0635, LFML04_0660-LFML04_0658, LFML04_1791-LFML04_1789) encoding a membrane efflux pump for Cd^{2+}, Zn^{2+}, Co^{2+} and two other *czcAB*-like genes (LFML04_1448, LFML04_1447) without *czcC* were identified. Two integral membrane pumps of putative Co/Zn/Cd cation transporter (LFML04_0906, LFML04_1494) were also found.

Oxidative stress response

L. ferriphilum is more resistant to high redox potential than other bioleaching bacterium (Rawlings *et al*., 1999) with well-developed oxidative stress defense mechanisms to produce non-enzymatic antioxidants, and specific enzymes to remove active oxidants, or to repair the damage caused by oxidants. Phytoene synthase (LFML04_2158, LFML04_2160) and phytoene dehydrogenase (LFML04_0980) were identified, which catalyze the synthesis of ζ-carotene from phytoene. One thioredoxin reductase (LFML04_2379), six thioredoxins (LFML04_

0373, LFML04_1075, LFML04_1107, LFML04_1299, LFML04_2341, LFML04_2342), and one glutaredoxin (LFML04_0610) were located. Several enzymes capable of detoxifying the reactive oxygen molecules were identified. They are alkyl hydroperoxide reductase subunit c (AhpC) (LFML04_2086, LFML04_0565, LFML04_0389) which decreases the levels of organic hydroperoxides and H_2O_2; rubrerythrin (LFML04_0428), which works as a hydrogen peroxide reductase and catalyzes the NAD(P)H peroxidase reaction and the rubredoxin (Rub) peroxidase reaction (Kurtz, 2006); cytochrome c peroxidase (LFML04_2087), which converts H_2O_2 to H_2O; and NADH oxidase (LFML04_1842, LFML04_2045), which eliminates oxygen molecules in the environment (Kawasaki *et al*., 2004). One gene (LFML04_0827) in relation to alkyl hydroperoxide reductase or thiol specific antioxidant was found. But no genes for catalase and superoxide dismutase were detected.

The active oxygen molecules which escape from the defense mechanisms can lead to damage to DNA and protein. In this connection, *L. ferriphilum* has some DNA repair (supplementary 5) and protein repair mechanisms. DnaK (LFML04_1192) and GroEL (LFML04_1106) are induced by H_2O_2 and O_2^- (Morgan *et al*., 1986; Walkup and Kogoma, 1989), which decrease the amount of misfolded protein under conditions of oxidative stress. The peptide methionine sulfoxide reductase (LFML04_0614) can reduce the oxidized form of methionine back to normal methionine (Cabreiro *et al*., 2006).

DNA repair

In the genome of *L. ferriphilum* ML-04 several genes for DNA repair were identified, which are involved in base excision repair, mismatch repair, nucleotide excision repair, photoreactivation, recombination repair and SOS-inducible repair (Supplementary data Table 5). The genome contains all the enzymes needed in base excision repair. But the enzymes involved in mismatch repair are not complete. Methyl-directed mismatch repair protein (MutH), which splits the 5′ end of GATC in the unmethylated strand, and DNA exonuclease (*Exo*I, *Exo*X) do not exist. The enzymes of nucleotide excision repair are complete, which also repair the DNA in transcription. Deoxyribodipyrimidine photolyase encoded by *phr* (LFML04_2009) plays a role in photoreactivation. The genes *recD*, *recB*, and *recC* (LFML04_2023-LFML04_2025) are arrayed near one another and function in recombination repair together with *recA* (LFML04_1118). When cells suffer severe DNA injury, SOS-response transcriptional repressor (LexA, LFML04_0944, LFML04_0993) and RecA will start an SOS-inducible repair (Moat *et al*., 2002).

