Meta-genomic analysis of *Halobacillus trueperi* S61 isolated from the Qarhan Salt Lake

Shuo Shen

Qinghai University

Wei Li (✉ lwbabylw@163.com)

Qinghai University

Jian Wang

Qinghai University

Ruiting Xu

Wuhan Metware Biotechnology Co., Ltd

Article

Keywords:

Posted Date: July 26th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1377706/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Halophilic microbial as prospective resources of biotechnology due to the advantages of flexible survivability. Qarhan Salt Lake is the second largest Salt Lake in the world which contains rich-unique extremophiles and deserved in-depth exploration. Present study first time isolated novel strain *Halobacillus trueperi* S61 from Qarhan Salt Lake and performed whole-genome sequencing through combined third-generation PacBio and second-generation Illumina technology. The whole genome of *Halobacillus trueperi* S61 identified 57549 total reads and consists a complete circular chromosome of 4047887 bp with 43.86% GC content without gaps. Total number of 139 non-coding RNA (included 86 tRNA, 30 rRNA and 23 sRNA), 16 gene islands with 260275 bp and two prophages (with 82682 length) were predicted. In addition, the whole genome of *Halobacillus trueperi* S61 summarized basic annotation for 3982 protein-coding genes, 3980, 3667, 2998 and 2303 unigenes were annotated with Nr, Swissport, KOG and KEGG database. Combined with advanced analysis, 561 carbohydrate enzymes and 4416 pathogen host interactions related genes were identified. The protein function of *Halobacillus trueperi* S61 was mainly focus on biological processes, and the protein function was mainly distributed in gene transcription and amino acids, and carbohydrates metabolism. The complete whole genome sequence assembly and annotation of novel strain *Halobacillus trueperi* S61 isolated from Qarhan Salt Lake mainly focus on protein biological processes and antibiotic resistance, provides a potential resource for biotechnology.

Introduction

The marine environment as broadly source of microbes, contains numerous halotolerant or psychrophilic microorganisms¹⁻². Microorganisms inevitable evolve physiological and genomic adaptability to adopt extreme conditions, among which Halophilic microbes' advantages of flexible survivability is valuable characteristics and prospects in biotechnology³⁻⁴. For instance, *Halobacillus* members are important sources for the halotolerant extracellular enzymes production in industry, *Halobacillus trueperi* RSK CAS9 was optimized lipase production in marine fish industry⁵⁻⁶.

Noteworthy, *Halobacillus trueperi* are considered as moderately halophilic ($0.5-2.5 \text{ mol}\cdot\text{L}^{-1}$, aerobic and heterotrophic) and first isolated from the Great Salt Lake (Utah)⁷. Since then, *Halobacillus trueperi* attracted more attention and researchers have successively carried out, Lu et al.⁸ isolated *Halobacillus trueperi* from the salt water in the western Himalayas, and reported that *H. trueperi* DSM10404 could be accumulate glycine, glutamate and betaine as salt-tolerant compatible solutes. Gupta et al.⁹ isolated *Halobacillus trueperi* SS1 from Lunsu saltwater and Kharangate-Lad et al.¹⁰ isolated *Halobacillus trueperi* MXM-16 from mangrove plant litter which capable to produce hydroxamate siderophore and carotenoid pigment to chelate iron. *Halobacillus trueperi* ATCC 700077 isolated from the solid and liquid salinities in southern Sahara region of Tunisia and reported as major ecosystem-adaptive microorganism¹¹. Although several halophilic bacteria have been widely reported, while the unique characteristics presented in different natural environments. Importantly, Qarhan Salt Lake as the second largest Salt Lake in the world

and largest in China, contained rich and unique halophilic microbial resources which need to be in-depth explored with broad prospects¹².

With the development of biotechnology, researchers identified the organisms with emerged techniques¹³. Currently, whole-genome sequencing as novel and culture-independent technology for explore the genetic diversity and evolutionary history of microorganisms¹⁴. Various genomic project has been carried out¹⁵⁻²⁰. Although the second-generation Illumina sequencing has large sequencing throughput, high sequencing accuracy and low cost, while its read length is relatively short (150-400bp). The third-generation single-molecule real-time sequencing Pacbio advantaged in ultra-long sequencing read length (average of 10-12 kb and longest exceed 30 kb), high throughput (One SMRT cell on the RS II platform generate 0.5~1Gb data, while one SMRT cell on the Sequel platform generate 5~10Gb data), no GC bias (unrequire PCR amplification) and direct detection of various types of DNA methylation. The latest advanced sequencing technology will be provided novel insights into the metabolic profile of microorganisms²¹⁻²³.

