

Unveiling the Genomic Sequences of Biocontrol Bacterial Agents Associated with *Solanum* Species in Tunisia

Manel CHAOUACHI^{1*}, Faouzia HADDADA¹, Bilel KHIARI¹, Cécile BEN^{2,3}, Laurent GENTZBITTEL^{2,3}, Naceur DJEBALI^{1**}

¹ Laboratory of Bioactive Substances, Centre of Biotechnology of Borj Cedria, BP 901, Hammam-Lif, 2050, Tunisia.

² Laboratoire Ecologie Fonctionnelle et Environnement, Université de Toulouse, CNRS, Toulouse, France.

³ Project Center for Agro Technologies, Skolkovo Institute of Science and Technology, Bolshoy Boulevard 30 bld.1, Skolkovo, Moscow Region 121205, Russian Federation.

Corresponding authors: *manelchaouachi812@gmail.com; [** dnaceur@yahoo.fr](mailto:**dnaceur@yahoo.fr)

Abstract

This study unveils the complete genome sequences obtained through whole-genome shotgun sequencing of four biocontrol agents originating from *Solanum* species samples gathered in Northern Tunisia. These genome sequences serve to enrich the current array of genomics data and provide valuable insights into the bacterial diversity intertwined with plants. Moreover, they offer a perspective into essential genes pivotal for their endophytic behaviors, interactions with host plants, and their effectiveness in biocontrol against plant pathogens.

Introduction

Endophytes are microorganisms that inhabit the internal tissues of plants without displaying any apparent indications of infection. These bacterial communities play a potential role in supporting plant well-being and growth by aiding in nutrient absorption and potentially suppressing harmful plant pathogens (Vejan et al., 2016; Afzal et al., 2019; Chaouachi et al., 2021; Marzouk et al., 2021). Numerous research studies have illustrated the concept of a beneficial partnership between plants and microorganisms, where bacteria release enzymes or substances outside their cells, enhancing mineral nourishment and consequently influencing plant health and growth. In return, plants provide nourishment to these beneficial microorganisms (Glick, 2015; Santoyo et al., 2016; Zhao et al., 2022). Endophytic Bacteria residing offer cost-effective and environmentally friendly options. These bacteria have found broad application in methods like biological control (Ayed et al., 2021; Chaouachi et al., 2021; Marzouk et al., 2021; Chaouachi et al., 2022) and the development of plant

biofertilizers (Vargas-Díaz et al., 2019; Bakhtiyarifar et al., 2021 ; Martínez-Hidalgo et al., 2021). In spite of numerous comprehensive reviews that outline functional models of plant growth-promoting bacteria, the precise mechanisms underlying the effectiveness of these bacteria remain elusive, necessitating further in-depth exploration. In the realm of natural product research, genome sequencing has acquired significant importance. The comprehensive analysis provided by whole-genome sequencing (WGS) offers a valuable approach for the identification of genes accountable for the synthesis of natural products (Belaouni et al., 2022). These BGC sequences can serve as predictive tools for deducing potential structures of the resultant natural product (Bauman et al., 2021), enabling the evaluation of novel compounds and the differentiation of compounds within a strain collection (Guo et al., 2020; Hwang et al., 2021). In addition, it allows for categorizing genes linked to plant growth-promoting (PGP) activities, simultaneously shedding light on their molecular and operational mechanisms (Guo et al., 2020). These encompassed nitrogen fixation, the synthesis of plant hormones, management of both biotic and abiotic stresses, the induction of resistance mechanisms, and effective root colonization (Guo et al., 2020).

Four endophytic bacterial strains AR11 isolated from *Solanum tuberosum* (unpublished data), DHT2, H1F1 and H2FL2 isolated from *S. lycopersicum* L. (Chaouachi et al., 2021). The choice of these endophytic bacteria was made based on their demonstrated antifungal capabilities against *Botrytis cinerea* strains via volatile organic compounds emission.