pH tolerance

L. ferriphilum grows in an extremely acidic environment (pH<3), and has acquired some mechanisms to defend against this extreme condition. The inside positive potential $\Delta\psi$ of the cell generated by cation transporters can inhibit proton influx, and the potassium ions play an important role in maintaining $\Delta\psi$ (Baker-Austin and Dopson, 2007). In the genome of *L. ferriphilum* ML-04, one potassium transporting ATPase KdpCAB (LFML04_0551, LFML04_0552, LFML04_0554), one potassium channel protein TrkA (LFML04_2322), and one voltage-gated potassium channel protein (LFML04_2010) were located. Proton transporters can remove excessive protons from the cytoplasm (Michels and Bakker, 1985). One Na^+/H^+ antiporter (LFML04_1999, LFML04_2001), and two putative MotA/TolQ/ExbB proton channel proteins (LFML04_1154, LFML04_2017) were found. Two decarboxylases, glutamate decarboxylase (LFML04_0731) and arginine decarboxylase (LFML04_1110) were identified, which help maintain normal intracellular pH by consuming excessive protons in the process of decarboxylation (Castanie-cornet *et al*., 1999). Several chaperonins GroEL (LFML04_1106), ClpC (LFML04_1171), ClpB (LFML04_1038, LFML04_2405) and DnaK (LFML04_1192) were dispersed in the genome, which help cells endure the acidic environment (Leverrier *et al*., 2004).

Signal transduction

Several two-component systems (TCS) including the typical one of "chemotaxis" were found in *L. ferriphilum* ML-04. These are used in sensing the external environment and transmiting the signals into cells for adaptation to the environment as described in the following.

Two-component system

TCS is a mechanism found broadly in bacteria for signal transduction. The system is typically composed of a histidine kinase (HK) and a response regulator (RR). 13 HK genes, 18 RR genes and one hybrid HK carrying both HK and RR (Supplementary data Table 6) were found by blastp and CDD analyses (http://www.ncbi.nlm.nih.gov/Structure/cdd/docs/cdd_search.html). Among the above genes, there were 12 pairs of HK and RR genes, one orphan HK gene and five orphan RR genes. It is interesting that there is an RR gene (LFML04_0755) lying immediately upstream of the hybrid HK (LFML04_0754). The TCSs in *L. ferriphilum* ML-04 may participate in nitrogen fixation and metabolic regulation (LFML04_0057, LFML04_0058), potassium transport (LFML04_0561, KdpD) and bacterial chemotaxis.

Flagellum formation and chemotaxis

Several gene clusters and dispersed genes for flagellum synthesis (Supplementary data Table 4) were found. Bacteria tend to escape from harm through the mechanism of chemotaxis. The chemotaxis genes *cheA* (LFML04_0250), *cheY* (LFML04_0249), *cheW* (LFML04_0253), *cheV* (LFML04_1052), *cheR* (LFML04_0247, LFML04_0248), *cheB* (LFML04_0251), *cheV* (LFML04_1052), and *cheZ* (LFML04_0953) were detected (Szurmant and Ordal, 2004). Several methyl-accepting chemotaxis sensory transducers are dispersed in the genome, but *cheRYAB*, *mcp*, and *cheW* are in one cluster (LFML04_0247-LFML04_0253) located between the genes encoding flagella biosynthetic proteins and flagella motor proteins. There have been several reports about chemotaxis of bioleaching bacteria. In *Leptospirillum ferrooxidans*, Fe^{2+}, Ni^{2+}, and Cu^{2+} are attractants while aspartate is a repellent (Acuña *et al*., 1992). *Thiobacillus caldus* responds chemotactically to sulphur gradients (Edwards *et al*., 2000). This chemotaxis mechanism and flagellum rotation could help *L. ferriphilum* ML-04 respond to the gradients of metal ions, ores, and organic compounds, and direct it to a beneficial growth environment.