Based on first time isolated novel strain from Qarhan Salt Lake and identified as *Halobacillus trueperi* S61, this study performed whole-genome sequencing through combined exert the advantage of third-generation PacBio and second-generation Illumina technology, to obtained completer and more accurate and reliable genome assembly and metabolic diversity for understanding the genome characteristics of *Halobacillus trueperi* S61, aimed at contributed to genetics and potential biotechnology applications.

Materials And Methods

Sample collection and pretreatment as well as microbe isolation. Fresh water and soil were collected from Qarhan Salt Lake located at Qinghai, Tibet Plateau, China. Fifteen sampling points were selected and each point was collected according to the five-point sampling method. All samples kept in portable refrigerator and transported to the laboratory for pretreatment. Pretreatment of water samples was mixed water samples and filtered by filter with 0.22 aperture and the bacteria were enriched on the filter membrane under sterile conditions. When the amount of water reaches 30 ml, removed the filter membrane and placed in a glass test tube contained 3 ml sterile seawater, which is 10^{-1} water sample. Pretreatment of soil samples was taken appropriate amount of soil sample, air dried (about 20 d) and treated at 120 °C for 1 h. Then weigh 10g treated soil sample and added 90ml sterile seawater, put into the sterilized triangle flask with glass balls and fully shake 30min, then absorb the supernatant after standing. The supernatant was 10^{-1} soil sample. Diluted the above 10^{-1} samples successively to 10^{-2} and 10^{-3} times, and 150ul gradient samples were taken and coated on ATCC213 medium plate ($\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 10g, $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ 0.2g, peptone 2.5g, yeast extract 10g, KCl 5g, NaCl 30g, and agar powder 12g, then add distilled water until 1000 mL and adjust pH about 7.2-7.4) with three repetitions. Followed by placed upside down the coated plate in incubator and cultivated at 28°C and 37°C respectively. After observed colonies, pick up single colony and purification to obtained purified strains.

Molecular Identification and genome analysis of *Halobacillus trueperi* S61. *Halobacillus trueperi* S61 was isolated and purified by dilution-plate enrichment culture. Morphologically, it was Gram-positive, spherical, capsular, and peritrichous flagella, with size of 0.6-0.8 μm $\times$ 0.4-0.6 μm . Molecular identification was performed from DNA extraction according to Sangon Biotech Co., Ltd. column bacterial DNA extraction kit procedures (Shanghai, China). The bacterial universal primers were F27 and P1541 (5'-AGAGTTTGATCCTGGCTCAGG-3' and 5'-AAGGAGGTGGTGGTATGCCGCA-3'). The reaction conditions were 94°C denaturation for 45s, 50 °C annealing for 45 s, 72 extensions for 75 s, and 50 μl reaction system 30 cycles. The PCR products were detected by agarose gel electrophoresis, and the sequence results obtained by cloning then uploaded sequence to <https://www.ezbiocloud.net> for comparison. Finally, the matched genus was determined and named *Halobacillus trueperi* S61 (preservation number GDMCC No: 60078).

The genome sequencing carried out by third-generation PacBio and the second-generation Illumina with assistance of Genedenovo Biotechnology Co., Ltd (Guangzhou, China). Firstly, extract the DNA and quality control by concentration and electrophoresis test. Then, RS II and Sequel of Pacific Biosciences were used for amplification based on Single Molecule Real Time. After constructed library, perform quality detection by Qubit and Agilent 2100 was used to evaluate the insert size and carried out PacBio sequencing. At the same time, Illumina sequencing was performed using Hiseq X10 after library construction and detection.

Genome assembly and function annotation. Genome assembly was performed by third-generation sequencing data, and then use the second-generation data to correct the assembly results. Genome component analysis and functional annotation were analyzed based on the corrected assembly results. Firstly, quality control of sequencing data, filtered the raw data of Pacbio and Illumina sequencing to obtain clean data. Then, genome assembly by use of Falcon to splice and assemble the third-generation sequencing reads, compare the original sequencing reads to the assembly results, calculate the coverage and GC distribution to evaluate the assembly quality. According to the assembled genome sequences, combined with the prediction results of coding genes, the genome circle diagram was drawn to comprehensively display the characteristics of the genome.

Followed by genome component analysis, using National Center for Biotechnology Information Search database for coding gene prediction, using RNAmmer, tRNAscan, cmscan to compare Rfam database for non-coding ribosomal RNA (rRNA), transfer RNA (tRNA) and small RNA (sRNA) prediction. Further predict the Interspersed and Tandem repeat sequence of bacterial genome by Repeat Masker and TRF software. Use CRISPR finder to perform Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) prediction on the genome. Use Transposon PSI (version: 20100822), Island Viewer4 and Phage_Finder to perform transposons, Gene Islands (GIs) and prophage prediction on the genome.

