Data Article

# Draft genome sequence data of a chromium reducing bacterium, *Bacillus licheniformis* strain KNP

Pankaj Kumar Arora [a,*], Rupali Mishra [a], Rishabh Anand Omar [a], Raj Shekhar Saroj [a], Alok Srivastava [b], Sanjay Kumar Garg [b], Vijay Pal Singh [b]

[a] *Department of Microbiology, Babasaheb Bhimrao Ambedkar University, Lucknow 226025, India*
[b] *Department of Plant Science, Faculty of Applied Sciences, MJP Rohilkhand University, Bareilly, India*

## A B S T R A C T

A chromium-reducing bacterium designated as strain KNP was isolated from a sample collected from a tannery effluent of Kanpur, India. Phylogenetic analysis based on the 16S rRNA gene sequences revealed that strain KNP belonged to the *Bacillus* genus and showed 100% similarity with *Bacillus licheniformis*. Furthermore, average nucleotide identity and digital DNA-DNA hybridization between strain KNP and its closely related strains confirmed its affiliation with *Bacillus licheniformis* species. Whole-genome sequencing of *Bacillus licheniformis* KNP was performed using the Illumina Hiseq platform. Here, we present the draft genome sequence of *Bacillus licheniformis* KNP. The total size of the draft assembly was 4,280,093 bp, distributed into 21 contigs with an N50 value of 4,186,229. The genome has 45.9% G + C content, 4255 coding sequences and 86 putative RNA genes. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession

---

* Corresponding author.
  *E-mail address:* arora484@gmail.com (P.K. Arora).
  *Social media:* 🐦 (P.K. Arora)

JACDXS000000000. The version described in this paper is version JACDXS010000000.

## Specifications Table

| | |
|---|---|
| Subject | Biochemistry, Genetics and Molecular Biology (General) |
| Specific subject area | Genomics and Microbiology |
| Type of data | Draft genome sequence data in FASTA format, table and figures |
| How data were acquired | Whole genome sequence of *Bacillus licheniformis* KNP was sequenced with Illumina HiSeq system. |
| Data format | Raw, analyzed and assembled genome sequences |
| Parameters for data collection | Strain KNP reduced hexavalent chromium up to concentrations of 1000 ppm. Genomic DNA was isolated from a pure culture of *Bacillus licheniformis* KNP. |
| Description of data collection | Whole-genome sequencing, assembly, and annotation |
| Data source location | *Bacillus licheniformis* KNP was isolated from a tannery effluent sample collected from Kanpur (26.4670°N 80.3500°E), India. |
| Data accessibility | Data is publicly available at NCBI Genbank from the following links: https://www.ncbi.nlm.nih.gov/nuccore/JACDXS000000000 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA645298 https://www.ncbi.nlm.nih.gov/biosample/SAMN15501749 |

## Value of the Data

- Whole Genome sequence of *Bacillus licheniformis* strain KNP could provide valuable information about chromium resistance and its transformation.
- The data of this article could be useful for scientists and students working in the field of environmental microbiology, environmental biotechnology, genomics and genetic engineering.
- This genome data could be valuable resource for comparative genomic analysis among *Bacillus licheniformis* strains.
- Based on genome data, *Bacillus licheniformis* strain KPN could be a potential strain for study of heavy metal stress, bacterial chemotaxis, and various enzymes production.

## 1. Data Description

*Bacillus licheniformis* KNP was isolated from a tannery effluent sample collected from Kanpur, India. Strain KNP reduced hexavalent chromium completely within 48 h when it was grown in nutrient media containing 1000 mg/L potassium dichromate under shaking conditions (200 rpm). The genome features of strain KPN were summarized in Table 1. The assembled genome of *Bacillus licheniformis* KNP comprised 21 contigs with a total size of 4,280,093 bp and N50 value of 4,186,229. The genome G + C content was 45.9%. Based on the genome annotation, a total of 4434 genes were predicted in which, 4255 of them were responsible for coding specific proteins while 86 and 93 of them were coded for RNA genes (77 tRNAs, 5 ncRNAs, 4 16S-23S-5S rRNAs) and pseudogenes, respectively. A circular map of genome of *Bacillus licheniformis* KNP was represented in Fig. 1.