Secretion systems

Bacteria use secretion systems as strategies for communica-

tion with the environment, and do their utmost to manipulate the environmental response for their benefit. Several types of secretion systems were found in the genome of *L. ferriphilum* ML-04. The type I secretion system consists of an ATP-binding cassette (ABC), an adaptor or membrane fusion protein (MFP) and an outer membrane protein (OMP). It transfers unfolded protein across both membranes of the bacterium in a single step. In the genome of *L. ferriphilum* ML-04, two gene clusters (LFML04_1132, LFML04_1133, LFML04_1137; LFML04_1317-LFML04_1319) are regarded as a putative type I secretion system. Four pairs of genes (LFML04_0190, LFML04_0191; LFML04_0636, LFML04_0637; LFML04_1550, LFML04_1551; LFML04_1790, LFML04_1791) encoding secretion protein, HlyD and outer membrane efflux proteins were identified, but the ATP-binding cassette (ABC) was not found. Three genes of the single secretion protein HlyD (LFML04_2033, LFML04_2281, LFML04_2400) were dispersed in the genome (Delepelaire, 2004).

A type II system, with Sec, Tat and Gsp pathways, was found. Sec and Tat pathways are responsible for the translocation of a substrate across the inner membrane, while the Gsp pathway is responsible for the translocation across the outer membrane. Whether the Sec or Tat pathway is chosen depends on the signal peptides of the substrates. The genes annotated for the Sec pathway are *secG* (LFML04_0073), *secA* (LFML04_0928), *ftsY* (LFML04_1042), *ffh* (LFML04_1095), *yajC*, *secD*, *secF* (LFML04_1175-LFML04_1177), *secY* (LFML04_1573), *secE* (LFML04_1594), and *yidC* (LFML04_2493). However, *secB* which encodes a protein export chaperone was not identified. The genes annotated for the Tat pathway are *tatC* (LFML04_2146), *tatA/E* (LFML04_2147), and another *tatA/E* (LFML04_1846). TatB was not found, so the Tat system is of the AC type in *L. ferriphilum* (Blaudeck *et al*., 2005). The proteins annotated for the Gsp pathway were the general secretion pathway protein JI (LFML04_1376, LFML04_1377), protein K (LFML04_0961), and protein GFEDC (LFML04_0966-LFML04_0970).

The type III secretion system of the flagella export apparatus were identified in the genome of *L. ferriphilum* ML-04 for transportation of flagella proteins across both membranes before assembly. This system consists of the integral membrane components FliO, FliP, FliQ, FliR, FlhB, FlhA (LFML04_0238-LFML04_0243), export ATPase (FliI, LFML04_0230), and its regulator FliH (LFML04_0229), and chaperone FliS (LFML04_2295) (Macnab, 2004).

In conclusion, the genome sequence of *L. ferriphilum* ML-04 has been determined, which is the first intact and pure culture genome sequence for *Leptospirillum* spp.. Through genome analysis, it is obvious that special metabolic pathways enable it to grow chemoautotrophically, and horizontal gene transfer played an important role in environmental adaptation of *L. ferriphilum* ML-04, a conclusion supported by the existence of two genomic islands and the mobile element of arsenic resistance. In addition, the complicated cell membrane, EPS biosynthesis ability and numerous genes encoding proteins that confer resistance to heavy metal ions, acidity and oxidative stress help it to adapt to the extreme growth conditions. The complete genome of *L. ferriphilum* deepens our understanding of *Leptospirillum* spp. and its metabolism and adaptation to the bioleaching environment.

Acknowledgements

This work was supported by the grants from the National Basic Research Program (2010CB630902), the National Natural Science Foundation (31070034, 30800011), and the Knowledge Innovation Program of CAS (2AKSCX2-YW-JS401) of China. The authors would like to thank Dr. Pamela Holt for proofreading the manuscript and Dr. Wei Liu for useful discussions.

References

Acuña, J., J. Rojas, A.M. Amaro, H. Toledo, and C.A. Jerez. 1992. Chemotaxis of *Leptospirillum ferrooxidans* and other acidophilic chemolithotrophs: comparison with the *Escherichia coli* chemosensory system. *FEMS Microbiol. Lett.* 75, 37-42.

