Data in Brief 39 (2021) 107505

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Genome sequence data of *Bacillus amyloliquefaciens* L-17

Meriem Zaidi-Ait Salem, Elisabeth Girbal-Neuhauser, Yassine Nait Chabane*

Laboratoire de Biotechnologies Agroalimentaire et Environnementale (LBAE) URU 4565, UPS, Université de Toulouse, IUT Paul Sabatier, 24 rue d'Embaquès, Auch F-32000, France

ARTICLE INFO

Article history:
Received 21 June 2021
Revised 15 October 2021
Accepted 18 October 2021
Available online 23 October 2021

Keywords:
Complete genome
Bacillus amyloliquefaciens
Biofilm
antimicrobials

ABSTRACT

Bacillus amyloliquefaciens L-17 strain was isolated from a sample of chicken feathers. Here, we report complete genome sequence data of *B. amyloliquefaciens* L-17. The size of the genome is 3,933,788 bp which harbours 4001 coding Sequences. The BioProject has been deposited at NCBI GenBank. The GenBank accession numbers are PRJNA727793 for the BioProject, CP074391.1 for the chromosome, GCA_018363035.1 for GenBank assembly accession and SAMN19035411 for the BioSample.

* Corresponding author.
E-mail address: yassine.nait-chabane@iut-tlse3.fr (Y. Nait Chabane).

https://doi.org/10.1016/j.dib.2021.107505

Specifications Table

Subject	Biochemistry, Genetics and Molecular Biology
Specific subject area	Genomics and Microbiology
Type of data	Genome sequence data in FASTA format, table and figures
Data acquisition	Whole genome sequence of Bacillus *amyloliquefaciens* L-17 sequenced using a PacBio Sequel II System
Data format	Raw, analyzed and assembled genome sequences
Parameters for data collection	Genomic DNA was isolated from a pure culture of Bacillus *amyloliquefaciens* L-17
Description of data collection	Whole-genome sequencing, assembly, and annotation
Data source location	Bacillus amyloliquefaciens L-17, provided by the Culture Collection of the Laboratoire de Biotechnologies Agroalimentaire et Environnementale (culture collection WDCM 1016, LBAE-UPS, Auch, France), was isolated from a sample of chicken feathers from a local farm in west south of France
Data accessibility	Data are publicly available at NCBI Genbank https://www.ncbi.nlm.nih.gov/bioproject/PRJNA727793/ https://www.ncbi.nlm.nih.gov/assembly/GCA_018363035.1 https://www.ncbi.nlm.nih.gov/biosample/SAMN19035411/ https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA727793

Value of the Data

- Based on genome data, *Bacillus amyloliquefaciens* UPS-LBAE strain L-17 *could* be a potential strain for study of biofilm formation, and for investigation in biocontrol agents against plant pathogens and enzymes production.
- The data of this article could be useful for scientists working in the field of environmental microbiology, environmental biotechnology, genomics and genetic engineering.
- This genome data could be a valuable resource for comparative genomic analysis among *Bacillus amyloliquefaciens* strains.

1. Data Description

Bacteria under the Operational Group *Bacillus amyloliquefaciens* (OGBa) are Gram-positive, rod-shaped, and endospore-forming. Taxonomically, the OGBa belongs to the *Bacillus subtilis* species complex group, in *Bacillaceae* family. Currently, the OGBa comprises four bacterial species: *Bacillus amyloliquefaciens, Bacillus siamensis, Bacillus velezensis* and *Bacillus nakamurai* [1]. It was reported that *Bacillus* isolates from plants or soil are closely related to but distinct from *B. amyloliquefaciens* type strain DSM7^T; it is the case of *B. amyloliquefaciens* strain GB03 [2], and *B. amyloliquefaciens subsp. plantarum* type strain FZB42 which recently requalified as *Bacillus velezensis* FZB42 [3–4]. *B. velezensis* FZB42 has been shown to promote plant growth and is widely used in different commercial formulations as biofertilizers and biocontrol agents against plant pathogen [3–5].