Additionally, analyzed basic and advanced function annotation. The basic function annotation included non-redundant protein database and SwissProt by use blastp and diamond to compare the amino acid sequence encoded by the gene with the database, as well as through Kyoto encyclopedia of genes and

genomes (KEGG), Gene Ontology (GO), Cluster of Orthologous Groups of proteins (COG) database to obtain the annotation results corresponding to the genes and classify accordingly. For advanced analysis, Use Pfam Scan (<https://www.ebi.ac.uk/Tools/pfa/pfamscan/>) perform protein families database of alignments and hidden Markov models (Pfam) to provide complete and accurate protein family and domain classification information. Use blastp for pathogen host interactions (PHI-base) and carbohydrate-active enzymes database (CAZy) analysis. Use SignalP 4.1 to analyze the protein sequence of the predicted gene, identify the signal protein, and predict the transmembrane protein and effector protein through TMHMM and EffectiveT3. Use resistance gene identifier (RGI), blastn and antismash 4.1.0 to predict antibiotic resistance ontology, virulence factors of pathogenic bacteria (VFDB) and secondary metabolism gene clusters.

Results And Discussion

The Genome assembly and component characteristics of the genome *Halobacillus trueperi* S61. The second-generation sequencing technology combined with the third-generation sequencing technology was used for deep sequencing of *Halobacillus trueperi* S61, and a detailed map of a circular chromosome was obtained. The whole genome of *Halobacillus trueperi* S61 identified 57549 total reads by QC-PacBio and consists a complete circular chromosome of 4047887 bp with 43.86% GC content without gaps (Fig.1). The GC-Depth demonstrated that the strains *Halobacillus trueperi* S61 shown Poisson distribution, GC without obvious bias and appeared a scattered area between 20-30% GC content, which may be affected by mitochondrial DNA (Fig. 2a). Gupta et al.²⁴ identified the genome *Halobacillus trueperi* SS1 has 4329 sequences with 4.14Mbp and 42.15% GC content as well as 35 RNA genes.

The genome *Halobacillus trueperi* S61 predicted total number of 3982 nucleotides with total length 3567510 and 44.57% GC content. The non-coding RNA (ncRNA) was non-encode proteins while perform various biological functions in life activities at the RNA level²⁵. In this study, 139 non-coding RNA was identified included 86 tRNA, 30 rRNA and 23 sRNA (Table 1). Among them, the largest number was tRNA that sequence length accounts for 0.16% of the total sequence length, indicated the important role of tRNA contributed to expression and regulation in *Halobacillus trueperi* S61 cells (Fig. 2b). The repeated sequence as component of the gene regulatory network which affects the evolution, heredity, and mutation of life²⁶. This study predicted total 58 interspersed repeats with 3909 bp and five types, the largest number of elements were 28 and 18 in LINEs and SINEs with 1856 bp and 1149 bp. While less proportion in DNA and LTR elements with 497bp and 218bp. In addition, three types of transposons (helitronORF and LINE) were predicted.

The clustered regularly interspaced short palindromic repeats (CRISPR) as genetic weapon or natural immune system of most bacteria and archaea, since resistance to extraneous plasmids and phage sequence²⁷. Two kinds of CRISPRs were predicted in the genome *Halobacillus trueperi* S61, Crispr 1 (AGAAAACAAAACCAACAATCAGCTG) and Crispr 2 (TGATGGGAATCGAACCCACGACAT) indicated that strain *Halobacillus trueperi* S61 provide the corresponding acquired immunity to the host through CRISPR pathway. Gene islands (GI) considered as mobile genetic elements due to related with various biological

functions, especially the horizontal transfer of genes²⁸. These predicted GI regions might be contained *Halobacillus trueperi* S61 antibiotic resistance genes and bacteriostatic gene fragments. Total 16 gene islands with 260275 bp were predicted in the whole genome of *Halobacillus trueperi* S61, which may support microbial adaptability to distinct abiotic stresses and antimicrobial resistance environments. In addition, Prophage as carrier of genetic information could be integrated with infected microbes' genome after infection. Previous studies found that bacteriophages were able to dissolve certain pathogenic microbes which may beneficial for diseases curing, while at the same time may also dissolve beneficial or other harmful microbes. Thus, it is widely used as carrier for horizontal transfer of beneficial microbes²⁹. Present study identified two prophages with total length 82682 in *Halobacillus trueperi* S61, contained 43.86 CDs of 44 and 61 genes with 44.85 and 38.43% GC respectively. Therefore, we speculate that *Halobacillus trueperi* S61 has ability to lysis pathogen which need further explored.