Baker-Austin, C. and M. Dopson. 2007. Life in acid: pH homeostasis in acidophiles. *Trends Microbiol.* 15, 165-171.

Barreto, M., E. Jedlicki, and D.S. Holmes. 2005. Identification of a gene cluster for the formation of extracellular polysaccharide precursors in the chemolithoautotroph *Acidithiobacillus ferrooxidans*. *Appl. Environ. Microbiol.* 71, 2902-2909.

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573-580.

Blaudeck, N., P. Kreutzenbeck, M. Muller, G.A. Sprenger, and R. Freudl. 2005. Isolation and characterization of bifunctional *Escherichia coli* TatA mutant proteins that allow efficient Tat-dependent protein translocation in the absence of TatB. *J. Biol. Chem*. 280, 3426-3432.

Cabreiro, F., C.R. Picot, B. Friguet, and I. Petropoulos. 2006. Methionine sulfoxide reductases: relevance to aging and protection against oxidative stress. *Ann. NY Acad. Sci*. 1067, 37-44.

Castanie-cornet, M., T.A. Penfound, D. Smith, J.F. Elliott, and J.W. Foster. 1999. Control of acid resistance in *Escherichia coli*. *J. Bacteriol.* 181, 3525-3535.

Coram, N.J. and D.E. Rawlings. 2002. Molecular relationship between two groups of the genus *Leptospirillum* and the finding that *Leptosphillum ferriphilum* sp. nov. dominates South African commercial biooxidation tanks that operate at 40°C. *Appl. Environ. Microbiol.* 68, 838-845.

Dahl, C., A. Schulte, Y. Stockdreher, C. Hong, F. Grimm, J. Sander, R. Kim, S. Kim, and D.H. Shin. 2008. Structural and molecular genetic insight into a widespread sulfur oxidation pathway. *J. Math. Biol.* 384, 1287-1300.

Delepelaire, P. 2004. Type I secretion in gram-negative bacteria. *Biochim. Biophys. Acta.* 1694, 149-161.

Dixon, R. 1998. The oxygen-responsive NIFL-NIFA complex: a novel two-component regulatory system controlling nitrogenase synthesis in γ-Proteobacteria. *Arch. Microbiol.* 169, 371-380.

Ducluzeau, A., S. Ouchane, and W. Nitschke. 2008. The cbb3 oxidases are an ancient innovation of the domain bacteria. *Mol. Biol. Evol.* 25,1158-1166.

Edwards, K.J., P.L. Bond, and J.F. Banfield. 2000. Characteristics of attachment and growth of *Thiobacillus caldus* on sulphide minerals: a chemotactic response to sulphur minerals? *Environ. Microbiol.* 2, 324-332.

Eichhorn, E., J.R. van der Ploeg, and T. Leisinger. 2000. Deletion analysis of the *Escherichia coli* taurine and alkanesulfonate transport systems. *J. Bacteriol.* 182, 2687-2695.

Goltsman, D.S.A., V.J. Denef, S.W. Singer, N.C. VerBerkmoes, M. Lefsrud, R.S. Mueller, G.J. Dick, and *et al*. 2009. Community genomic and proteomic analysis of chemoautotrophic, iron-oxidizing "*Leptospirillum rubarum*" (Group II) and "*Leptospirillum ferrodiazotrophum*" (Group III) in acid mine drainage biofilms. *Appl. Environ. Microbiol.* 75, 4599-4615.

Hanson, T.E. and F.R. Tabita. 2001. A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from *Chlorobium tep-*

idum that is involved with sulfur metabolism and the response to oxidative stress. *Proc. Natl. Acad. Sci. USA* 98, 4397-4402.

Hutchins, A.M., J.F. Holden, and M.W.W. Adams. 2001. Phosphoenolpyruvate synthetase from the hyperthermophilic archaeon *Pyrococcus furiosus. J. Bacteriol.* 183, 709-715.

