Whole-genome sequencing of *Bacillus Safensis* B7 isolated from Diplotaxis harra growing in the arid region of Tunisia

Maissa Ben-Jabeur[1*], Zayneb kthiri[1], Walid Hamada[1]

1 Laboratory of Genetics and Plant Breeding, National Institute of Agronomy of Tunis, University of Carthage. 43, Ave. Charles Nicolle, 1082 Tunis, Tunisia

*Corresponding author: maissa.benjabeur@gmail.com

Abstract

This study reports the complete genome sequence (CGS) of *Bacillus safensis* B7 isolated from the leaves of *Diplotaxis harra*, which was sampled from the Mednine province situated in the arid region of Tunisia. The endophytic strain has shown potential in plant growth promotion, and CGS will provide valuable insights into implications of genes required for its endophytic behavior and interactions with host plants, which facilitates its development and application as a biostimulant agent.

Introduction

Plants provide a wide range of niches for endophytic organisms that can penetrate and colonize the plant tissues and promote plant growth and resilience to biotic and abiotic challenges throughout the plant's life cycle (Santoyo et al. 2016). To seek potential candidates for biostimulant application in agricultural systems, the quest for well-adapted endophytic bacteria co-existing with native plants growing in extreme environments should be rewarding. *Diplotaxis harra* (Forssk.) Boiss, under the vernacular name "El harra", is a common annual species distributed in many regions of arid Tunisia in often sandy or gypseous soils with elevated salinity (Tlig et al., 2008; Akrout et al., 2010). There is scarce information in the literature reporting the microbiome co-existing with *D. harra*. Only one recent study reported the richness of their rhizospheric soil in beneficial fungal microorganisms (Mohamed et al., 2022). In this study, we sought to isolate potential endophytes present within the wild plant *Diplotaxis harra*. We report here the isolation and the whole genome sequence data of the bacterium, which we identified as a Tunisian strain of *Bacillus safensis* named B7.

Material and Methods

Collection of plant material

For the isolation of endophytic bacteria, leaves of *Diplotaxis harra* were collected from randomly selected wild plants from the arid region of Tunisia (Bou ghrara, Mednine, 33°31'58.0"N, 10°40'20.2"E). The soil is characterized as an arid, native saline soil (Omar et al., 2020). Samples were placed in clean plastic bags, brought to the laboratory, and used for further experimental purposes.

Isolation and purification of bacterial endophytes

Leaves were surface-sterilized by a stepwise washing procedure through the following protocol: 70% alcohol for 1 min, sodium hypochlorite (2,5% Cl^-) for 4 min, ethanol for 30 s, and 3 rinses in sterile distilled water, according to Costa et al. (2012). Afterwards, surface-sterilised leaves were macerated in buffer solution (0,9% NaCl) using a sterile mortar and pestle. Serial tenfold dilutions of the tissue extract were prepared in sterilized distilled water. From each dilution, 100 µl were spread on 10% tryptic soy agar (TSA)

plates and incubated at 28 ºC. Emerging bacterial colonies were purified through repetitive single-colony isolation.

Whole-genome sequencing of bacteria

The DNA extraction of B7, the Illumina genome sequencing, and assembly were conducted by the sequencing service provider microbesNG (The BioHub Birmingham, UK) according to their protocols accessible at https://microbesng.com/. Briefly, genomic DNA was subjected to whole-genome sequencing on an Illumina NovaSeq 6000 (Illumina, San Diego, USA) using a 250 bp paired end protocol. Reads were trimmed using Trimmomatic version 0.3 and then were de novo assembled into contigs using SPAdes version 3.7. The complete genome sequence of B7 was annotated using Prokka version 1.11. The preliminary taxonomic classification of the genome sequence was performed using Kraken. Prediction of protein coding features and tRNA was performed utilizing Prodigal version 2.6, while rRNA prediction was carried out using ARAGORN version 1.2.

Precise species identification and phylogenetic analysis

To confirm the taxonomic assignment generated from Kraken, the core genes 16S, rpoB, pycA, and aroE were used for the identification of the B7 strain. The sequences of the core genes were extracted from the assembled genome of the B7 strain by using the Standalone BLAST program. These sequences were further aligned and compared with the sequences of *Bacillus* species by using the NCBI Basic Local Alignment Search Tools, nucleotide (BLASTn) program available at NCBI (http:www.ncbi.nlm.nih.gov). Alignments of the sequences with the highest homologous hits mined from BLASTn and the gene sequences extracted from the assembled genome were performed using the MUSCLE program in MEGA version 11. Afterwards, a phylogenetic tree was constructed using the maximum likelihood method with 100 bootstrap replicates in MEGA. The phylogenetic tree was visualized with iTOL version 6.