The Basic function Annotation of genome *Halobacillus trueperi* S61. The whole genome sequence of *Halobacillus trueperi* S61 has summarized basic annotation for 3982 protein-coding genes. In order to improve the functional prediction, 3980, 3667, 2998 and 2303 unigenes were annotated with Nr, Swissport, KOG and KEGG database.

Specifically, in the whole genome sequence of *Halobacillus trueperi* S61, 3668 genes have been annotated with COG function database (Fig. 3). The protein function mainly distributed in 9.95% amino acid transport and metabolism (E), 8.02% carbohydrate transport and metabolism (G), 8.53% transcription (K), the number of genes were 365, 294, and 313 respectively. There were 13.71% genes for general function prediction only, and 9.13% of gene protein functions were unknown which require for further evaluation. In addition, we found that 107 genes were involved in secondary metabolites related biosynthesis transport and catabolism and 222 genes related with inorganic ion transport and metabolism (P), while other categories account less proportion. There were 7829 genes with GO annotation function (Fig. 4) and demonstrated that biological process account for 52% and dominant by metabolic process, cellular and single-organism process (1016, 953 and 752), as well as localization and biological regulation (318 and 298). Additionally, molecular function account for 23% and affiliated to catalytic activity and binding (888 and 657), followed by transporter activity and nucleic acid binding transcription factor activity (114 and 103). As to part of cellular component account for 25% and most distributed in membrane, membrane part and cell (537, 484 and 376). This result indicated that the gene products of strain *Halobacillus trueperi* S61 was mainly focus on biology process.

The KEGG pathway annotation could be identified the functional genes up or down-regulated in target metabolites³⁰⁻³¹. In this study, total 3672 genes of *Halobacillus trueperi* S61 were annotated with KEGG and annotated as five types included 80.94% metabolism, 8.66% environmental information processing, 5.80% genetic information processing, 4.36% cellular processes, 0.25% organismal systems. In the metabolic pathway, 602 related with metabolic, 274 and 215 related with biosynthesis of secondary metabolites and antibiotics as well as 183 related with microbial metabolism in diverse environments, 117 biosynthesis of amino acids and 96 carbon metabolisms identified. Furthermore, 76 genes were related with microbial viability (replication and repair) and reflected to 6 KEGG pathways, indicated

Halobacillus trueperi S61 may play a role in homologous recombination, mismatch repair, DNA replication, base excision repair, nucleotide excision repair, and non-homologous end-joining metabolic pathways³². Additionally, Nr database annotated total number of 3980 genes and among which 88.32% matched with *Bacillus subtilis* (number of 3515), 134 and 130 genes belong to species *Bacillus sp. EGD-AK10* and *Streptococcus pneumoniae*, followed by 66 *Bacillus sp. YP1*, 22 *Bacillus sp. CMAA 1185*, 17 *Bacillus sp. LM 4-2*, 13 *Bacillus* and *Bacillus sp. JS*. Overall, the GO, COG, KEGG, and Nr annotations of the protein-coding genes indicated that the protein function of *Halobacillus trueperi* S61 was mainly focus on biological processes, and the protein function was mainly distributed in gene transcription and amino acids, and carbohydrates metabolism.

In addition, bacterial comparative genomic analysis was performed between genome *Bacillus velezensis* (Hal61) and the target genome *Halobacillus trueperi* (Htr), *Halobacillus litoralis* (Hli), *Halobacillus kuroshimensis* (Hku), *Halobacillus dabanensis* (Hda), *Rossellomorea vietnamensis* (Rvi). The results of parallel collinearity and two-dimensional collinearity showed that the coverage of Rvi and Hal61 was higher than other genomes (Fig. 5 ac), and the gene family analysis showed that Total genes of Htr, Hli, Hku, Hda, Rvi and Hal61 were 3820, 4065, 3872, 3981, 4319, and 3982 with gene families of 474, 463, 506, 407, 1393, and 1290 (Fig. 5 b). Overall, *Bacillus velezensis* is more closely related to *Rossellomorea vietnamensis*.

