3 Biotech (2018) 8:223
https://doi.org/10.1007/s13205-018-1242-y

GENOME REPORTS

Whole genome shotgun sequence of *Bacillus paralicheniformis* strain KMS 80, a rhizobacterial endophyte isolated from rice (*Oryza sativa* L.)

Kannepalli Annapurna[1] · Venkadasamy Govindasamy[1] · Meenakshi Sharma[1] · Arpita Ghosh[2] · Surendra K. Chikara[2]

Received: 20 January 2018 / Accepted: 3 April 2018 / Published online: 20 April 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Bacillus paralicheniformis strain KMS 80 (MTCC No. 12704) is an isolate from the root tissues of rice (*Oryza sativa* L.) that displays biological nitrogen fixation and plant growth promoting abilities. Here, we report the complete genome sequence of this strain, which contains 4,566,040 bp, 4424 protein-coding genes, 8692 promoter sequences, 67 tRNAs, 20 rRNA genes with six copies of 5S rRNAs along with a single copy of 16S–23S rRNA and genome average GC-content of 45.50%. Twenty one genes involved in nitrogen metabolism pathway and two main transcriptional factor genes, *glnR* and *tnrA* responsible for regulation of nitrogen fixation in *Bacillus* sp. were predicted from the whole genome of strain KMS 80. Analysis of the ~ 4.57 Mb genome sequence will give support to understand the genetic determinants of host range, endophytic colonization behaviour as well as to enhance endophytic nitrogen fixation and other plant beneficial role of *B. paralicheniformis* in rice.

Keywords Rice (*Oryza sativa* L.) · Endophyte · Diazotroph · *Bacillus paralicheniformis* · Illumina · NGS · Whole genome sequencing

Plants are known to harbour rich communities of commensal and mutualistic bacteria. Soil bacteria in close proximity to plant roots (termed the rhizosphere) have been shown to have a profound effect on plant nutrient acquisition, disease suppression, and growth. Plants selectively recruit and moderate their associated root microbiomes. Identifying the composition and maintenance of the plant associated microbiome is not only important to strengthen our basic understanding of plant–microbe interactions; it is of great agronomic and ecological importance as well (Smith 2015). Among the different kind of plant–microbe interactions, beneficial "endophytic interactions" occur, wherein diverse group of bacteria and fungi colonize plant inter and intracellular spaces (Reinhold-Hurek and Hurek 2011; Hirsch et al. 2012). The structural composition of endophytic bacterial communities depends on genotype of the host plant, the kind of tissue and the vegetation stage. The species composition may also be significantly influenced by plant stress and soil types (Govindasamy et al. 2017; Kumar et al. 2017). In our study, we purified 110 isolates with different morphologies from surface sterilized root–tissues of nine genotypes of rice (*Oryza sativa* L.) grown in alluvial soil of ICAR-IARI research Farm, PUSA campus New Delhi (Annapurna et al. 2017). Among them, an isolate of a Gram-positive, facultatively anaerobic, rod-shaped, motile, endospore forming bacterium was identified as a *Bacillus licheniformis* strain by 16S rRNA gene sequencing. It grew on N_2 free mineral medium, showed nitrogenase activity in acetylene reduction assay and was found to be *nif*H positive. Hence, this was selected for whole genome sequencing. *B. licheniformis* is a bacterium commonly found in the soil. Some of which have been identified previously as *Bacillus subtilis*. *B. licheniformis* has been further categorized into two distinct groups as Group 1 and Group 2 (Madslien et al. 2012; Dhakal et al. 2014). Based on the whole genome analysis, our isolate was named *Bacillus paralicheniformis* strain KMS 80 (MTCC No. 12704). *B. paralicheniformis* has been designated as a new member branched from *B. licheniformis* Group 2, which represents a novel species within the genus *Bacillus* as proposed by Dunlap et al. (2015). Genomic DNA from *B. paralicheniformis* strain KMS 80 was extracted from exponential growth cultures (1 mL, $A_{600} = 0.5$) using the NucleoSpin® DNA extraction Kit for DNA (NucleoSpin®, Germany).

