

Whole-genome sequence and genomic analysis of *Saccharothrix longispora* strain MB29, a potential biocontrol agent

Khaoula Bouznada¹, Hadj Ahmed Belaouni¹, Samira Tata¹, Eliška Peňázová², Jana Čechová², Akila Berraf-Tebbal², Aleš Eichmeier², Abdelghani Zitouni¹ and Atika Meklat¹

1: Laboratoire de Biologie des Systèmes Microbiens (LBSM), Ecole Normale Supérieure de Kouba, Algiers, Algeria.

2: Faculty of Horticulture, Mendeleum-Institute of Genetics, Mendel University in Brno, Valtická 334, 69144 Lednice, Czech Republic

khaoula.bouznada@g.ens-kouba.dz

This study describes the high-quality draft genome of *Saccharothrix longispora* strain MB29, a biocontrol agent found to be active against the durum wheat pathogen *Bipolaris sorokiniana*. The Illumina MiniSeq platform was used to sequence the entire genome, yielding 1,588,534 x 2 paired-end (2 x 150-bp) reads. 750 contigs (637 contigs 1000 bp) were assembled using the SPAdes (3.12.0) tool. *Saccharothrix longispora* strain MB29 has a genome length of 8,449,678 bp and a GC percentage of 73.15 (N50 = 23603, L50 = 117). The Quast tool (5.0.2) was used to evaluate the quality of the assembly. The Bowtie2 tool (2.4.2) was used to map the reads (mapped reads = 98.25 percent, coverage = 31.69 11.63). The annotation was performed using Prokka (1.14.0) (CDS: 7547, rRNA: 4, tRNA: 68). The genome quality was determined using CheckM (1.0.18), Busco (5.2.2), and the Patric portal's genome quality-control pipeline (3.6.10). They all advocate for a high-quality near-complete draft genome (CheckM: completeness = 100%, contamination = 2.67%; Busco: 357 total genes, 2 duplicated, 1 fragmented, 1 missing; Patric genome quality report: genome quality = good, coarse consistency = 98.5%, fine consistency = 95.2%). Phylogenomic analysis using the Type (Strain) Genome Server (TYGS) confirmed the strain's identification as *Saccharothrix longispora*. Speciality genes (antibiotics resistance/virulence) were detected using Abricate (1.0.1); the Patric pipeline (3.6.10); and the ShortBRED + RGI (Resistance Gene Identifier) under the Edge portal (2.4.0). Numerous biosynthetic gene clusters (BGCs) were predicted using Antismash (6.0.1) (58 BGCs). The present genome will be subsequently the object of genomic mining aiming to decipher the plant growth/protection mechanisms of *Saccharothrix longispora* strain mb29.

Keywords: *Saccharothrix*, High-quality draft genome, genomic analysis, annotation, BGC.