The Advanced function Annotation of genome *Halobacillus trueperi* S61. The carbohydrate enzymes (CAZymes) as essential factor when pathogens pass through the primary barrier cell wall after the host attacked, and contributed to carbohydrates, glycoconjugates biosynthesis and decomposition³⁰. The *Halobacillus trueperi* S61 contains a total of 561 CAZymes (Fig. 6). Among them, glycoside hydrolases (GH) had highest content accounted for 35.29% of all carbohydrate enzymes, and the rest were glycosyltransferases (GT), carbohydrate-binding modules (CBM), carbohydrate esterases (CE), auxiliary activities (AAs), and polysaccharide lyases (PL), which accounted for 31.37%, 19.61%, 12.66%, 0.89% and 0.18%, respectively. Importantly, GT and GH play vital role in the metabolism process since GT associated with nucleotide and amino sugar metabolism and GH associated with glycogen, maltose, and N-acetylglucosamine degradation, that might be favorable to nutrient acquisition and maintain structure for survivability of strain *Halobacillus trueperi* S61 in salt sea³². Woo et al.³³ annotated *Halobacillus mangrovi* KTB 131 genome and pointed out that the most strains distributed in secondary metabolite biosynthesis, catabolism and transport. Additionally, secreted proteins involved enzymes, antibodies and some hormones, total of 3982 proteins were predicted with 273 signal, 138 transmembrane and 135 secret proteins. Effector protein as critical point in bacterial secretion systems, pathogens secrete effector proteins into the extracellular or host cells through TNSS (type N secretion systems, type I-VII), which affect various important activities such as immune response and cell death in the cell process and caused pathological reactions. There was identified four symbol of effectors included yfjA (Hal61 00834), yueC (Hal61 03072), yueB (Hal61 03073), yukC (Hal61 03075), which all belong to *Bacillus subtilis*.

The pathogenic host interaction gene database pathogen host interactions (PHI) included diverse pathogenic genes related to different types of hosts, its crucial to find target genes for drug

intervention²⁹. Through gene annotation, strain *Halobacillus trueperi* S61 has a total of 4416 PHI-related genes, which most dominant pathogen species distributed in *Burkholderia glumae* that caused bacterial grain rot disease with function of DNA gyrase (bacterial topoisomerase II), followed by *Flavobacterium psychrophilum* (DNA gyrase), *Cryptococcus neoformans* (GTP Biosynthesis), *Bacillus anthracis* (Tellurite Resistance), *Pectobacterium wasabiae* (Posttranscriptional regulator) which caused bacterial cold-water disease, meningoencephalitis, anthrax and soft rot. Among which, 742 pathogenic factor genes come from *Magnaporthe oryzae* (related to *Magnaporthe grisea*), 367, 262, 216 and 208 pathogenic related with *Fusarium graminearum* (related to *Gibberella zeae*), *Aspergillus fumigatus*, *Alternaria alternata* and *Candida albicans*. Moreover, the virulence factors of pathogenic bacteria (VFDB) database annotated 15 factors in form of *Listeria monocytogenes*, *Legionella pneumophila Philadelphia*, *Chlamydia trachomatis*, *Salmonella enterica*, *Escherichia coli*, *Bacillus anthracis*, *Bacillus anthracis*, and *Mycobacterium tuberculosis*. In addition, the prediction results of the secondary metabolism gene cluster showed ten gene cluster types, comprised *nrps*, *terpene*, *nrps-transatpks-otherks*, *t3pks*, *lantipeptide*, and *sactipeptide-head_to_tail*.

CARD was used to associate antibiotic modules and their targets, resistance mechanisms, and gene mutations²⁸. There was predicted 11 efflux pump complex or subunit conferring antibiotic resistance included *ImrB*, *ykkD*, *TaeA*, *sav1866*, *ykkC*, *ImrD*, *TriC*, *bmr*, and *blt*. Four antibiotic inactivation enzymes included *aadK*, *VgbC*, *rphB*, *BLA1*, and *mphI*, as well as antibiotic target protection protein (*mfd*). Antibiotic resistant gene *Enterococcus faecium* *cls* conferring resistance to daptomycin, antibiotic resistant *fabI*, *mecA*, *Bacillus subtilis* *mprF*, *Escherichia coli* *EF-Tu* mutants conferring resistance to kirromycin, *Staphylococcus aureus* *rpoB* mutants conferring resistance to rifampicin, *Mycobacterium tuberculosis* intrinsic *murA* conferring resistance to Fosfomycin as well as determinant of resistance to nucleoside antibiotic (*tmrB*). Treves et al.³⁴ evaluated the draft genome of *Halobacillus* sp. BBL2006 identified 4331 open reading frames and comprised of heavy metals and antibiotic resistance genes. Although the coding genes were annotated from different databases while the phenomenon reflected was consistent, mainly distributed in protein biological processes and antibiotic resistance which provides a potential resource for biotechnology.