✉ Kannepalli Annapurna
annapurna96@gmail.com

[1] Division of Microbiology, ICAR-Indian Agricultural Research Institute, New Delhi 110012, India

[2] Eurofins Genomics India Private Limited, Bangalore 560048, Karnataka, India

Quality and quantity of genomic DNA was checked using 0.8% agarose gel and Nano-drop 2000 (Thermo Scientific Inc, USA) by determining the A260/280 ratio. DNA concentration was checked by Qubit® 3.0 Fluorometer (Thermo Scientific Inc, USA) for library preparation.

Genome sequencing was performed at Eurofins Genomics India Pvt. Ltd. with the paired-end sequencing libraries prepared using TruSeq® Nano DNA Library Prep Kit for illumina (2 × 150 NextSeq-500 libraries). The mean of the library fragment size distributions of 463 bp were subjected to end-repair followed by adapter ligation to the fragments. The ligated products were size selected using AMPure XP beads and size-selected products were used in PCR amplification using the index primer. The PCR amplified libraries were analyzed in Tape Station 4200 (Agilent Technologies, USA) using High sensitivity D1000 Screen Tape assay kit as per manufacturer instructions. The PE illumina library with an insert size ranging from 245 to 894 bp was sequenced using 2 × 150 bp chemistry in NextSeq-500. A total of 2.6 Gb data representing high-quality reads of 8,994,174 were generated on NextSeq-500. High-quality paired-end short reads of *B. paralicheniformis* strain KMS 80 obtained from illumina NextSeq-500 were assembled into scaffolds using SPAdes (Version: 3.7.1) with default parameters (Bankevich et al. 2012).

In total, the assembly of 58 scaffolds resulted in a genome size of 4,566,040 bp with an average scaffold size of 55,193 bp and N_{50} value of 1,056,939 bp. The gene prediction was performed with the help of Prodigal (version 2.60) using default parameters (Hyatt et al. 2000; Dhakal et al. 2014). A total of 4598 genes in the range of 882–19,082 bp were predicted with an average size of 60 bp. We also predicted 92 RNA genes (20 rRNA, 67 tRNA, 1 tmRNA and 4 other RNA) using RNAmmer (version 1.2) and ARAGON algorithm (Laslett et al. 2004). A genome average GC-content of 45.50 mol% was estimated (Table 1). Predicted genes were mapped on reference canonical pathways and classified using an automated KASS server in KEGG (http://www.genome.jp/kaas-bin/kaas_main). Functional annotation of the genes was performed using the BLASTx program. Majority of genes showed homology with *Bacillus* sp. genome. Gene ontology distribution revealed 1221 genes responsible for biological process, 1315 genes responsible for molecular function and 899 genes responsible for cellular component synthesis, out of a total 4424 annotated genes. Twenty one genes involved in nitrogen metabolism pathway were annotated with the aim to identify genes for nitrogen fixation and their regulatory elements. Important regulatory genes like *glnA* (Type I glutamate–ammonia ligase/glutamine synthetase), *glnL* (two-component transcriptional response regulator), *glnR* (transcriptional factor/regulator), *glnT* (sodium: glutamine symporter), *tnrA* (nitrogen sensing transcriptional regulator) and several other elements

Table 1 Genome features of *Bacillus paralicheniformis* strain KMS 80 (MTCC No. 12704)

Features	Chromosome
Length	4,581,036 bp
G+C content (%)	45.50
Genes	4598
CDS	4424
rRNA genes	20 (5S, 16S & 23S)
tRNA genes	67
tmRNA gene	1
Other RNA	4
Promoter sequences	8692
Plasmid	1
CRISPR repeats	0
Phage elements	3

responsible for nitrogen metabolism were predicted from the whole genome sequence of *B. paralicheniformis* strain KMS 80. It has been reported that regulation of nitrogen fixation in *Bacillus* sp. is mediated by two transcriptional factor proteins, *gln*R and *tnr*A, that belong to the merR protein family (Wray and Fisher 2007, 2008). These transcriptional factors were annotated from the whole genome of *B. paralicheniformis* strain KMS 80 as CDS_3675 (*glnR*) in the scaffold-3 and CDS_507 (*tnrA*) in the scaffold-1, which regulates the expression of N-fixation genes in response to changes in nitrogen availability (Wray and Fisher 2007, 2008). The Multiple Sequence Alignment (MSA) and molecular phylogenetic analysis of *glnR* (Transcriptional regulator) and *nifH* (encodes nitrogenase iron protein also called component-2 of nitrogenase enzyme/protein) genes of *B. paralicheniformis* strain KMS 80 was carried out which showed wide genetic diversification (Fig. 1).

