Draft genome sequence of *Paenibacillus algorigonticola* sp. nov., an antimicrobial-producing strain

Liying Zhu\(^a\), Qian Wu\(^b\), Qiu Liu\(^b\), Xian Yu\(^b\), Ling Jiang\(^c\), He Huang\(^b\)

\(^a\) College of Sciences, Nanjing Tech University, Nanjing 210009, PR China
\(^b\) College of Biotechnology and Pharmaceutical Engineering, Nanjing Tech University, Nanjing 210009, PR China
\(^c\) College of Food Science and Light Industry, Nanjing Tech University, Nanjing 210009, PR China

**Abstract**

*Paenibacillus algorigonticola* sp. nov. is isolated from a cold spring sample from Xinjiang Uygur Autonomous Region (China), a novel strain that can produce antimicrobial substance against human pathogenic bacteria and fungi, including *Staphylococcus aureus* and *Candida albicans*. Here we report a 7.60-Mb assembly of its genome sequence and other useful information, including the coding sequences (CDSs) responsible for the biosynthesis of antibacterial factors, anaerobic respiration and several immune-associated reactions. Also, prospective studies on *P. algorigonticola* sp. nov. in the cold spring might offer a potential source for the discovery of bioactive compounds with medical value. The data repository is deposited on the website [http://www.ncbi.nlm.nih.gov/nuccore/LAQO00000000](http://www.ncbi.nlm.nih.gov/nuccore/LAQO00000000) and the accession number is LAQO00000000.

**Keywords:**
- *Paenibacillus algorigonticola*
- Cold spring
- Antimicrobial factors
- Genome analysis

**Specifications**

| Organism/cell/tissue type | *Paenibacillus algorigonticola* sp. nov. XJ259\(^T\) |
|---|---|
| Strain | *Paenibacillus algorigonticola* sp. nov. XJ259\(^T\) |
| Sequencer or array type | Illumina HiSeq 2000 |
| Data format | Raw and processed |
| Experimental factors | DNA extracted from a wild-type strain, no treatment |
| Experimental features | Draft genome sequencing of *Paenibacillus algorigonticola*, assembly and annotation |
| Consent | N/A |
| Sample source location | a cold spring sample from Xinjiang Uygur Autonomous Region (China) |

**1. Direct link to deposited data**

[http://www.ncbi.nlm.nih.gov/nuccore/LAQO00000000](http://www.ncbi.nlm.nih.gov/nuccore/LAQO00000000)

A large-scale metagenomic-based study revealed that the microbes with antibiotic resistance genes are abundant and diverse in nature [1]. However, the widespread misuse or abuse of antibiotics in clinical practice has led to a large number of antibiotic-resistant pathogenic bacteria such as *Staphylococcus aureus* and *Candida albicans*. Recently, our group has isolated from a cold spring sample from Xinjiang Uygur Autonomous Region (China) with a novel species of *Paenibacillus algorigonticola* sp. nov., named XJ259\(^T\) (=CGMCC 1.10223\(^T\) = JCM16598\(^T\)), which can produce water-soluble constituents with the significant inhibitory activity against both *S. aureus* and *C. albicans* [2]. Comparisons with 16S rRNA gene sequences as shown in Fig. 1 revealed that the novel strain (1519 nt) had the highest similarity to *Paenibacillus xinjiangensis* B538\(^T\) (96.6%) [3]. However, the phylogenetic distances from recognized species (Fig. 2) indicated that *P. algorigonticola* is not affiliated to any of these recognized species and the proportion of saturated straight-chain fatty acid C\(_{16:0}\) was relatively high [2]. We can therefore conclude that this strain represents a novel species of the genus *Paenibacillus*. In consequence, investigation of the genetic information and characteristics of *P. algorigonticola* is desired to further investigate its mechanism of metabolic regulation. Knowledge of the genome sequence and bioinformatics will be of great help in this regard.

Here we present the draft genome sequence of strain *P. algorigonticola* XJ259\(^T\) obtained using Illumina HiSeq 2000 system. The reads were assembled with SOAPdenovo [4,5], the version is 2.04, and the sequence was annotated using the RAST annotation server [Fig. 3] [6], and the KEGG metabolic pathway was also constructed [Fig. 4]. A library containing 500-bp inserts was constructed. Sequencing was performed based on the paired-end strategy of 478-bp reads to produce 1161 Mb of filtered sequences, representing a 132-fold coverage of the genome. The sequence
of *P. algorifonticola* is 7,495,641 bases with a G + C content of 48.39%, which was assembled into 115 contigs and 101 scaffolds. It contains 7048 open reading frames (ORFs), 70 tRNA genes, and 4 rRNA genes (Table 1) identified by Glimmer 3.02 [7], Genemark [8], tRNAscan-SE [9], and RNAmmer [10].

According to the genomic analysis of the strain *P. algorifonticola*, we identified 6 neutrophil extracellular traps (NETs)-associated proteins, including the calcium ion binding protein and cytoplasmic calprotectin protein complex (e.g., Mrp8/14-complex or S100A8/A9) with potent antimicrobial properties [11]. We also identified 6 ORFs relate to the immune-associated reactions, including the nucleotide-binding oligomerization domain (NOD)-like receptor signaling pathway, antigen processing and immune-responsive domain IPR011614. The strain *P. algorifonticola* can also utilize other complex immune responses that lead to production of soluble effectors, including 14 antimicrobial peptides (AMPs), or to

---

**Fig. 1.** Multiple sequences alignment of *Paenibacillus* (P: *Paenibacillus telluris*; 2P: *Paenibacillus* sp. LNUB461; 3P: *Paenibacillus telluris* strain JS01-08; 4P: *Paenibacillus xinjiangensis* strain B538; 5P: *Paenibacillus castaneae* strain Ch-32; 6P: *Paenibacillus algorifonticola*).
activation of complement to inhibit activity of key metabolic enzymes (e.g., malate dehydrogenase and succinate dehydrogenase), which can directly damage the pathogen [12,13]. Additionally, *P. algorifonticola* have the ability to keep alive in the anaerobic condition since there were 8 ORFs related to the anaerobic dehydrogenases and anaerobic regulatory protein. Except for the biosynthesis of antimicrobial substances, the trehalose, vitamin B12 and B2 biosynthesis were also annotated in *P. algorifonticola*, further studies will be performed to confirm their functions, and a complete genome sequence will be included in the future to reveal the unique molecular characteristics of strain *P. algorifonticola*.

