XIII Argentine Congress of Bioinformatics and Computational Biology
XIII International Conference of the Iberoamerican Society of Bioinformatics
III Annual Meeting of the Ibero-American Artificial Intelligence Network for Big BioData

Genome sequencing and characterization of *Bacillus velezensis* strain NKG50

Fernandez, FD[1]; Sardo MF[1]; Monteoliva, M[2]; Valetti, L[1]

1-Instituto de Patología Vegetal (CIAP-INTA) 2-Instituto de Fisiología y Recursos Genéticos Vegetales (CIAP-INTA). Av 11 de septiembre 4755, Córdoba (X5020ICA)-Argentina

fernandez.franco@inta.gob.ar

Background

Bacillus velezensis is a Gram-positive bacterium that colonizes the plant rhizosphere and has been widely studied for its growth-promoting and biological control effects on fungi, bacteria, and nematodes

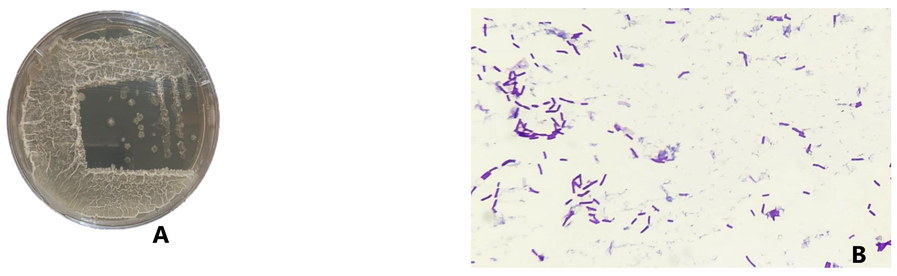
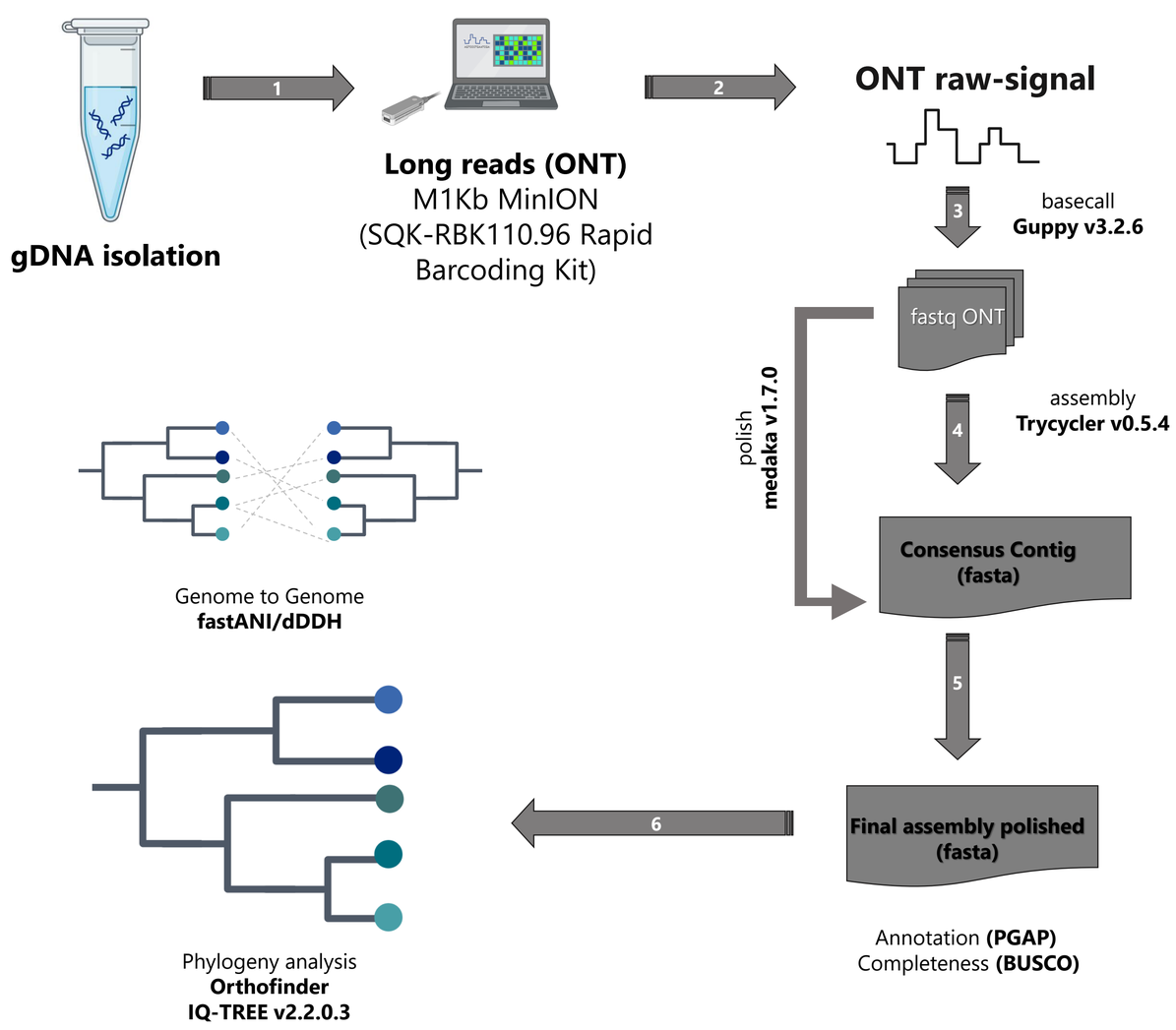