In addition, a total of 8692 promoter sequences were predicted using BPROM. One intact prophage region and three incomplete regions were identified using PHASTER (http://phaster.ca/server/) and there were no CRISPRs identified by CRISPRFinder. The information presented here will enable further study of the genetic and functional characteristics of rice-endophytic colonizing diazotroph, *B. paralicheniformis* strain KMS 80, and enhancing its genetic potential on biological N-fixation through biotechnological interventions.

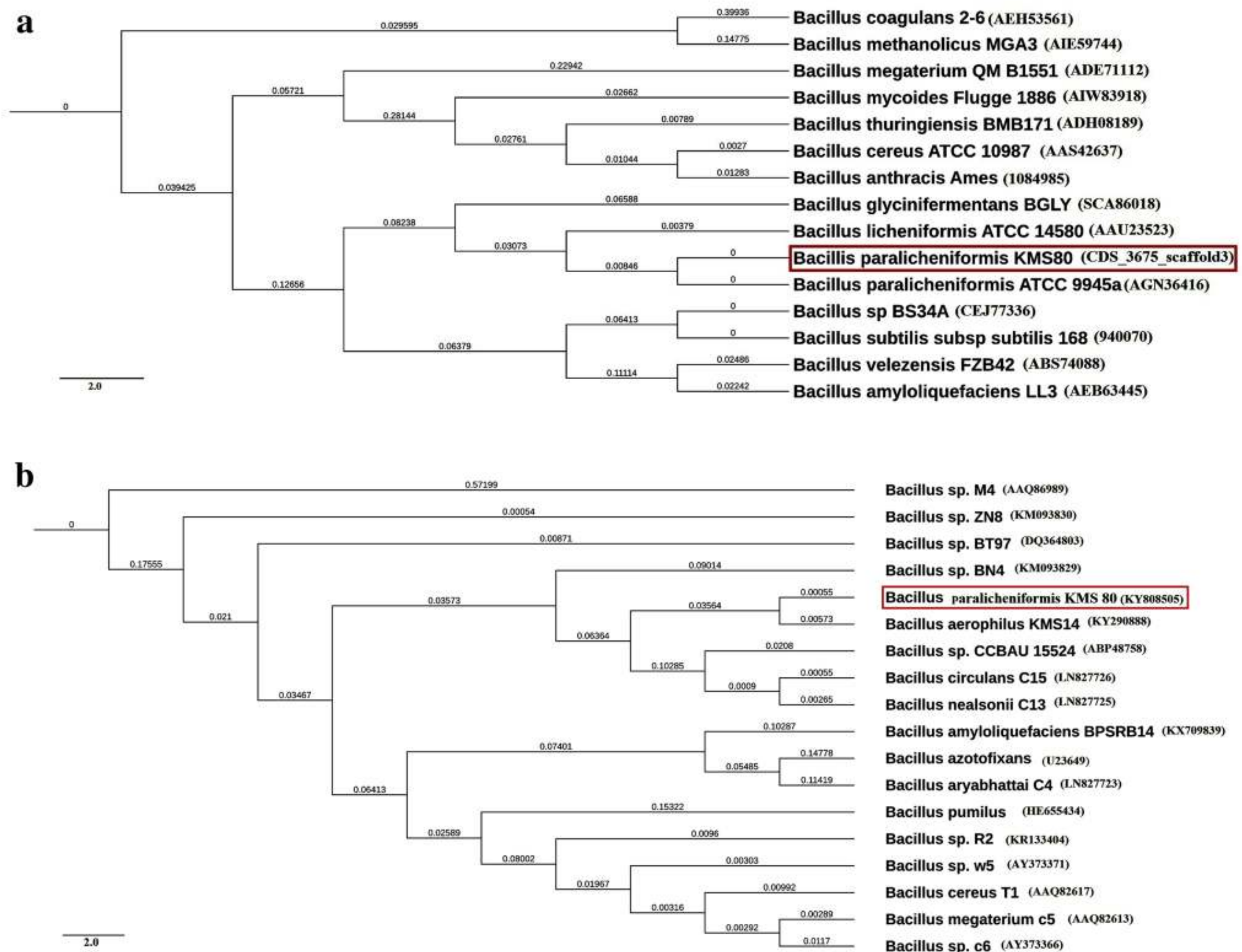

Fig. 1 Molecular phylogeny of *glnR* **(a)** and *nifH* **(b)** gene sequences of *Bacillus paralicheniformis* strain KMS 80 (MTCC No. 12704). The *glnR* and *nifH* gene sequences of *Bacillus paralicheniformis* strain KMS 80 (highlighted branch in red colour) was used in a BlastN search with default parameters against NCBI-nucleotide/protein databases. The phylogenetic tree is generated with phylogeny program (http://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny/) using default parameters and then visualized in FigTree (http://tree.bio.ed.ac.uk/software/figtree/). The taxon named with strain and gene sequence append with accession number in parentheses