Previously, the identification of these isolates was achieved through analysis of the 16S RNA gene (unpublished data; Chaouachi et al., 2021). Nevertheless, in this current study, we present the Whole Genome Sequencing (WGS) of these bacterial strains, which offers benefits including enhanced recognition of bacterial species, heightened diversity detection, and improved predictive capacity for functional genes.

Material and Methods

Bacterial isolates

AR11 strain was isolated from the roots of *Solanum tuberosum*'s variety Annabelle according to Kalai-Grami et al. (2014) protocol. DHT2 , H1F1 and H2FL2 strains were isolated from stem, leaves and flower of *Solanum lycopersicum* L. according to Chaouachi et al. (2021).

Sequencing of bacterial genomic DNA

Bacterial genomic DNA sequencing procedures encompassed several steps. Bacterial culturing, DNA extraction, genome sequencing, and assembly were managed by microbesNG, adhering to their established protocols accessible at <https://microbesng.com/>. The acquisition of genomic DNA sequences from the bacterial isolates was facilitated through a sequencing service provider called MicrobesNG, situated in Birmingham, UK (<https://microbesng.com>). To ensure the proper handling of bacterial samples, guidelines provided by MicrobesNG were followed during the shipping process.

The sequencing reads underwent trimming using Trimmomatic version 0.39 (Bolger et al., 2014), employing a sliding window quality cutoff of Q15. Quality evaluation was carried out through a combination of in-house scripts, along with Samtools version 1.4 ([git://github.com/samtools/samtools.git](https://github.com/samtools/samtools.git)), BedTools version 2.18 (Quinlan and Hall, 2010), and bwa-mem (Li and Durbin, 2009) software. Subsequently, the sequence reads were assembled into contigs employing Quast software version 5.0.2 (Gurevich et al., 2013). The annotation of genomes was accomplished using Prokka 1.14.3 (<https://github.com/tseemann/prokka>). Prediction of protein coding features and tRNA was performed utilizing Prodigal version 2.6 (Hyatt et al., 2010), while rRNA prediction was carried out using ARAGORN version 1.2 (Laslett and Canback, 2004).

Results

We conducted genome sequencing on four biocontrol agents that were obtained from *S. lycopersicum* and *S. tuberosum*. The sequence reads were assembled using Quast, resulting in varying numbers of contigs, ranging from 42 to 142, with the largest contig measuring 1,128,577 base pairs. The cumulative contig lengths of the genome assemblies for our four strains—AR11, DHT2, H1F1 and H2FL2—are as follows: 4,168,360 bp; 5,928,369 bp; 4,285,063 bp and 4,056,308 bp respectively.

The genomic GC contents for these bacterial strains are as follows: 46,2 % for AR11, 35,28 % for DHT2, 45,69 % for H1F1 and 46,39 % for H2FL2. The total gene count, encompassing protein-coding genes, tRNA genes, and rRNA genes for each isolate, can be found in Table 1. Based on the ANI test and adhering to the current taxonomic nomenclature, the results indicate a sequence identity exceeding 99% between the submitted genome sequences and the *Bacillus* genus. The genomes of strains AR11, DHT2, H1F1 and H2FL2 exhibit identity to

distinct *Bacillus* species type genomes, displaying relatively lower identity to *B. licheniformis*, *B. pseudomycooides* and *B. velezensis* species.

Data availability

This whole-genome shotgun has been deposited at DDBJ/ENA/GenBank under accession numbers [JAPHWB000000000](#), [JAPHWH000000000](#), [JAPHWC000000000](#) and [JAPHWG000000000](#). Raw sequence reads have been deposited in the NCBI Sequence Read Archive under BioProject number PRJNA761700 and run numbers SRS15725548, SRS15725551, SRS15725549 and SRS15725550.

Statement on continuing work

The datasets are being shared prior to formal publication and should be considered preliminary. We encourage and welcome feedback from the community. Please get in touch with ND (dnaceur@yahoo.fr) for more information.