Conclusions

The whole genome assembly and annotation of novel strain *Halobacillus trueperi* S61 isolated from Qarhan Salt Lake were performed. The genome *Halobacillus trueperi* S61 predicted total number of 3982 nucleotides with 3567510 length and 44.57% GC content. The 3668 genes have been annotated with COG and protein function mainly distributed in 9.95% amino acid transport and metabolism (365 genes). There were 7829 genes annotated with GO annotation function and biological process account for 52% and dominant by metabolic process (1016 genes). Total 3672 genes were annotated with KEGG and dominant by metabolism (80.94%), and Nr database annotated 3980 genes and 88.32% matched with *Bacillus subtilis*. The *Bacillus velezensis* is more closely related to *Rosellomorea vietnamensis*. Overall, the strain *Halobacillus trueperi* S61 was mainly focus on biological processes.

Abbreviations

Not applicable

Declaration

Ethics approval and consent to participate

Not applicable

Consent for publication

None

Availability of data and materials

<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA795248>

Accession: PRJNA795248

Accession: PRJNA795248 ID: 795248

Prediction of secondary metabolites

Prediction of secondary metabolites

Accession	PRJNA795248
Data Type	Raw sequence reads
Scope	Multispecies
Submission	Registration date: 6-Jan-2022 Qinghai Academy of Agriculture and Forestry Sciences
Relevance	Agricultural

Competing interests

Authors have no competing interest

Funding

General Project of Natural Science Foundation of Qinghai Science and Technology Department (2019-ZJ-914), National Modern Agricultural Technology System (CARS-10).

Authors' contributions

Dr. Shuo Shen: He has design whole experiment and contribution to data correction and formal analysis. In addition, he write whole manuscript.

Prof. Wei Li: He has conceptualization, design and supervision of review and also writing - original draft this article. In addition, funding acquisition and project administration.

Prof. Jian Wang: He has conceptualization, design and supervision of review and also writing - original draft this article. In addition, funding acquisition and project administration.

Mr. Ruiting Xu: He has conceptualization, design and supervision of review and also writing - original draft this article. In addition, funding acquisition and project administration.

Acknowledgements

We are grateful for the support of funding General Project of Natural Science Foundation of Qinghai Science and Technology Department (2019-ZJ-914), National Modern Agricultural Technology System (CARS-10).

Authors information (optional)

Dr. Shuo Shen: fjzss@126.com

Prof. Wei Li: lwbabylw@163.com

Prof. Jian Wang: qhxnss@126.com

Mr. Ruiting Xu: 1091932214@qq.com

*Corresponding authors: Prof. Wei Li and Prof. Jian Wang.

E-mail: lwbabylw@163.com

References

1. Park, Y., Choi, T., Han, Y., Song, H., Park, J., Bhatia, S., Gurav, R., Choi, K., Kim, Y., & Yang, Y. Effects of osmolytes on salt resistance of *Halomonas socia* CKY01 and identification of osmolytes-related genes by genome sequencing. *J. Biotechnol.* 322, 21-28 (2020).
2. Poli, A., Finore, I., Romano, I., Gioiello, A., Lama, L., & Nicolaus, B. Microbial diversity in extreme marine habitats and their biomolecules. *Microorganisms.* 5, 25 (2017).
3. Hong, J., Song, H., Moon, Y., Hong, Y., Bhatia, S., Jung, H., Choi, T., Yang, S., Park, H., & Choi, Y. Polyhydroxybutyrate production in halophilic marine bacteria *Vibrio proteolyticus* isolated from the Korean peninsula. *Bioprocess. Biosyst. Eng.* 42, 603-610 (2019).
4. Zhang, H., Fang, J., Zhang, H., & Cao, J. Complete genome sequence of a psychrotolerant and piezotolerant bacterium *Parasedimentitalea marina* W43T, isolated from deep sea water of the New

