

## Draft genome sequence of a thermostable, alkaliphilic $\alpha$ -amylase and protease producing *Bacillus amyloliquefaciens* strain KCP2

Vimalkumar S. Prajapati<sup>1</sup> · Sanket Ray<sup>2</sup> · Jitendra Narayan<sup>3</sup> · Chaitanya C. Joshi<sup>4</sup> · Kamlesh C. Patel<sup>2</sup> · Ujjval B. Trivedi<sup>2</sup> · R. M. Patel<sup>1</sup>

Received: 19 July 2017 / Accepted: 6 October 2017 / Published online: 13 October 2017  
© Springer-Verlag GmbH Germany 2017

**Abstract** *Bacillus amyloliquefaciens* strain KCP2 was isolated from municipal food waste samples collected in Vallabh Vidyanagar, Gujarat, India. Strain KCP2 is noteworthy due to its ability to produce a thermostable, alkaliphilic  $\alpha$ -amylase and a protease. These enzymes have importance in several industrial processes including bread making, brewing, starch processing, pharmacy, and textile industries. Whole genome sequencing of strain KCP2 showed that the estimated genome size was 3.9 Mb, the G + C content was 46%, and it coded for 4113 genes.

**Keywords** *Bacillus amyloliquefaciens* · Ion torrent · NGS · Whole genome sequencing · Amylases · Protease

Here, we report on a draft genome sequence of *Bacillus amyloliquefaciens* strain KCP2 which produces a thermostable alkaliphilic  $\alpha$ -amylase and protease, useful in bread making, brewing, starch processing, pharmacy, and

textile industries. Strain KCP2 was isolated from municipal food waste samples collected in Vallabh Vidyanagar, Gujarat, India (Prajapati et al. 2015a). The cells of the strain are aerobic motile rods arranged singly or in chains and stain Gram positive. The strain is oxidase and catalase positive, hydrolyses gelatine and casein, starch, and Tween 40 and Tween, suggesting that it is proteolytic, amylolytic, and lipolytic, respectively. The  $\alpha$ -amylase produced by strain KCP2 has been used in the saccharification and fermentation of raw corn starch and food waste for the production of bioethanol (Prajapati et al. 2015b). The strain is also positive in Vogel-Proskauer test and grows on glucose, mannitol, glycerol, glycogen, salicin, cellobiose, fructose, galactose, maltose, ribose, sorbitol, sucrose, and D-xylose while producing acid (Wang et al. 2008; Borriss et al. 2011).

Bacterial DNA from strain KCP2 was extracted and purified (Prajapati et al. 2015b), and single end sequencing was performed using a 318 chipset and 400 bp chemistry

**Accession number(s)** The whole genome data have been deposited at DDBJ/ENA/GenBank under the accession number NMRK00000000. The version described in this paper is version NMRK01000000.

✉ Vimalkumar S. Prajapati  
vimalprajapati100@gmail.com

<sup>1</sup> Aspee Shakilam Biotechnology Institute, Navsari  
Agricultural University, Surat, Gujarat 395 007, India