A total of 3,819,059,724 reads were produced from paired-end sequencing of a genomic library with an average insert size of 9348.32 bp resulted in 1 circular contig with a total length of 3,933,788 bp. The N50 and maximum contig length were 387,471 bp and 784,095 bp, respectively. Automatic genome annotation, performed using the RAST server [6], predicted 113 RNA genes and 4001 coding sequences (Table 1, Fig. 1)

As shown in Fig. 1, proteins distribution is mostly related to metabolism, sporulation, adaptation and cofactors production. Moreover, based on RAST data, more than 14% of total proteins (587) are hypothetical, suggesting the potential of this bacterium to provide more information on its possible applications, but also the further work on the functional annotation of this strain. More interesting, the presence of proteins involved in biofilm formation such as TasA considered as a major proteinaceous component in addition to exopolysaccharides (EPS) in related *Bacillus*

Table 1
Characteristics of genome assembly of *B. amyloliquefaciens* L-17.

Number of contigs	1
Genome size (bp)	3,933,788
Number of Contigs (with PEGs)	1
GC content (%)	46.63
Number of Coding Sequences	4001
Number of subsystems	326
RNAs Genes	113

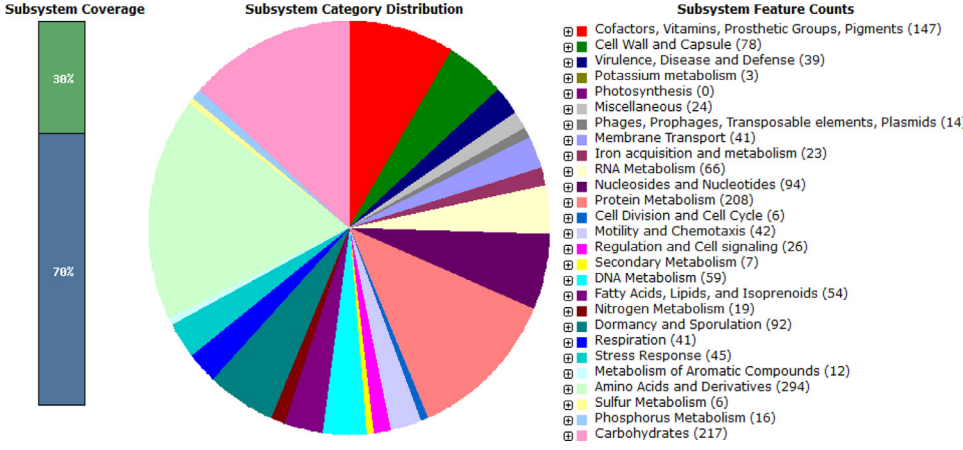

Fig. 1. Subsystem statistics information of *B. amyloliquefaciens* L-17 using RAST annotation server (RAST: Rapid Annotation Subsystem Technology version 2.2).

subtilis species [7,8], reflects the ability of *B. amyloliquefaciens* L-17 to adapt but also to persist the environment.

The potential of *B. amyloliquefaciens* L-17 to produce secondary metabolites, was predicted using anti SMASH server v.5.0 [9]. The results showed that the genome contained 12 gene clusters coding for enzymes involved in the biosynthesis of bacteriocins, non-ribosomal peptides synthases (NRPS), thiopeptide, siderophores, betalactone, terpenes, lanthipeptides-class-ii, polyketides synthases (PSK), transAT-PKS and type III polyketides. Four of the NRPS gene clusters exhibited 100% similarity with bacillaene, fengycin and difficidin gene clusters, respectively. An NRP (lipopeptide) identified as surfactin gene cluster exhibits 82% similarity. Moreover, bacilisin and macrolactin H gene clusters showed also 100% similarity.

Using the ContEst16S software, it has been indicated that the draft genome assembly did not have contamination of another prokaryotic genome. Analysis based on the 16S rRNA gene sequences revealed that *B. amyloliquefaciens* L-17 was affiliated to OGBa but this did not allow to distinguish it from other members of the group especially between conspecific complex consisting of *B. velezensis*, and *B. amyloliquefaciens*. Owing to limitations of the 16S rRNA gene comparative phylogenies, utilisation of housekeeping genes, such as DNA gyrase subunits A and B (*gyrA* and *gyrB*), signal transduction histidine kinase CheA (*cheA*) and RNA polymerase β-subunit (*rpoB*), allowed to improve speciation within the OGBa [10,11]. Otherwise, these phylogenetic analyzes revealed that *Bacillus amyloliquefaciens* L-17 is closely related to the *Bacillus amyloliquefaciens* species within the OGBa (Fig 2).