- Britain trench. *Mar. Geno.* 61, 100915 (2022).
5. Treves, D., Francis, J., & Kirchner, G. Draft genome sequence of the moderately halophilic bacterium *Halobacillus* sp. BBL2006. *Data. Brief.* 21, 2410-2413 (2018).
 6. Sathishkumar, R., Ananthan, G., Iyappan, K., & Stalin, C. A statistical approach for optimization of alkaline lipase production by ascidian associated-*Halobacillus trueperi* RSK CAS9. *Biotechnol. Rep.* 8, 64-71 (2015).
 7. Spring, S., Ludwig, W., Marquez, M., Ventosa, A., & Schleifer, K. *Halobacillus* gen. nov. with description of *Halobacillus litoralis* sp. nov. and *Halobacillus trueperi* sp. nov., and transfer of *Sporosarcina halophila* to *Halobacillus halophilus* comb. nov. *Int. J. Syst. Bacteriol.* 46, 492-496 (1996).
 8. Lu, W., Zhao, B., Feng, D., & Yang, S. Cloning and characterization of the *Halobacillus trueperi* *bethH* gene, encoding the transport system for the compatible solute glycine betaine. *FEMS. Microbiol. Lett.* 235, 393-399 (2004).
 9. Gupta, S., Sharma, A., Dev, A., Baumler, D., & Sourirajan, A. Draft Genome Sequence of *Halobacillus trueperi* SS1, Isolated from Lunsu, a Saltwater Body in the Northwest Himalayas. *Microbiol. Resour. Ann.* 8, e01710-18 (2019).
 10. Kharangate-Lad, A., & Bhosle, S. Studies on Siderophore and Pigment Produced by an Adhered Bacterial Strain *Halobacillus trueperi* MXM-16 from the Mangrove Ecosystem of Goa, India. *Indian. J. Microbiol.* 56, 461-466 (2016).
 11. Rivadeneyra, M., Parraga, J., Delgado, R., Cormenzana, A., & Delgado, G. Biomineralization of carbonates by *Halobacillus trueperi* in solid and liquid media with different salinities. *Fems. Microbiol. Ecol.* 48, 39-46 (2004).
 12. Li, R., Liu, C., Jiao, P., Liu, P., Tang, D., & Wang, S. The effect of solvent chemistry on potassium dissolution extraction from low-grade solid potash ore in Qarhan Salt Lake, China. *Appl. Geochem.* 115, 104550 (2020).
 13. Shaikh, K., Kumar, P., Nesamma, A., Abdin, M., & Jutur, P. Hybrid genome assembly and functional annotation reveals insights on lipid biosynthesis of oleaginous native isolate *Parachlorella kessleri*, a potential industrial strain for production of biofuel precursors. *Algal. Res.* 52, 102118 (2020).
 14. Thirugnanasambandam, R., Inbakandan, D., Abraham, L., Kumar, C., Sundaram, S., Subashni, B., Vasantharaja, R., Kumar, A., Kirubakaran, R., Khan, S., & Balasubramanian, T. De novo assembly and annotation of the whole genomic analysis of *Vibrio campbellii* RT-1 strain, from infected shrimp: *Litopenaeus vannamei*. *Microb. Pathogenesis.* 113, 372-377 (2017).
 15. Wang, J., Yang, C., Zhang, C., Mao, X., & Li, Z. Complete genome sequence of the *Clostridium difficile* LCL126. *Bioengineered.* 12, 745-754 (2021).
 16. Xu, P., Zhang, X., Su, H., Liu, X., & Hong, G. Genome-wide analysis of PYL-PP2C-SnRK2s family in *Camellia sinensis*. *Bioengineered.* 11, 103-115 (2020).
 17. Wang, Y., Ma, L., He, J., Liu, Z., Weng, S., Wang, L., He, J., & Guo, C. Whole genome sequencing and comparative genomic analyses of *Planococcus alpinumensis* MSAK28401T, a new species isolated from Antarctic krill. *BMC. Microbiol.* 21, 288 (2021).