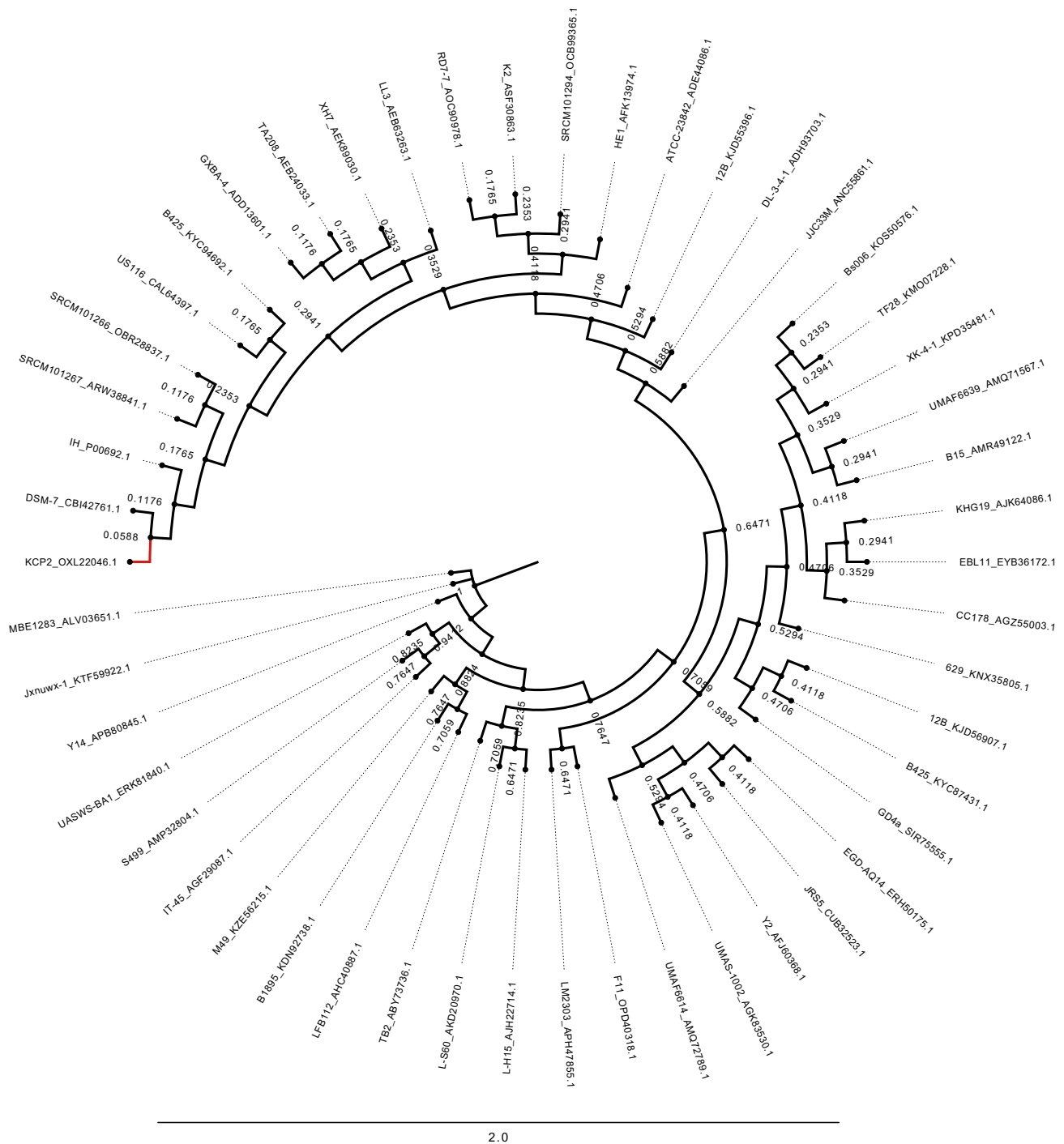
<sup>2</sup> Department of Biosciences, Satellite Campus, Sardar Patel  
University, Bakrol, Gujarat, India

<sup>3</sup> Research Unit in Environmental and Evolutionary Biology,  
University of Namur, Namur, Belgium

<sup>4</sup> Department of Animal Biotechnology, College of Veterinary  
Science and Animal Husbandry, Anand Agricultural  
University, Anand, Gujarat, India

**Table 1** *Bacillus amyloliquefaciens* strain KCP2 genome assembly data statistics from PGAP (v4.2) pipeline

Items	Count
Total Genes	4113
Total CDS	3999
Coding Genes	3681
Coding CDS	3681
Genes (RNA)	114
rRNAs	9, 9, 6 (5S, 16S, 23S)
tRNAs	85
ncRNAs	5
Total Pseudo Genes	318
tmRNA	1



**Fig. 1** Phylogenetic analysis of the  $\alpha$ -amylase protein sequence from *Bacillus amyloliquefaciens* strain. The  $\alpha$ -amylase protein sequence from *B. amyloliquefaciens* strain KCP2 (highlighted branch in red colour) was used in a Blast search with default parameters against protein databases of *B. amyloliquefaciens* strain. The phylogenetic tree is generated with simple phylogeny ([http://www.ebi.ac.uk/Tools/](http://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny/)