The 16S rRNA gene sequence of strain KNP was deposited to NCBI Genbank under accession number MW265434. Phylogenetic analysis based on the 16S rRNA gene sequences revealed that strain KNP was affiliated to *Bacillus* (Fig. 2) and exhibited 100% 16S rRNA gene similarity with *Bacillus licheniformis* ATCC 14580. Furthermore, genome-based taxonomic analysis showed that strain KNP exhibited high average nucleotide identity (ANI) value (99.54%) as well as high digital DNA-DNA hybridization (dDDH) value (96.50%) with *Bacillus licheniformis* (Table 2). Based on the

**Table 1**
Genome features of *Bacillus licheniformis* strain KNP.

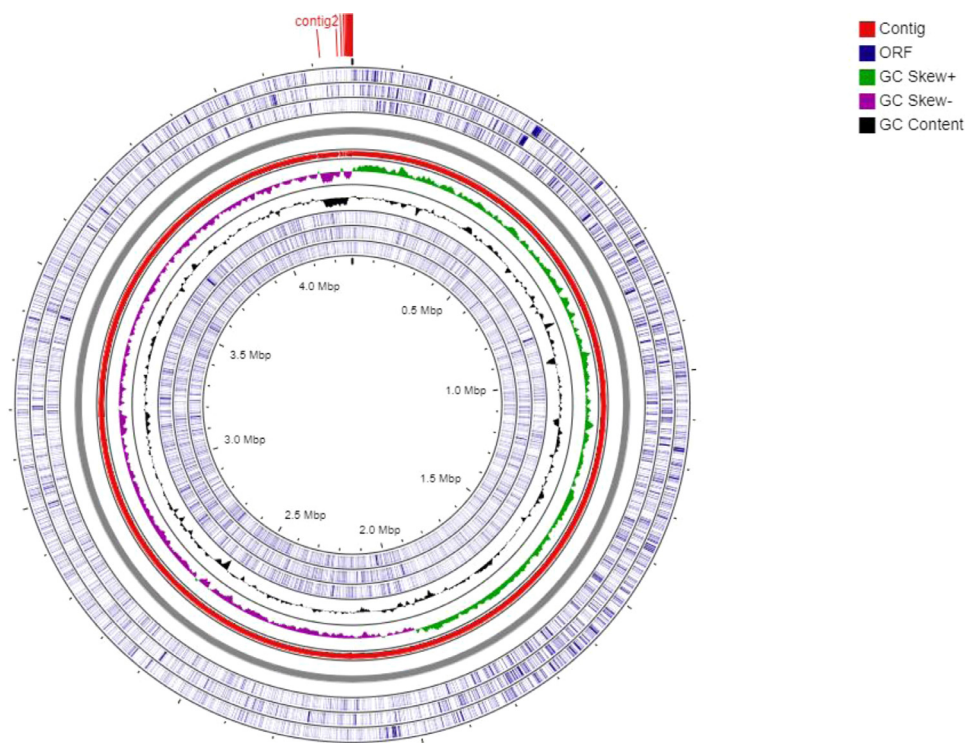| Features | Value | Percentage |
|---|---|---|
| Number of contigs | 21 | 100 |
| Genome size (bp) | 4,280,093 bp | 100 |
| G + C | 1,964,563 bp | 45.9 |
| Genes (total) | 4434 | 100 |
| Protein coding genes | 4255 | 95.96 |
| RNA genes | 86 | 1.94 |
| 5S rRNA gene | 2 | 0.05 |
| 16S rRNA gene | 1 | 0.02 |
| 23S rRNA gene | 1 | 0.02 |
| tRNAs | 77 | 1.74 |
| ncRNAs | 5 | 0.11 |
| Pseudo Genes (total) | 93 | 2.10 |



**Fig. 1.** Circular map of *Bacillus licheniformis* strain KNP. The following rings were included: i) ORFs (blue color); (ii) Contigs (red color); iii) Positive GC Skew (green color); (iv) Negative GC Skew (violet color); (v) G + C content (black color). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cutoff values on species delimitation established for ANI (> 95–96%) [1] and dDDH value (> 70%) [2], strain KNP was strongly affiliated to *Bacillus licheniformis.*