Jeans, C., S.W. Singer, C.S. Chan, N.C. VerBerkmoes, M. Shah, R.L. Hettich, J.F. Banfield, and M.P. Thelen. 2008. Cytochrome 572 is a conspicuous membrane protein with iron oxidation activity purified directly from a natural acidophilic microbial community. *ISME J.* 2, 542-550.

Jiao, Y. and D.K. Newman. 2007. The *pio* operon is essential for phototrophic Fe(II) oxidation in *Rhodopseudomonas palustris* TIE-1. *J. Bacteriol.* 189, 1765-1773.

Kawasaki, S., J. Ishikura, D. Chiba, T. Nishino, and Y. Niimura. 2004. Purification and characterization of an H_2O-forming NADH oxidase from *Clostridium aminovalericum*: existence of an oxygen-detoxifying enzyme in an obligate anaerobic bacteria. *Arch. Microbiol.* 181, 324-330.

Kerscher, L., S. Nowitzki, D. Oesterhelt. 1982. Thermoacidophilic archaebacteria contain bacterial-type ferredoxins acting as electron acceptors of 2-oxoacid:ferredoxin oxidoreductases. *Eur. J. Biochem.* 128, 223-230.

Kurtz, D.M., Jr. 2006. Avoiding high-valent iron intermediates: Superoxide reductase and rubrerythrin. *J. Inorg. Biochem.* 100, 679-693.

Leigh, J.A. and J.A. Dodsworth. 2007. Nitrogen regulation in bacteria and archaea. *Annu. Rev. Microbiol*. 61, 349-377.

Leverrier, P., J.P. Vissers, A. Rouault, P. Boyaval, and G. Jan. 2004. Mass spectrometry proteomic analysis of stress adaptation reveals both common and distinct response pathways in *Propionibacterium freudenreichii*. *Arch. Microbiol*. 181, 215-230.

Levican, G., J.A. Ugalde, N. Ehrenfeld, A. Maass, and P. Parada. 2008. Comparative genomic analysis of carbon and nitrogen assimilation mechanisms in three indigenous bioleaching bacteria: predictions and validations. *BMC Genomics* 9, 581.

Li, B., J. Lin, S. Mi, and J. Lin. 2010. Arsenic resistance operon structure in *Leptospirillum ferriphilum* and proteomic response to arsenic stress. *Bioresour. Technol*. 101, 9811-9814.

Macnab, R.M. 2004. Type III flagellar protein export and flagellar assembly. *Biochim. Biophys. Acta.* 1694, 207-217.

Mai, X. and M.W. Adams. 1996. Characterization of a fourth type of 2-keto acid-oxidizing enzyme from a hyperthermophilic archaeon: 2-ketoglutarate ferredoxin oxidoreductase from *Thermococcus litoralis*. *J. Bacteriol.* 178, 5890-5896.

Maklashina, E., D.A. Berthold, and G. Cecchini. 1998. Anaerobic expression of *Escherichia coli* succinate dehydrogenase: functional replacement of fumarate reductase in the respiratory chain during anaerobic growth. *J. Bacteriol*. 180, 5989-5996.

Michels, M. and E.P. Bakker. 1985. Generation of a large, protonophore-sensitive proton motive force and pH difference in the acidophilic bacteria *Thermoplasma acidophilum* and *Bacillus acidocaldarius*. *J. Bacteriol.* 161, 231-237.

Moat, A.G., J.W. Foster, and M.P. Spector. 2002. Microbial Physiology, fourth ed. Wiley-Liss, New York, NY, USA.

Morgan, R.W., M.F. Christman, F.S. Jacobson, G. Storz, and B.N. Ames. 1986. Hydrogen peroxide-inducible proteins in *Salmonella typhimurium* overlap with heat shock and other stress proteins. *Proc. Natl. Acad. Sci. USA* 83, 8059-8063.