Figure 1: *Bacillus velezensis* strain NKG50 growing in TSA medium (A); and gram stain (100x) (B)

In previous works, the chickpea endophytic isolate ***Bacillus velezensis* strain NKG50** (Figure 1) was selected due to its biocontrol capacity against different pathogens of agricultural importance *(Thecaphora frezzi*, *Ascochyta rabiei* and *Fusarium sp.*), highlighting the production of different **antifungal metabolites**. Considering the beneficial effect of this microorganism, the sequencing and study of its genome will allow determining the beneficial potential of the strain and the **potential pgpr mechanisms**, which means an important contribution to the development of a bioinoculant to improve crop productivity

This study aims to sequence and characterize the genome of *Bacillus velezensis* strain NKG50 and conduct comparative genomics studies among representative species within the *Bacillus* clade.

Materials and Methods



Results

The genome sequencing of ***Bacillus velezensis* strain NKG50** generated 1.3 Gb of ONT reads (72.589 reads, N50 19 kb). The final genome was assembled into a single circular contig, comprising 4.123.916bp (GC% 46.00) (Figure 2). In the annotation process, 4.048 predicted CDSs (including 462 hypothetical proteins), 9 complete rRNA operons (6 5S, 6 16S, and 6 23S), and 85 tRNAs were annotated in the chromosome. Based on **BUSCO analyses**, the assembly was found to be **94.4%** complete (with 425 complete and single-copy BUSCO groups out of 450 total BUSCO groups from the bacilalles_odb10 lineage).