Acknowledgements

The sequencing of the bacterial strains was supported by GetGenome and The Sainsbury Laboratory, Norwich, UK, with contributions from the Gatsby Charitable Foundation, the Biotechnology, Biological Sciences Research Council (BBSRC) and The University of East Anglia. We would like to thank Pr. Sophien Kamoun for the development of GetGenome and we extend our appreciation to Dr. James Canham and Dr. Joe Win for providing technical assistance.

References

- Afzal I, Shinwari ZK, Sikandar S, Shahzad S (2019). Plant beneficial endophytic bacteria: mechanisms, diversity, host range and genetic determinants. *Microbiol Research*, 221, 36 - 49. <https://doi.org/10.1016/j.micres.2019.02.001>
- Ayed, A., Kalai-Grami, L., Ben Slimene, I., Chaouachi, M., Mankai, H., Karkouch, I., ... & Limam, F. (2021). Antifungal activity of volatile organic compounds from *Streptomyces* sp. strain S97 against *Botrytis cinerea*. *Biocontrol Science and Technology*, 31(12), 1330-1348.
- Bakhtiyarifar, M., Enayatizamir, N., & Mehdi Khanlou, K. (2021). Biochemical and molecular investigation of non-rhizobial endophytic bacteria as potential biofertilisers. *Archives of Microbiology*, 203, 513-521.

Bauman KD, Butler KS, Moore BS, Chekan JR (2021). Genome mining methods to discover bioactive natural products. *Nat Prod Rep*, 38, 2100–2129. <https://doi.org/10.1039/d1np00032b>

Belaouni HA, Compant S, Antonielli L, Nikolic B, Zitouni A, Sessitsch A (2022). In-depth genome analysis of *Bacillus* sp. BH32, a salt stress-tolerant endophyte obtained from a halophyte in a semiarid region. *Appl Microbiol Biotechnol* 106, 3113–3137. <https://doi.org/10.1007/s00253-022-11907-0>

Chaouachi, M., Marzouk, T., Aouini, J., Alaya, A. B., Khiari, B., & Djebali, N. (2022). Prospecting bacterial volatile organic compounds antifungal activities against postharvest diseases. *JOURNAL OF OASIS AGRICULTURE AND SUSTAINABLE DEVELOPMENT*, 4(3), 86-99.

Chaouachi, M., Marzouk, T., Jallouli, S., Elkahoui, S., Gentzbittel, L., Ben, C., Djébal, N. (2021). Activity assessment of tomato endophytic bacteria bioactive compounds for the postharvest biocontrol of *Botrytis cinerea*. *Postharvest Biology and Technology*, 172, 111389. <https://doi.org/10.1016/j.postharvbio.2020.111389>

Glick, B. R. (2015). Beneficial plant-bacterial interactions (pp. 1-28). Heidelberg: Springer.

Guo DJ, Singh RK, Singh P, Li DP, Sharma A, Xing YX, Song XP, Yang LT, Li YR (2020) Complete genome sequence of *Enterobacter roggenkampii* ED5, a nitrogen fixing plant growth promoting endophytic bacterium with biocontrol and stress tolerance properties, isolated from sugarcane root. *Front Microbiol* 11,580081. <https://doi.org/10.3389/fmicb.2020.580081>

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072-1075. <https://doi.org/10.1007/978-3-030-44368-9>

Hwang HH, Chien PR, Huang FC, Hung SH, Kuo CH, Deng WL, Chiang EI, Huang CC (2021) A plant endophytic bacterium, *Burkholderia seminalis* strain 869T2, promotes plant growth in *Arabidopsis*, Pak Choi, Chinese amaranth, lettuces, and other vegetables. *Microorganisms* 9, 1703. <https://doi.org/10.3390/microorganisms9081703>

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, 11(1), 1-11. <https://doi.org/10.1186/1471-2105-11-119>

Kalai-Grami, L., Ben Slimane, I., Mnari-Hattab, M., Rezgui, S., Aouani, M.A., Hajlaoui, M.R., Limam, F., 2014. Protective effect of *Bacillus amyloliquefaciens* against infections of *Citrus aurantium* seedlings by *Phoma tracheiphila*. *World J. Microbiol. Biotechnol.* 30, 529–538.