[phylogeny/simple\\_phylogeny/](http://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny/)) program using default parameters and then visualized in FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>). The taxon named after strain and appended accession number (<StrainName>\_<SequenceAccessionNumber>). The scale represents the number of differences between sequences (root age 1.0)

on an Ion Torrent Personal Genome Machine (PGM) housed at the Department of Animal Biotechnology College of Veterinary Science and Animal Husbandry, Anand Agricultural University, Gujarat, India using the standard

protocols described by the manufacturer (Roche Diagnostics Ltd., United Kingdom). In total, 2,590,427 reads were obtained, which were corrected with Pollux to 2,554,487 reads (Marinier et al. 2015). The genome was assembled

using SPAdes version 3.10.1 (Bankevich et al. 2012) and MIRA version 4.0.2 assembler (Chevreux et al. 2004). The two assemblies and reads were used to re-verify the longest and reliable contigs using custom Perl scripts. The assembled draft genome of *B. amyloliquefaciens* strain KCP2 consisted of 34 scaffolds, a N50 of 439,503 bp with the largest scaffold of 979,499 bp. The genome size was estimated to be 3,906,932 bp which had a GC content of 46%. BUSCO2 analysis (Simão et al. 2015) using bacterial lineage data set in the presence of conserved orthologous genes among species of the genus indicated 98% completeness. Gene annotation was performed using NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova et al. 2016), which identified 3681 protein-coding sequences (Table 1). The nearest alpha amylase sequence was found from *B. amyloliquefaciens* strain DSM-7, a soil bacterium characterized by its enormous potential to produce extracellular enzymes of industrial importance including amylases and proteases. The PGAP annotation confirmed the presence of amylase in the genome. The Multiple Sequence Alignment (MSA) and phylogenetic analysis of  $\alpha$ -amylase show wide diversification (Fig. 1). Further in silico study of  $\alpha$ -amylase may provide novel thermostability signatures. The genome sequence and their annotation reported here would also be useful genetic resource for engineering metabolism of *B. amyloliquefaciens* and improve their usability. The comparative genomics analysis of this strain with 52 sequenced genomes of *B. amyloliquefaciens* is in progress. We expect to yield a better understanding of *B. amyloliquefaciens* evolution through genome dynamics, population structure, and phylogenies of species groups.

#### Compliance with ethical standards

**Conflict of interest** The authors declare that there is no conflict of interest regarding the publication of this article.

#### References

- Bankevich A, Nurk S, Antipov D et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. doi:10.1089/cmb.2012.0021
- Borriess R, Chen X-H, Rueckert C et al (2011) Relationship of *Bacillus amyloliquefaciens* clades associated with strains DSM 7T and FZB42T: a proposal for *Bacillus amyloliquefaciens* subsp. *amyloliquefaciens* subsp. nov. and *Bacillus amyloliquefaciens* subsp. *plantarum* subsp. nov. based on complete genome sequence comparisons. *Int J Syst Evol Microbiol* 61:1786–1801. doi:10.1099/ijs.0.023267-0
- Chevreux B, Pfisterer T, Drescher B et al (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res* 14:1147–1159. doi:10.1101/gr.1917404
- Marinier E, Brown DG, McConkey BJ (2015) Pollux: platform independent error correction of single and mixed genomes. *BMC Bioinformatics* 16:10. doi:10.1186/s12859-014-0435-6
- Prajapati V, Trivedi U, Patel KC (2015a) Bioethanol production from the raw corn starch and food waste employing simultaneous saccharification and fermentation approach. *Waste and Biomass Valorization* 6:191–200. doi:10.1007/s12649-014-9338-z
- Prajapati VS, Trivedi UB, Patel KC (2015b) A statistical approach for the production of thermostable and alklophilic alpha-amylase from *Bacillus amyloliquefaciens* KCP2 under solid-state fermentation. *3. Biotech* 5:211–220. doi:10.1007/s13205-014-0213-1
- Simão FA, Waterhouse RM, Ioannidis P et al (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. doi:10.1093/bioinformatics/btv351
- Tatusova T, DiCuccio M, Badretin A et al (2016) NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. doi:10.1093/nar/gkw569
- Wang L-T, Lee F-L, Tai C-J, Kuo H-P (2008) *Bacillus velezensis* is a later heterotypic synonym of *Bacillus amyloliquefaciens*. *Int J Syst Evol Microbiol* 58:671–675. doi:10.1099/ijs.0.65191-0