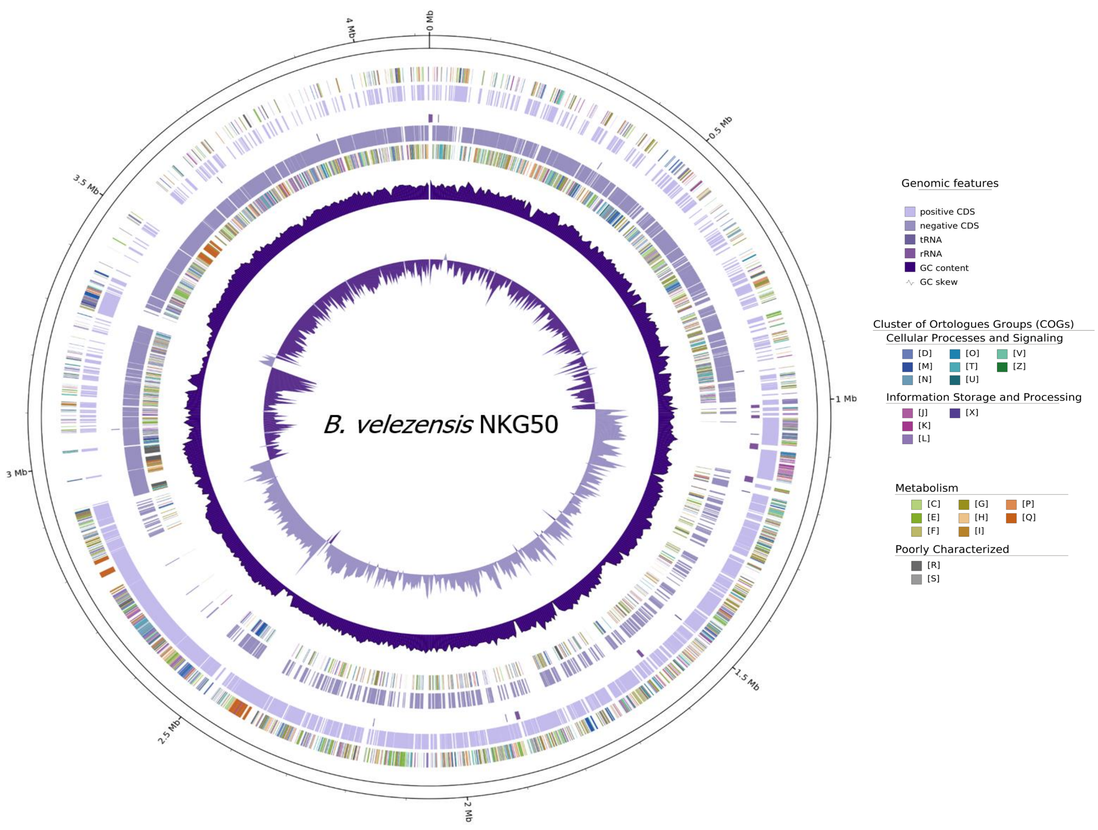


Figure 2: Circular map representation of ***Bacillus velezensis* strain NKG50** complete genome. Labelling from outside to inside: Contigs, COGs on the forward strand, CDS, tRNAs, and RNAs on the foward strand, CDS, tRNAs, and RNAs on the reverse strand, COGs on the reverse strand, GC content and GC skew. .

All genome sequences were subjected to digital DNA-DNA hybridization (dDDH) and average nucleotide identity (ANI) evaluation to obtain genome distance metrics. The ***B. velezensis* strain NKG50 genome** showed values >96% in ANI comparison and >70% for dDDH (Figure 4) within the *B. velezensis* co-especies clade (Species number 5 to 10) , confirming the taxonomic identification as ***B. velezensis*** provided by phylogenomics analysis.

The phylogenomic analysis involved comparing the NKG50 with **18 selected genomes** from other Bacillus species. The study of orthologs identified 58 single-copy genes shared among all species studied. The phylogenetic tree constructed from the DNA sequence of these **58 SCGs (40,806 bp)** showed that the NKG50 genome clustered within the *B. velezensis* species with high support (Figure 3)