### 2. Nucleotide sequence accession numbers

This whole genome shotgun project has been deposited at DDBJ/EMBL/GenBank under accession number LAQ00000000. The version described in this paper is the first version, with accession number LAQ001000000.

#### Conflict of interest

The authors declare that there is no conflict of interest on any work published in this paper.
Acknowledgments

This work was supported by the National Science Foundation for Young Scholars of China (21225628, 21406111), the National High Technology Research and Development Program of China (2012AA022101, 2012AA021700), the National Natural Science Foundation of China for Young Scholars (21406111), the Natural Science Foundation of Jiangsu Province (BK20131406, BK20130917), and the Natural Science Foundation for Colleges and Universities in Jiangsu Province (14KJB530003).

References

[1] N. Joseph, C. Se’bastien, S. Pascal, et al., Large-scale metagenomic-based study of antibiotic resistance in the environment. Curr. Biol. 24 (2014) 1096–1100.
[2] Q.Y. Tang, N. Yang, J. Wang, et al., Paenibacillus algorifonticola sp. nov., isolated from a cold spring. Int. J. Syst. Evol. Microbiol. 61 (2011) 2167–2172.
[3] J.M. Lim, C.O. Jeon, D.J. Park, et al., Paenibacillus xinjiangensis sp. nov., isolated from Xinjiang province in China. Int. J. Syst. Evol. Microbiol. 56 (2006) 2579–2582.
[4] Li, et al., De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 20 (2010) 265–272.
[5] D.R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18 (2008) 821–829.

[6] R.K. Aziz, D. Bartels, A.A. Best, et al., The RAST server: rapid annotations using subsystems technology. BMC Genomics 9 (75) (2008), http://dx.doi.org/10.1186/1471-2164-9-75.

[7] A.L. Delcher, K.A. Bratke, E.C. Powers, et al., Identifying bacterial genes and endosymbiont DNA with glimmer. Bioinformatics 23 (2007) 673–679.

[8] J. Besemer, A. Lomsadze, M. Borodovsky, GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29 (2001) 2607–2618.

[9] K. Lagesen, P. Hallin, E.A. Rødland, et al., Rfam: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35 (2007) 3100–3108.

[10] T.M. Lowe, S.R. Eddy, TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25 (1997) 955–964.

[11] C.F. Urban, D. Erment, M. Schmid, et al., Neutrophil extracellular traps contain calprotectin, a cytosolic protein complex involved in host defense against Candida albicans. PLoS Pathog. 5 (10) (2009) e1000639.

[12] S. Marc, F.E. Joachim, Interplay between Candida albicans and the antimicrobial peptide armory. Eukaryot. Cell 13 (8) (2014) 950–957.

[13] G. Mariusz, N. Joanna, F. Grzegorz, et al., Antimicrobial photodynamic therapy with fulleropyrrolidine: photoinactivation mechanism of Staphylococcus aureus, in vitro and in vivo studies. Appl. Microbiol. Biotechnol. 99 (9) (2015) 4031–4043.

Fig. 1 (continued).
Fig. 1 (continued).

Paenibacillus sp. XWS-29 (JQ617905.1)
Paenibacillus sp. SaMR6 (JQ806430.1)
Paenibacillus sp. SQ2-36-3 (KM252921.1)
Paenibacillus sp. A12 (KF479531.1)
Paenibacillus sp. rif200865 (FJ527667.1)
Paenibacillus xinjiangensis strain B538 (NR_043221.1)
Paenibacillus castaneae strain Ch-32 (NR_044403.1)
Paenibacillus algorifonticola (NR_108602.1)
Paenibacillus telluris (AM745265.1)
Paenibacillus sp. JS01-08 (AM162314.1)
Paenibacillus sp. LNUB461 (KF997862.1)
Paenibacillus sp. IHB B4030 (HM233996.1)
Paenibacillus agaridevorans strain DSM 1355 (HQ290361.1)

Fig. 2. Phylogenetic tree analysis of Paenibacillus algorifonticola in this study and other strains belonging to the genus Paenibacillus (MEGA 3.1).
**Table 1**

General features of *Paenibacillus algorifonticola* sp. nov. draft genome.

| Attributes                  | Value  |
|-----------------------------|--------|
| Total sequence length (bp)  | 7,495,641 |
| Gene number                 | 7048   |
| Gene length/Genome (%)      | 85.86  |
| Gene average length (bp)    | 913    |
| GC content (%)              | 48.39  |
| N50 contig length (bp)      | 200,917|
| N90 contig length (bp)      | 57,078 |
| Number of contig            | 115    |
| Number of scaffolds          | 101    |
| tRNAs                       | 70     |
| rRNAs (5S, 16S, 23S)        | 4      |
| sRNA                        | 5      |
| Protein-coding genes        | 5364   |
| Genes assigned to COGs      | 3376   |
| Number of TFR               | 143    |