Laslett, D., Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic acids research*, 32(1), 11-16. <https://doi.org/10.1093/nar/gkh152>

Li, H., Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, 25(14), 1754-1760. <https://doi.org/10.1093/bioinformatics/btp324>

Martínez-Hidalgo, P., Flores-Félix, J. D., Sánchez-Juanes, F., Rivas, R., Mateos, P. F., Santa Regina, I., ... & Velázquez, E. (2021). Identification of canola roots endophytic bacteria and analysis of their potential as biofertilizers for canola crops with special emphasis on sporulating bacteria. *Agronomy*, 11(9), 1796.

Marzouk, T., Chaouachi, M., Sharma, A., Jallouli, S., Mhamdi, R., Kaushik, N., & Djébali, N. (2021). Biocontrol of *Rhizoctonia solani* using volatile organic compounds of solanaceae seed-borne endophytic bacteria. *Postharvest Biology and Technology*, 181, 111655.

Quinlan, A. R., Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841-842. <https://doi.org/10.1093/bioinformatics/btq033>

Santoyo, G., Moreno-Hagelsieb, G., del Carmen Orozco-Mosqueda, M., Glick, B. R. (2016). Plant growth-promoting bacterial endophytes. *Microbiological research*, 183, 92-99. <https://doi.org/10.1016/j.micres.2015.11.008>

Vargas-Díaz, AA; Ferrera-Cerrato, R; Silva-Rojas, HV; Alarcón, A (2019). Isolation and evaluation of endophytic bacteria from root nodules of *Glycine max* L. (Merr.) and their potential use as biofertilizers. *Spanish Journal of Agricultural Research*, Volume 17, Issue 3, e1103. <https://doi.org/10.5424/sjar/2019173-14220>

Vejan P, Abdullah R, Khadiran T, Ismail S, Nasrulhaq Boyce A (2016) Role of plant growth promoting rhizobacteria in agricultural sustainability—a review. *Molecules* 21, 573. <https://doi.org/10.3390/molecules21050573>

Zhao D, Ding Y, Cui Y, Zhang Y, Liu K, Yao L, Han X, Peng Y, Gou J, Du B, Wang C (2022) Isolation and genome sequence of a novel phosphate-solubilizing rhizobacterium *Bacillus altitudinis* GQYP101 and its effects on rhizosphere microbial community structure and functional traits of corn seedling. *Curr Microbiol* 79, 249. <https://doi.org/10.1007/s00284-022-02944-z>

Table 1: Summary statistics for Biocontrol agents genomes assembled from Illumina reads.

Bacterial strains	Host plants	Biosamples	# contigs (>=1000bp)	largest contig	total length	GC (%)	mean coverage	N50	CD S	tRNA	rRNA	GenBank accession (Assembly)	GenBank accession (reads)
<i>Bacillus licheniformis</i> AR11	<i>Solanum tuberosum</i>	SAMN31677948	24	664416	4177609	46,2	119,814	392589	4,558	79	11	JAPHWB000000000	SRS15725548
<i>Bacillus pseudomycoides</i> DHT2	<i>Solanum lycopersicum L.</i>	SAMN31677951	63	939586	5944900	35,28	71,9981	427108	6,431	107	16	JAPHWH000000000	SRS15725551
<i>Bacillus velezensis</i> H1F1	<i>Solanum lycopersicum L.</i>	SAMN31677949	23	643720	4291851	45,69	66,1125	517572	4,532	86	14	JAPHWC000000000	SRS15725549
<i>Bacillus velezensis</i> H2FL2	<i>Solanum lycopersicum L.</i>	SAMN31677950	13	1128577	4062032	46,39	55,4613	971828	4,135	84	12	JAPHWG000000000	SRS15725550