Functional annotation of genome of strain KPN revealed presence of three chromate transporters that involve in chromium resistance through chromate efflux mechanism [3]. Furthermore, putative genes related to the chromium reduction such as nitroreductase [4], quinone reductase [5], and azoreducatase [6] were also identified in the genome of strain KNP. In addition, genes coding the proteins involved in arsenic resistance and reduction [7] such as arsenate
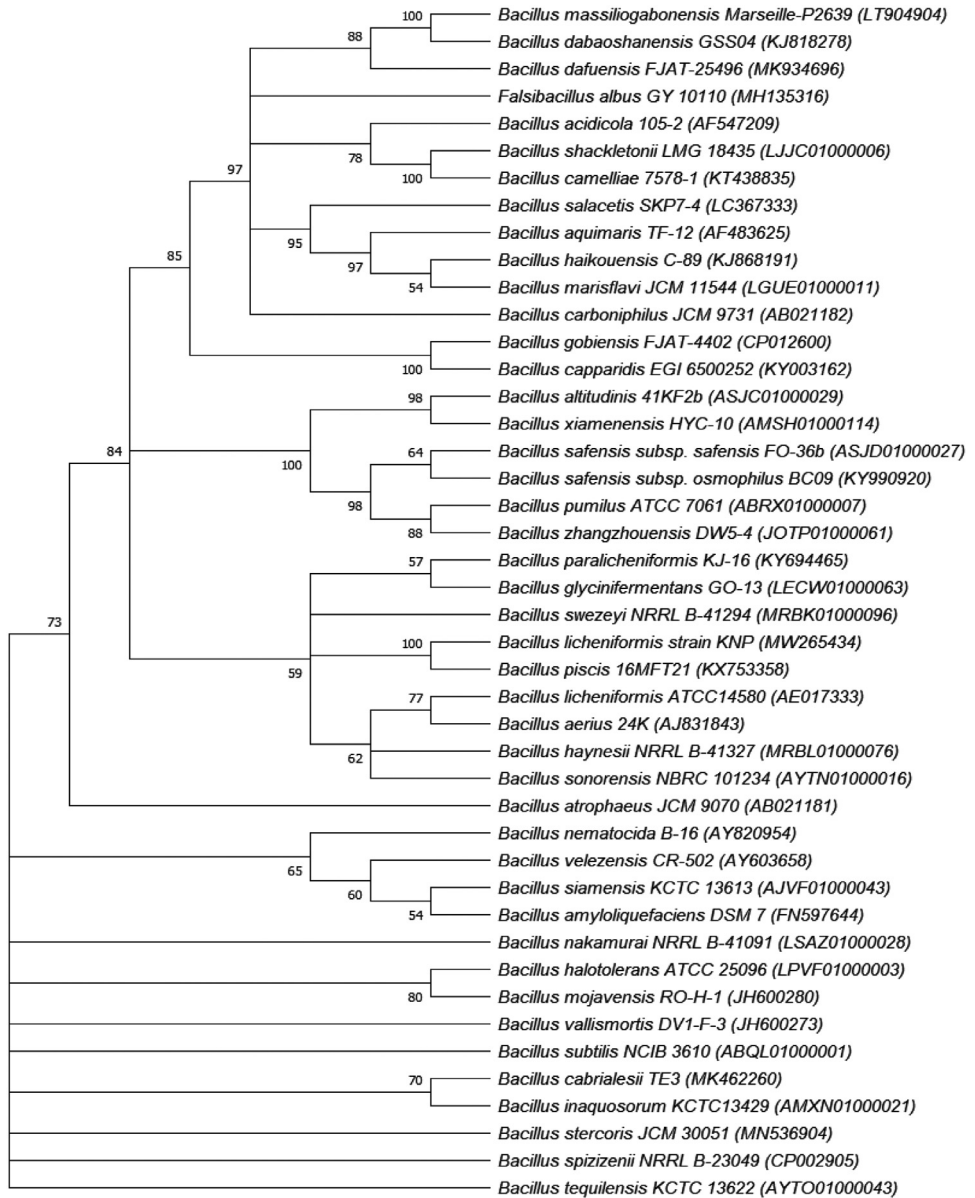
**Fig. 2.** Neighbor-joining tree based on the16S rRNA gene sequence (1325 nt) showing the relationships among strain KNP and its related type strains The numbers at branch nodes indicate bootstrap percentages derived from 1000 replications; only values >50 % are shown.