Results

Genomic features

The assembly of the sequence reads resulted in 11 contigs, with the largest contig measuring 1,816,339 bp. The cumulative contig length of the genome assembly accounted for 3,708,130 bp with up to 105 X mean coverage. The genomic GC content is determined to be 41,47%. The other genome features are listed in Table 1.

Table 1. Summary statistics for the *Bacillus safensis* B7 genome

Biosample	SAMN37233912	**Mean coverage**	105,19
# contigs (>= 1000 bp)	11	**CDS**	3,740,0
Largest contig	1,816,339	**tRNA**	75
Total length	3,708,130	**tmRNA**	1
GC (%)	41,47	**Gene bank accession (assembly)**	JAZBQW000000000
N50	1,000,500	**Gene bank accession (reads)**	SRS18799449

Phylogenetic analysis

The phylogenetic analyses displayed a difference in the phylogenetic trees constructed from 16S rRNA, rpoB, pycA, and aroE genes (Figure 1). The strain B7 branched with different *Bacillus* species for the 16S rRNA gene with 100% intraspecies similarity; therefore, 16S was unsuitable molecular marker for the differentiation of these closely related *Bacillus* species. Next, pycA was investigated, and the resultant

phylogenetic tree showed that B7 branched with *B. safensis* and *B. pumilus* for pycA, with intraspecies similarity ranging from 99,65 to 98,73%. Based on the phylogenetic analyses of rpoB and aroE gene sequences, both genes were able to vividly distinguish between *B. safensis* and *B. pumilus* which displayed high specificity for *B. safensis* with sequence similarity ranging from 99,45 to 98,9% for rpoB and 99 to 97% for aroE. These results suggest that rpoB and aroE could be used as marker molecules for the identification of *B. safensis*.