Nucleotide sequence accession number

The genomic DNA identity was confirmed by sequencing a near complete 16S rRNA gene and sequence submitted to NCBI-GenBank under the accession number KY355731. The presence of *nif*H gene was validated by PCR amplification, sequence identity confirmed and also submitted to NCBI-GenBank under the accession number KY808505. The complete chromosome sequence of *B. paralicheniformis* strain KMS 80 has been deposited in DDBJ/ENA/GenBank under the accession number MUEI00000000 (BioProject: PRJNA360479). The version described in this paper is the first version, MUEH01000000. *B. paralicheniformis* strain KMS 80 under the catalogue MTCC No. 12704 is available from CSIR-Institute of Microbial Technology (CSIR-IMTECH) Chandigarh, India.

Acknowledgments We thank the Indian Council of Agricultural Research-Incentivizing Research in Agriculture on Biological Nitrogen Fixation [ICAR-IRA (BNF)], IARI project code: 21–47 for funding this work. The facility provided by the Director of ICAR-IARI, New Delhi is gratefully acknowledged.

Compliance with ethical standards

Conflict of interest The authors declare that they do not have any conflict of interest.

Ethical approval This article does not contain any studies with animals performed by any of the authors.

References

Annapurna K, Govindasamy V, Sharma M, Rajrana Y, Swarnalakshmi K, Paul S, Gosh A, Chikara SK (2017) Draft genome sequence of *Pseudomonas stutzeri* Strain KMS 55, an endophytic diazotroph isolated from rice roots. Genome Announcements: https://doi.org/10.1128/genomeA.00972-17

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477

Dhakal R, Seale RB, Deeth HC, Craven H, Turner MS (2014) Draft genome comparison of representatives of the three dominant genotype groups of dairy *Bacillus licheniformis* strains. Appl Environ Microbiol 80:3453–3462

Dunlap CA, Kwon SW, Rooney AP, Kim SJ (2015) *Bacillus paralicheniformis* sp. nov., isolated from fermented soybean paste. Int J Syst Evol Microbiol 65:3487–3492

Govindasamy V, Raina SK, George P, Kumar M, Rane J, Minhas PS, Vittal KPR (2017) Functional and phylogenetic diversity of cultivable rhizobacterial endophytes of sorghum [*Sorghum bicolor* (L.) Moench]. Antonie Van Leeuwenhoek 110:925–943

Hirsch PR, Mauchline TH (2012) Who's who in the plant root microbiome? Nature Biotechnol 30:961–962

Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2000) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform 11:119

Kumar U, Panneerselvam P, Govindasamy V, Kumar LV, Senthilkumar M, Banik A, Annapurna K (2017) Long-term aromatic rice cultivation effect on frequency and diversity of diazotrophs in its rhizosphere. Ecol Eng 101:227–236

Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16

Madslien EH, Olsen JS, Granum PE, Blatny JM (2012) Genotyping of *B. licheniformis* based on a novel multi-locus sequence typing (MLST) scheme. BMC Microbiol 12:230

Reinhold-Hurek B, Hurek T (2011) Living inside plants: bacterial endophytes. Curr Opin Plant Biol 14:435–443

Smith PA (2015) Untapped plant microbiome could help feed billions. Scientific Am 6th November.http://www.scientificamerican.com/article/untapped-plant-microbiome-could-help-feed-billions

Wray LV Jr, Fisher SH (2007) Functional analysis of the carboxy-terminal region of *Bacillus subtilis* TnrA, a MerR family member. J Bacteriol 189:20–27

Wray LV Jr, Fisher SH (2008) *Bacillus subtilis* GlnR contains an autoinhibitory C-terminal domain required for the interaction with glutamine synthetase. Mol Microbiol 68:277–285