Muro-Pastor, M.I., J.C. Reyes, and F.J. Florencio. 2005. Ammonium assimilation in cyanobacteria. *Photosynth. Res*. 83, 135-150.

Preiss, J. 1984. Bacterial glycogen synthesis and its regulation. *Annu. Rev. Microbiol.* 38, 419-458.

Quatrini, R., C. Appia-Ayme, Y. Denis, E. Jedlicki, D.S. Holmes, and V. Bonnefoy. 2009. Extending the models for iron and sulfur oxidation in the extreme Acidophile *Acidithiobacillus ferrooxidans*. *BMC Genomics* 10, 394.

Rawlings, D.E., H. Tributsch, and G.S. Hansford. 1999. Reasons why '*Leptospirillum*'-like species rather than *Thiobacillus ferrooxidans* are the dominant iron-oxidizing bacteria in many commercial processes for the biooxidation of pyrite and related ores. *Microbiology* 145, 5-13.

Seibold, G.M., M. Wurst, and B.J. Eikmanns. 2009. Roles of maltodextrin and glycogen phosphorylases in maltose utilization and glycogen metabolism in *Corynebacterium glutamicum*. *Microbiology* 155, 347-358.

Simon, J., R. Gross, O. Einsle, P.M.H. Kroneck, A. Kroger, and O. Klimmek. 2000. A NapC/NirT-type cytochrome *c* (NrfH) is the mediator between the quinone pool and the cytochrome *c* nitrite reductase of *Wolinella succinogenes*. *Mol. Microbiol.* 35, 686-696.

Singer, S.W., C.S. Chan, A. Zemla, N.C. VerBerkmoes, M. Hwang, R.L. Hettich, J.F. Banfield, and M.P. Thelen. 2008. Characterization of cytochrome 579, an unusual cytochrome isolated from an iron-oxidizing microbial community. *Appl. Environ. Microbiol.* 74, 4454-4462.

Szurmant, H. and G.W. Ordal. 2004. Diversity in chemotaxis mechanisms among the bacteria and archaea. *Microbiol. Mol. Biol. Rev.* 68, 301-319.

Tuffin, I.M., S.B. Hector, S.M. Deane, and D.E. Rawlings. 2006. Resistance determinants of a highly arsenic-resistant strain of *Leptospirillum ferriphilum* isolated from a commercial biooxidation tank. *Appl. Environ. Microbiol.* 72, 2247-2253.

Valdes, J., F. Veloso, E. Jedlicki, and D. Holmes. 2003. Metabolic reconstruction of sulfur assimilation in the extremophile *Acidithiobacillus ferrooxidans* based on genome analysis. *BMC Genomics* 4, 51.

van der Ploeg, J.R., E. Eichhorn, and T. Leisinger. 2001. Sulfonate-sulfur metabolism and its regulation in *Escherichia coli*. *Arch. Microbiol*. 176, 1-8.

Walkup, L.K.B. and T. Kogoma. 1989. *Escherichia coli* proteins inducible by oxidative stress mediated by the superoxide radical. *J. Bacteriol.* 171, 1476-1484.

Wong, H.C., A.L. Fear, R.D. Calhoon, G.H. Eichinger, R. Mayer, D. Amikam, M. Benziman, and *et al*. 1990. Genetic organization of the cellulose synthase operon in *Acetobacter xylinum. Proc. Natl. Acad. Sci. USA* 87, 8130-8134.

Yun, N.R., M. Yamamoto, H. Arai, M. Ishii, and Y. Igarashi. 2002. A novel five-subunit-type 2-oxoglutalate:ferredoxin oxidoreductases from *Hydrogenobacter thermophilus* TK-6. *Biochim. Biophys. Res. Commun.* 292, 280-286.

Zhang, Q., T. Iwaaki, T. Wakagi, and T. Oshima. 1996. 2-Oxoacid:ferredoxin oxidoreductase from the thermoacidophilic archaeon, *Sulfolobus* sp. strain 7. *J. Biochem.* 120, 587-599.