18. Chen, K., Wang, L., Chen, H., Zhang, C., Wang, S., Chu, P., Li, S., Fu, H., Sun, T., Liu, M., Yang, Q., & Zou, H. Complete genome sequence analysis of the peanut pathogen *Ralstonia solanacearum* strain Rs-P.362200. *BMC. Microbiol.* 21, 118 (2021).
19. Zhang, Z., Yu, Y., Wang, Y., Liu, X., Wang, L., Zhang, H., Liao, M., & Li, B. Complete genome analysis of a virulent *Vibrio scophthalmi* strain VSc190401 isolated from diseased marine fish half-smooth tongue sole, *Cynoglossus semilaevis*. *BMC. Microbiol.* 20, 341 (2020).
20. Edward, L., Ignatius, B., Joseph, M., Rees, J., Muchadeyi, F., Madoroba, E., & Heerden, H. Whole genome sequencing and identification of *Bacillus endophyticus* and *B. anthracis* isolated from anthrax outbreaks in South Africa. *BMC. Microbiol.* 18, 67 (2018).
21. Kang, Z., Zhang, J., Jin, P., & Yang, S. Directed evolution combined with synthetic biology strategies expedite semi-rational engineering of genes and genomes. *Bioengineered.* 6, 136-140 (2015).
22. Williamson, A., De Santi, C., Altermark, B., Karlsen, C., & Hjerde, E. Complete genome sequence of *Halomonas* sp. R5-57. *Stand. Genomic. Sci.* 11, 62 (2016).
23. Buermans, H., & den Dunnen, J. Next generation sequencing technology: advances and applications. *Biochim. Biophys. Acta. Mol. basis Dis.* 1842, 1932-1941 (2014).
24. Gupta, S., Sharma, P., Dev, K., Baumler, D., & Sourirajan, A. Draft Genome Sequence of *Halobacillus trueperi* SS1, Isolated from Lunsu, a Saltwater Body in the Northwest Himalayas. *Microbiol. Resour. Ann.* 8, e01710-18 (2019).
25. Chen, K., Wang, L., Chen, H., Zhang, C., Wang, S., Chu, P., Li, S., Fu, H., Sun, T., Liu, M., Yang, Q., Zou, H., & Zhuang, W. Complete genome sequence analysis of the peanut pathogen *Ralstonia solanacearum* strain Rs-P.362200. *BMC. Microbiol.* 21,118 (2021).
26. Li, X., Xu, S., Zhang, J., & Li, M. Assembly and annotation of whole-genome sequence of *Fusarium equiseti*. *Genomics.* 113, 2870-2876 (2021).
27. Zhang, X., Ruan, Y., Liu, W., Chen, Q., Gu, L., & Guo, A. Whole genome sequencing and genome annotation of *Dermaococcus abyssi* strain HZAU 226 isolated from spoiled eggs. *Genomics.* 113, 1199-1206 (2021).
28. Lekota, K., Bezuidt, O., Mafofo, J., Rees, J., Muchadeyi, F., Madoroba, E., & Heerden, H. Whole genome sequencing and identification of *Bacillus endophyticus* and *B. anthracis* isolated from anthrax outbreaks in South Africa. *BMC. Microbiol.* 18, 67 (2018).
29. Zhang, Z., Yu, Y., Wang, Y., Liu, X., Wang, L., Zhang, H., Liao, M., & Li, B. Complete genome analysis of a virulent *Vibrio scophthalmi* strain VSc190401 isolated from diseased marine fish halfsmooth tongue sole, *Cynoglossus semilaevis*. *BMC. Microbiol.* 20, 341 (2020).
30. Yu, F., Song, J., Liang, J., Wang, S., & Lu, J. Whole genome sequencing and genome annotation of the wild edible mushroom, *Russula griseocarnosa*. *Genomics.* 112, 603-614 (2020).
31. Wang, Y., Ma, L., He, J., Liu, Z., Weng, S., Wang, L., He, J., & Guo, C. Whole genome sequencing and comparative genomic analyses of *Planococcus alpinumensis* MSAK28401T, a new species isolated from Antarctic krill. *BMC. Microbiol.* 21, 288 (2021).

32. Li, T., Zhang, X., Guo, L., Qi, T., Tang, H., Wang, H., Qiao, X., Zhang, M., Zhang, B., Feng, J., Zuo, Z., Zhang, Y., Xing, C., & Wu, J. Single-molecule real-time transcript sequencing of developing cotton anthers facilitates genome annotation and fertility restoration candidate gene discovery. *Genomics*. 113, 4245-4253 (2021).
33. Woo, M., Park, S., Park, K., Park, M., Kim, J., Lee, H., Sohn, J., Lee, D., Nam, G., Shin, K., & Lee, S. Draft genome sequence of the halophilic *Halobacillus mangrovi* KTB 131 isolated from Topan salt of the Jeon-nam in Korea. *Genomics. Data*. 14, 18-20 (2017).
34. Treves, D., Francis, J., III, & Kirchner, G. Draft genome sequence of the moderately halophilic bacterium *Halobacillus* sp. BBL2006. *Data. Brief*. 21, 2410-2413 (2018).

Tables

Table 1. The statistics of non-coding RNA prediction of *Halobacillus trueperi* S61.

Type	Number	Average length(bp)	Total length	In genome (%)
tRNA	86	77	6648	0.16
16S_rRNA	10	1538	15380	0.38
5S_rRNA	10	115	1150	0.03
23S_rRNA	10	2926	29260	0.72
sRNA	23	123	2847	0.07

Figures

Figure 1

The Genome *Halobacillus trueperi* S61 circle diagram.

Figure 2

The *Halobacillus trueperi* S61 subreads length distribution (a) and assembly result of GC-Depth distribution (b).

Figure 3

The function classification of *Halobacillus trueperi* S61 according to cluster of orthologous groups of proteins (COG) database.

Figure 4

The function classification of *Halobacillus trueperi* S61 according to gene ontology (GO) database.

Figure 5

Comparative genomic analysis of genome *Bacillus velezensis* (Hal61) and *Rosellomorea vietnamensis* (Rvi): Two-dimensional collinearity map (b) and Gene family (c).

Figure 6

The advanced function classification of *Halobacillus trueperi* S61 according to carbohydrate-active enzymes (CAZy) database.