reductase (thioredoxin), arsenical efflux pump membrane protein ArsB, arsenical pump-driving ATPase, arsenate reductase family protein, arsenite efflux transporter metallochaperone ArsD, arsenic transporter were also identified in the KNP genome. Several other genes involved in heavy metal resistance including heavy metal translocating P-type ATPase, divalent metal cation transporter, Nramp family divalent metal transporter, nickel ABC transporter, nickel/metallophore periplasmic binding protein, metal ABC transporter ATP-binding protein, metal ABC transporter

**Table 2**

16S rRNA gene similarity (> 98.7%), ANI and dDDH values of strain KNP with its closely related species.

| Closely related Species | 16S rRNA sequence similarity | OrthoANI value (%) | dDDH value (%) |
|---|---|---|---|
| *Bacillus licheniformis* ATCC 14,580 | 100.00% | 99.54% | 96.50% |
| *Bacillus paralicheniformis* KJ-16 | 99.92% | 94.65% | 57.70% |
| *Bacillus haynesii* NRRL B-41,327 | 99.92% | 95.65% | 64.10% |
| *Bacillus sonorensis* NBRC 101,234 | 99.84% | 81.82% | 24.40% |
| *Bacillus glycinifermentans* GO-13 | 99.84% | 80.93% | 23.60% |
| *Bacillus swezeyi* NRRL B-41,294 | 99.45% | 83.38% | 26. 20% |
| *Bacillus subtilis* NCIB 3610 | 98.90% | 72.53% | 18.50% |
| *Bacillus cabrialesii* TE3 | 98.90% | 72.73% | 18.30% |
| *Bacillus inaquosorum* KCTC 13,429 | 98.90% | 72.90% | 18.60% |
| *Bacillus stercoris* D7XPN1 | 98.90% | 72.60% | 18.50% |
| *Bacillus spizizenii* TU-B-10 | 98.90% | 72.80% | 19.10% |
| *Bacillus tequilensis* KCTC 13,622 | 98.82% | 72.83% | 18.10% |
| *Bacillus siamensis* KCTC 13,613 | 98.82% | 72.69% | 18.40% |
| *Bacillus amyloliquefaciens* DSM 7 | 98.74% | 72.49% | 19.20% |
| *Bacillus atrophaeus* NRRL NRS 213 | 98.74% | 72.68% | 18.10% |
| *Bacillus nakamurai* NRRL B-41,091 | 98.74% | 72.59% | 18.40% |
| *Bacillus halotolerans* FJAT-2398 | 98.74% | 72.61% | 18.40% |
| *Bacillus velezensis* NRRL B-41,580 | 98.68% | 72.67% | 18.40% |
| *Bacillus mojavensis* KCTC 3706 | 98.67% | 72.91% | 18.30% |

permease, metal ABC transporter substrate-binding protein, MerR family transcriptional regulator, copper-sensing transcriptional repressor CsoR, BlaI/MecI/CopY family transcriptional regulator, TetR/AcrR family transcriptional regulator, MarR family transcriptional regulator, GbsR/MarR family transcriptional regulator, LysR family transcriptional regulator, Transcriptional regulator MntR, Zn(II)-responsive metalloregulatory transcriptional repressor CzrA, MgtC/SapB family protein, Spx/MgsR family RNA polymerase-binding regulatory protein were also detected in the KNP genome. Moreover, genes related to bacterial chemotaxis were also identified in the KNP genome.

PHAST analysis [8] was performed to identify prophages in the genome of strain KNP. Total 5 prophage regions were identified, of which 3 regions were intact, 2 regions were incomplete. Intact regions of prophages were located between positions 124,430–169,526, 1,312,385–1,348,631 and 3,119,232–3,183,647 bp, respectively.

To predict the potential of *Bacillus licheniformis* KNP to produce secondary metabolites, the genome of this strain was analyzed with antiSMASH server v.5.0 [9]. The results showed that the genome contained eleven gene clusters coding for enzymes involved in the biosynthesis of bacteriocins, nonribosomal peptides, thiopeptide, siderophores, betalactone, terpenes, lanthipeptides, type III polyketides, lassopeptides, and tRNA-dependent cyclodipeptide. Two of the non-ribosomal peptide synthetase (NRPS) gene clusters exhibited 100% and 53% similarity with the lichenysin and bacillibactin gene clusters, respectively. Moreover, lanthipeptide and betalactone gene clusters showed 100% and 53% similarity with lichenicidin VK21A1/VK21A2 and fengycin gene clusters, respectively. Based on genome analysis, strain KPN has a potential strain for study of heavy metal stress, bacterial chemotaxis, and various enzymes production.