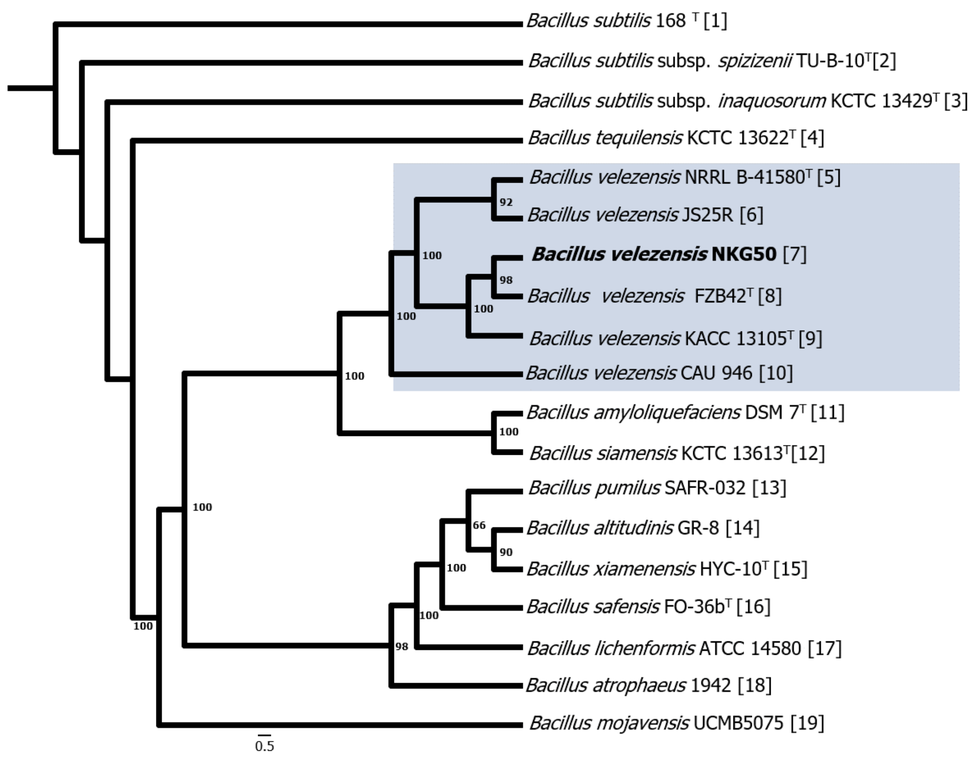


Figure 3: Maximum likelihood phylogenies. The numbers on the internal branches indicate the level of bootstrap support based on 1,000 resampling; only values >=60% are shown. The concatenated alignment contains 58 genes (Single-copy coding genes shared by all strains) and 40.806 aligned nucleotide sites. In light blue *B. velezensis* co-species clade. Bar: 0.5 substitutions per nucleotide position.

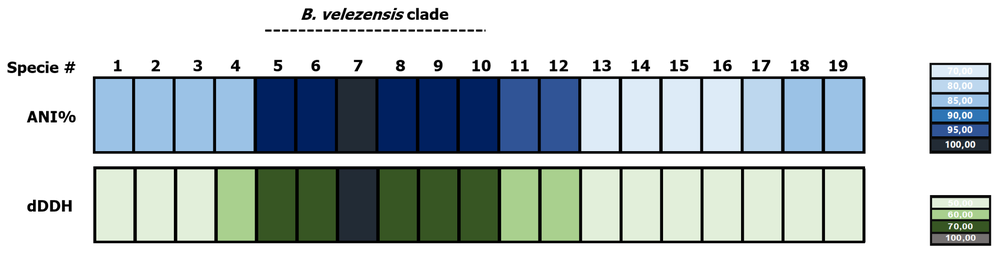


Figure 4 Genome-to-Genome comparison statistics of ***B. velezensis* strain NKG50** among all genomes analyzed in this study. The species numbers correspond to those described in the Phylogenetic tree. Average nucleotide identity (ANI) values were calculated using fastANI v1.1, and dDDH values were determined using the genome-to-genome distance calculator online tool (https://ggdc.dsmz.de/home.php). The scale of values is represented with different colors next to each bar.

Conclusions

This study provides a valuable genomic resource that enhances our understanding of the taxonomic identity, evolutionary relationships and biology of the ***B. velezensis* strain NKG50** along with its relevance as a biocontrol agent.

Fundings

PICT-2018-02410 , Proyectos INTA-2023: 2023-PD-L03-I084 and PD-L01-I087