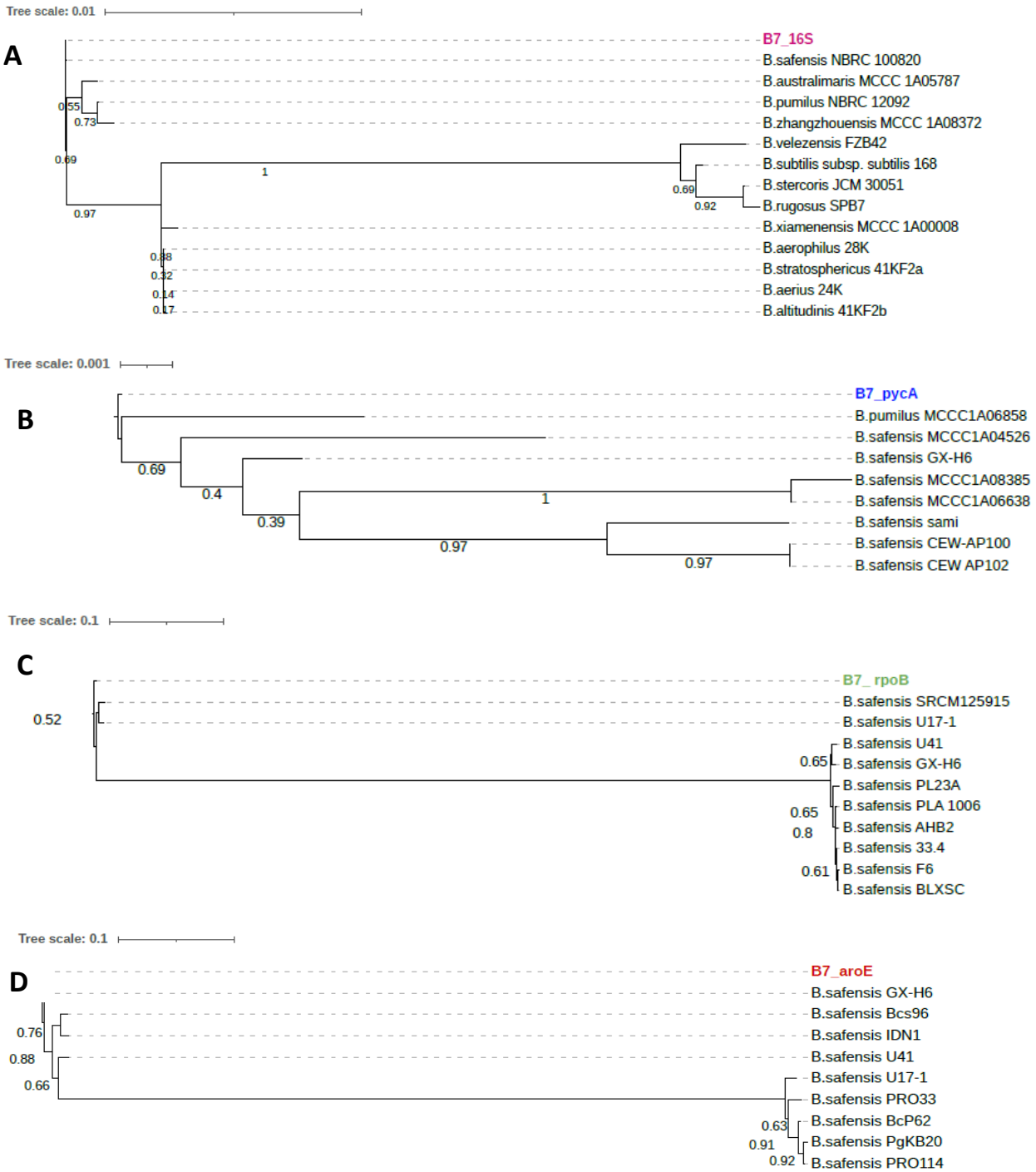


Fig 1. phylogenetic trees of *Bacillus safensis* B7. The phylogenetic trees are based on 16S (A), pycA (B), rpoB (C), and aroE (D) gene alignments obtained by MEGA 11 software using the maximum likelihood method.

Data availability

This whole-genome shotgun has been deposited at DDBJ/ENA/GenBank under the accession number JAZBQW000000000. Raw sequence reads have been deposited in the NCBI Sequence Read Archive under the BioProject number PRJNA1011052 and run numbers SRS18799449.

Statement on continuing work

The datasets are being shared prior to formal publication and should be regarded as preliminary.

Acknowledgements

The sequencing of the bacterial strains was supported by the GetGenome initiative of The Sainsbury Laboratory (TSL), Norwich, UK, with support from the Gatsby Charitable Foundation, the Biotechnology and Biological Sciences Research Council (BBSRC), and the University of East Angia. We would like to thank Pr. Sophien Kamoun for the development of the GetGenome program, and we are grateful to Dr. James Canham and Dr. Joe Win for their technical assistance.

References

Akrout, A., El Jani, H., Zammouri, T., Mighri, H., & Neffati, M. (2010). Phytochemical screening and mineral contents of annual plants growing wild in the southern of Tunisia. *Journal of Phytology*, *2*(1), 034-040.

Omar, Z., Bouajila, A., Bouajila, J., Rahmani, R., Besser, H., & Hamed, Y. (2020). Spectroscopic and chromatographic investigation of soil organic matter composition for different agrosystems from arid saline soils from Southeastern Tunisia. *Arabian Journal of Geosciences*, *13*, 1-13.

Costa, L. E. D. O., Queiroz, M. V. D., Borges, A. C., Moraes, C. A. D., & Araújo, E. F. D. (2012). Isolation and characterization of endophytic bacteria isolated from the leaves of the common bean (Phaseolus vulgaris). Brazilian Journal of Microbiology, 43, 1562-1575.

Tlig, T., Gorai, M., Neffati, M. (2008). Germination responses of Diplotaxis harra to temperature and salinity. *Flora-Morphology, Distribution, Functional Ecology of Plants*, *203*(5), 421-428.

Mohamed, A. H., Abd El-Megeed, F. H., Hassanein, N. M., Youseif, S. H., Farag, P. F., Saleh, S. A., ... & Abdel-Azeem, A. M. (2022). Native rhizospheric and endophytic fungi as sustainable sources of plant growth promoting traits to improve wheat growth under low nitrogen input. *Journal of Fungi*, *8*(2), 94.

Santoyo, G., Moreno-Hagelsieb, G., del Carmen Orozco-Mosqueda, M., & Glick, B. R. (2016). Plant growth-promoting bacterial endophytes. *Microbiological research*, *183*, 92-99.