## 2. Experimental Design, Materials and Methods

### 2.1. Collection of sample and isolation of chromium resistance bacteria

*Bacillus licheniformis* KNP was isolated from a tannery effluent sample collected from Kanpur (26.4670°N 80.3500°E), India. Briefly, a tannery effluent sample was collected in screw capped sterilized bottle from Kanpur, India. For isolation of hexavalent chromium resistance bacteria, 1 ml of water sample was inoculated in 500 ml Nutrient broth and 500 ppm potassium

dichromate for 72 h. The sample was serially diluted and poured into nutrient agar plate containing 500 ppm potassium dichromate and plates were inoculated at 30 °C for 48 h. A total eighteen morphotypes were selected and purified and preserved at −10 glycerol vials.

### 2.2. Hexavalent chromium transformation assay

All eighteen bacteria were screened for the reduction of hexavalent chromium at various concentrations (200–1200 ppm) of potassium dichromate by the diphenylcarbazide colorimetric method [10]. Out of eighteen bacteria, only one bacterium designated strain KNP showed the hexavalent chromium reduction up to concentrations of 1000 ppm.

### 2.3. Bacterial cultivation and DNA isolation

*Bacillus licheniformis* KNP was cultivated on nutrient agar plate at 37 °C. A single colony of strain KNP was grown overnight in Luria-Bertani media under shaking conditions. The culture was centrifuged and pellet was used for DNA extraction. Genomic DNA was extracted using the DNAminikit (Qiagen, Germantown, MD, USA), according to the manufacturer's instruction.

### 2.4. Genome sequencing, assembly, and annotation

A whole-genome sequencing library was constructed using the Nextera XT DNA library preparation kit, according to the manufacturer's instruction. The libraries were sequenced using the Hiseq platform (Illumina, San Diego, CA, USA), with 150-bp paired-end reads and 1900-fold genome coverage. FastQC was used to monitor the initial quality of the raw sequencing data [11]. The raw reads and adapter contam-inations were trimmed with Trim galore 0.6.5 [12] and primary assembly was performed using Unicycler version v0.4.8 [13]. Default parameters were used for all software unless otherwise specified. Annotation was performed using NCBI Prokaryotic Genome Automatic Annotation pipeline (PGAAP) [14]. The genome was analyzed with the PHAge Search Tool (PHAST) to identify prophages [8]. The potential secondary metabolite biosynthetic gene clusters (BGCs) were identified in the genome using antiSMASH v5.0 [9]. The graphical circular map of the complete genome was constructed and visualized using CGView Server [15].

### 2.5. 16S rRNA gene sequence and phylogenetic analysis

The 16S rRNA gene sequence of strain KNP (1325 nt) was retrieved from the draft genome sequence of strain KNP with RNAmmer [16]. The 16S rRNA gene sequence of strain KNP was analyzed by EzBiocloud to determinate its more closely related strains [17]. The 16S rRNA gene sequences of all closely related species were retrieved from EzBiocloud database. All sequences were aligned with ClustalW [18] and phylogenetic tree was constructed by the neighbour joining method with MEGA X software package [19].

### 2.6. Average nucleotide identity and digital DNA-DNA hybridization

Average nucleotide identity (ANI) between genomes of strain KNP and closely related species were determined by the OrthoANI algorithm [20] and digital DNA-DNA hybridization (dDDH) values were determined using genome-to-genome distance calculator (GGDC) 2.1 by BLAST [21].

## Data Availability

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JACDXS000000000. The version described in this paper is version JACDXS010000000.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

## References

[1] N.J. Varghese, S. Mukherjee, N. Ivanova, K.T. Konstantinidis, K. Mavrommatis, N.C. Kyrpides, A. Pati, Microbial species delineation using whole genome sequences, Nucl. Acids Res. 43 (2015) 6761–6771.

[2] J.P. Meier-Kolthoff, H.P. Klenk, M. Göker, Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age, Int. J. Syst. Evol. Microbiol. 64 (2014) 352–356.