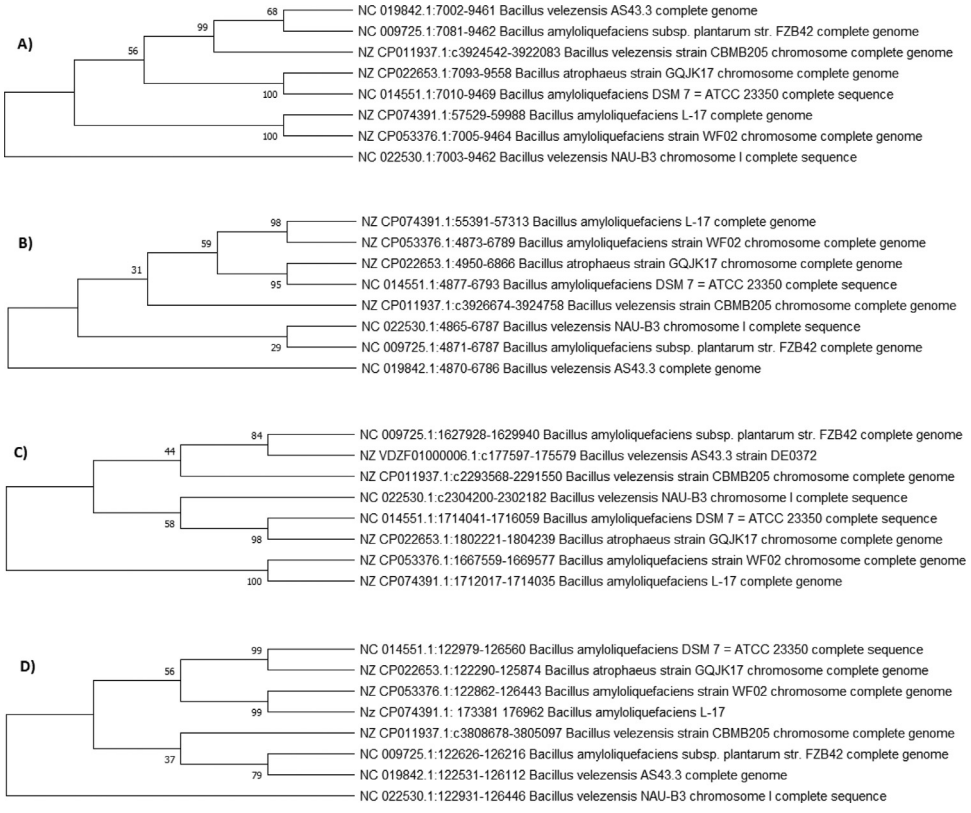

Fig. 2. Phylogenetic tree diagrams of *Bacillus amyloliquefaciens* L-17 strain generated using neighbour-joining based on (A) *gyrA* gene sequence (2460 pb), (B) *gyrB* gene sequence (1917 bp), (C) *CheA* gene sequence (2019 pb) and (D) *rpoB* gene sequence (3582 bp) show L-17 strain is closely related with the *Bacillus amyloliquefaciens* species. The numbers at branch nodes indicate percentages from 1000 bootstraps.

2. Experimental Design, Materials and Methods

2.1. Strain growth conditions and DNA isolation

Bacillus amyloliquefaciens L-17, provided by the Culture Collection of the Laboratoire de Biotechnologies Agroalimentaire et Environnementale (culture collection WDCM 1016, LBAE-UPS, Auch, France), was isolated from a sample of chicken feathers from a local farm in west south of France. Strain L-17 was grown in Trypticase soy agar (TSA) and Trypticase soy broth (TSB) for 24 h at 30 °C. Genomic DNA was extracted using a QIAamp DNA Stool Mini Kit (Qiagen, following manufacturer instructions). Extracted DNA was amplified by PCR with two 16S rDNA primer set (V2-4-8 and V3-6, 7-9) from the kit Ion 16S Metagenomics (ThermoFisher, following manufacturer instructions). Then, PCR were purified on magnetic beads (Agencourt, Beckman), and pooled. The library was prepared with the Ion Plus Fragment library kit (ThermoFisher, following manufacturer instructions). Sequencing was done on a 520 Ion Chip. 16S rDNA sequences were analyzed by Ion Reporter.

2.2. Genome sequencing, assembly, and annotation

The complete genome of *B. amyloliquefaciens* L-17 was carried out using a PacBio Sequel II System powered by Single Molecule Real-Time (SMRT) Sequencing technology (Pacific Biosciences, Menlo Park, CA, USA). De novo assembly was performed using PacBio Microbial Assembly Analysis Application SMRT Link v9.0, after quality trimming and filtering (about 893.66-fold coverage after pre-treatment of the reads),

The potential secondary metabolite biosynthetic gene clusters (BGCs) were identified in the genome using antiSMASH v5.0. The phylogenetic trees were constructed with the neighbour-joining method in MEGA 10.2.5.