[3] H. Thatoi, S. Das, J. Mishra, B.P. Rath, N. Das, Bacterial chromate reductase, a potential enzyme for bioremediation of hexavalent chromium: a review, J. Environ. Manag. 146 (2014) 383–399.

[4] Y.H. Kwak, D.S. Lee, H.B. Kim, Vibrio Harvey initroreductase is also a chromate reductase, Appl. Environ. Microbiol. 69 (2003) 4390–4395.

[5] S. Eswaramoorthy, S. Poulain, R. Hienerwadel, N. Bremond, M. D.Sylvester, Y.B. Zhang, C. Berthomiue, D.V.D. Lelie, A. Matin, Crystal structure of ChrR–a quinone reductase with the capacity to reduce chromate, PLoS ONE 7 (2012) e36017.

[6] G. Liu, J. Zhou, L.V. Hong, X. Xiang, J. Wang, M. Zhou, Q.V. Yuanyuan, Azoreductase from Rhodobactersphaeroides AS1.1737 is a flavodoxin that also functions as nitroreductase and flavin mononucleotide reductase, Appl. Microbiol. Biotechnol. 76 (2007) 1271–1279.

[7] S. Srivastava, K. Shukla K, Microbes are essential components of arsenic cycling in the environment: implications for the use of microbes in arsenic remediation, in: P. Arora (Ed.), Microbial Metabolism of Xenobiotic Compounds. Microorganisms for Sustainability, 10, Springer, Singapore, 2019, pp. 217–227, doi:10.1007/978-981-13-7462-3_10.

[8] Y. Zhou, Y. Liang, K.H. Lynch, J.J. Dennis, D.S. Wishart, PHAST: a fast phage search tool, Nucl. Acids Res. 39 (2011) W347–WW35.

[9] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S.Y. Lee, M.H. Medema, T. Weber, AntiSMASH 5.0: updates to the secondary metabolite genome mining pipeline, Nucl. Acids Res. 47 (2019) W81–W87.

[10] A. Lace, D. Ryan, M. Bowkett, J. Cleary, Chromium monitoring in water by colorimetry using optimised 1, 5-diphenylcarbazide method, Int. J. Environ. Res. Public Health 16 (2019) 1803.

[11] S. Andrews, FastQC: a quality control tool for high throughput sequence data (2010). https://www.bioinformatics.babraham.ac.uk/projects/fastqc.

[12] F. Krueger, Trim Galore: a wrapper tool around cutadapt and FastQC to consistently apply quality and adapter trimming to FastQfiles (2015). https://www.bioinformatics.babraham.ac.uk/projects/trim_galore.

[13] R.R. Wick, L.M. Judd, C.L. Gorrie, K.E. Holt, Unicycler: resolving bacterial genome assemblies from short and long sequencing reads, PLoS Comput. Biol. 13 (2017) e1005595.

[14] T. Tatusova, M. DiCuccio, A. Badretdin, V. Chetvernin, E.P. Nawrocki, L. Zaslavsky, A. Lomsadze, K.D. Pruitt, M. Borodovsky, J. Ostell, NCBI prokaryotic genome annotation pipeline, Nucl. Acids Res. 44 (2016) 6614–6624.

[15] J.R. Grant, P. Stothard, The CGView server: a comparative genomics tool for circular genomes, Nucl. Acids Res. 36 (2008) W181–W184.

[16] K. Lagesen, P. Hallin, E.A. Rødland, H.H. Staerfeldt, T. Rognes, D.W. Ussery, RNAmmer: consistent and rapid annotation of ribosomal RNA genes, Nucl. Acids Res. 35 (2007) 3100–3108.

[17] S.H. Yoon, S.M. Ha, S. Kwon, J. Lim, Y. Kim, H. Seo, J. Chun, Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies, Int J. Syst. Evol. Microbiol. 67 (2017) 1613–1617.

[18] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucl. Acids Res. 22 (1994) 4673–4680.

[19] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549.
[20] I. Lee, Y. Ouk Kim, S.C. Park, J. Chun, OrthoANI: an improved algorithm and software for calculating average nucleotide identity, Int. J. Syst. Evol. Microbiol. 66 (2016) 1100–1103.
[21] J.P. Meier-Kolthoff, A.F. Auch, H.-.P. Klenk, M. Göker, Genome sequence-based species delimitation with confidence intervals and improved distance functions, BMC Bioinform. 14 (2013) 60.