The whole genome of *B. amyloliquefaciens* L-17 was analyzed by ContEst16S (Contamination Estimator by 16S), in which 16S rRNA gene fragments from the query genome assemblies are screened the genome assembly contamination [11].

Ethics Statement

This work did not involve human subjects, animals, cell lines or endangered species. The present manuscript is the author's original work, which has not been previously published elsewhere

Declaration of Competing Interest

The authors declare that they do not have conflict of interest that could influence the work reported in this paper.

CRediT Author Statement

Meriem Zaidi-Ait Salem: Visualization, Investigation; **Elisabeth Girbal-Neuhauser:** Writing – review & editing, Supervision; **Yassine Nait Chabane:** Conceptualization, Methodology, Software, Data curation, Writing – original draft.

Acknowledgments

This research is supported by a grant from the Occitanie Regional Council and funding from Toulouse University.

References

[1] M.S. Ngalimat, R.S.R. Yahaya, M.M.A. Baharudin, S.M. Yaminudin, M. Karim, S.A. Ahmad, S. Sabri, A review on the biotechnological applications of the operational group *Bacillus amyloliquefaciens*, Microorganisms 9 (2021) 614, doi:10.3390/microorganisms9030614.

[2] S.K. Choi, H. Jeong, J.W. Kloepper, CM. Ryu, Genome Sequence of *Bacillus amyloliquefaciens* GB03, an active ingredient of the first commercial biological control product, Genome Announc. 2 (2014) e01092-14, doi:10.1128/genomeA.01092-14.

[3] B. Fan, J. Blom, K. HP, R. Borriss, *Bacillus amyloliquefaciens, Bacillus velezensis*, and *Bacillus siamensis* form an "operational group *B. amyloliquefaciens*" within the *B. subtilis* species complex, Front. Microbiol. 8 (2017) 22, doi:10.3389/fmicb.2017.00022.

[4] C. Dunlap, S.J. Kim, S.W. Kwon, A. Rooney, *Bacillus velezensis* is not a later heterotypic synonym of *Bacillus amyloliquefaciens, Bacillus methylotrophicus, Bacillus amyloliquefaciens subsp. plantarum* and '*Bacillus oryzicola*' are later heterotypic synonyms of of *Bacillus velezensis* based on phylogenomics, Int. J. Syst. Evol. Microbiol. 66 (2016) 1212–1217, doi:10.1099/ijsem.0.000858.

[5] S.P. Chowdhury, A. Hartmann, X. Gao, R. Borriss, Biocontrol mechanism by root-associated *Bacillus amyloliquefaciens* FZB42 -a review, Front. Microbiol. 6 (2015) 780, doi:10.3389/fmicb.2015.00780.

[6] RK. Aziz, D. Bartels, AA. Best, et al., The RAST server: rapid annotations using subsystems technology, BMC Genom. 9 (2008) 75, doi:10.1186/1471-2164-9-75.
[7] A. Diehl, Y. Roske, L. Ball, A. Chowdhury, et al., Structural changes of TasA in biofilm formation of *Bacillus subtilis*, PNAS 115 (13) (2018) 3237–3242 USA, doi:10.1073/pnas.1718102115.
[8] Á.T. Kovács, A. Dragoš, Evolved biofilm: review on the experimental evolution studies of *Bacillus subtilis* pellicles, J. Mol. Biol. 431 (23) (2019) 4749–4759, doi:10.1016/j.jmb.2019.02.005.
[9] K. Blin, S. Shaw, K. Steinke, R. R Villebro, et al., antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline, Nucleic Acids Res. 47 (2019) W81–W87, doi:10.1093/nar/gkz310.
[10] M.S. Ngalimat, S. Sabri, Taxonomic note: speciation within the operational group *Bacillus amyloliquefaciens* based on comparative phylogenies of housekeeping genes, Asia Pac. J. Mol. Biol. Biotechnol. 28 (2) (2020) 19–26, doi:10.35118/apjmbb.2020.028.2.02.
[11] I. Lee, M. Chalita, S.M. Ha, S.I. Na, S.H. Yoon, J. Chun, ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences, Int. J. Syst. Evol. Microbiol. 67 (6) (2017) 2053–2057, doi:10.1099/ijsem.0.001872.