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

Integration of Genomic Data Unveils Potential PGPT Activities in Antarctic Endophytic Bacteria

Orlowski, J.[12], Massot, F.[123], Basile, C.[123], Ruberto, L.[123]
[1]NANOBIOTEC-CONICET, [2]FFyB-UBA, [3]IAA

INTRODUCTION

In the context of an Antarctic soil bioremediation[1] project targeting hydrocarbon-contaminated soils, we isolated endophytic bacteria from one of the two indigenous plant species in the continent, *Deschampsia antarctica*. This plant species will play a crucial role in the setting-up of ecopiles[2] for on-site soil remediation.
The isolated bacteria were subjected to DNA Illumina sequencing, followed by genome assembly and gene/protein annotation. This study focuses on six isolates belonging to the *Agreia* genus, considering its high abundance in culture-dependent methods. By combining these isolated genomes with existing database entries, we constructed the *Agreia* pangenome, encompassing the gene pool of this genus. Subsequently, a bioprospecting and genome-mining effort was conducted to identify genes associated with plant growth-promoting traits (PGPT-proteins) activities, potentially enhancing the growth of these two plant species.
In the pursuit of bioremediation, understanding the genetic underpinnings of beneficial plant-microbe interactions is pivotal. The endophytic bacteria within indigenous Antarctic plants offer unique insight into their adaptive capabilities in extreme conditions.

MATERIALS AND METHODS

DNA extraction and Sequencing
_ 12 Raw Dataset, 6 of Agreia sp.
_ Format: fastq
_ Read Type: Paired-end
_ Sequencer Type: Illumina System
_ Read Length: 151 bp

Reads quality control
_Quality control with FastQC
_Removal of duplicate reads (Fastuniq)
_Cleaning with Trimmomatic:
_Filtering reads by minimum length (50 bp) Filtering reads by average quality (Phred score 30) Trimming bases from the ends of the reads Removal of "Nextera Transposase Adapters"

De novo Assembly
_With Unicycler
_Bandage for graph visualization
_Quast to evaluate the assemblies
_BBMap was used to align the read sequences (mapping)
_Samtools was used to calculate the sequencing depth from the alignments.
_IGV for mapping visualization
_Qualimap to analyse

Genome annotation
_Prokka, RAST, GTDB-Tk

Analysis
_Genome Contamination, taxonomic assignation and genome relatedness: ContEst16s, MIGA and Kraken
_Mauve (Multiple genome alignment)

Pangenome analysis
_Roary: Takes annotated assemblies and calculates the pan genome

Genome-mining
_PGPT-Pred tool was used for predicting plant growth-promoting traits (PGPTs) of single bacterial genomes. This tool is part of PLaBAse

RESULTS

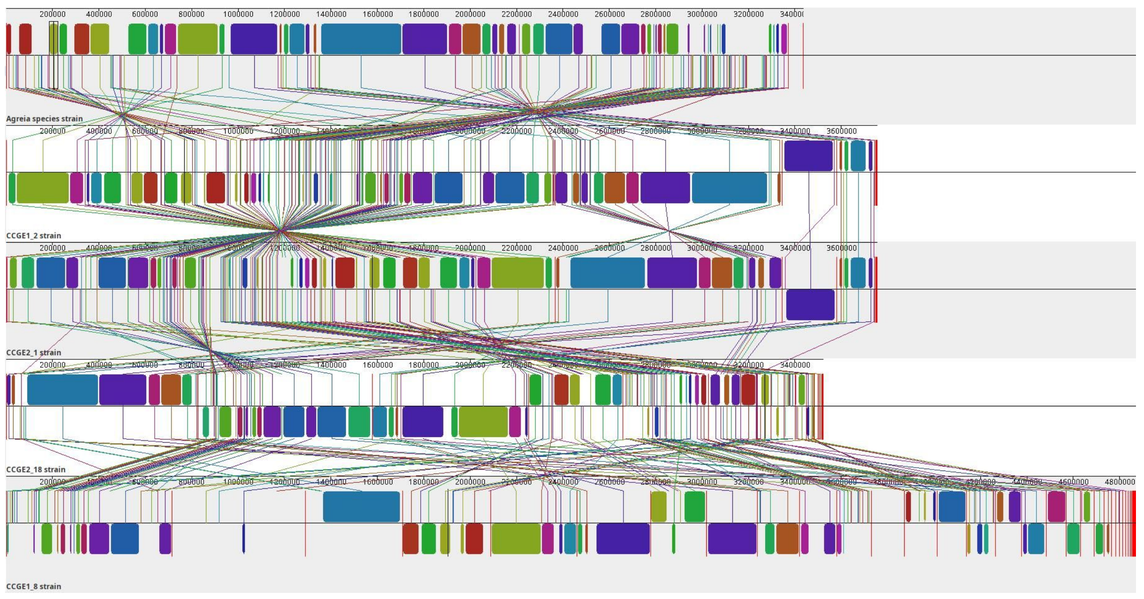

Figure1: multiple alignment of four representative *Agreia* genomes assembled, and a reference (first).

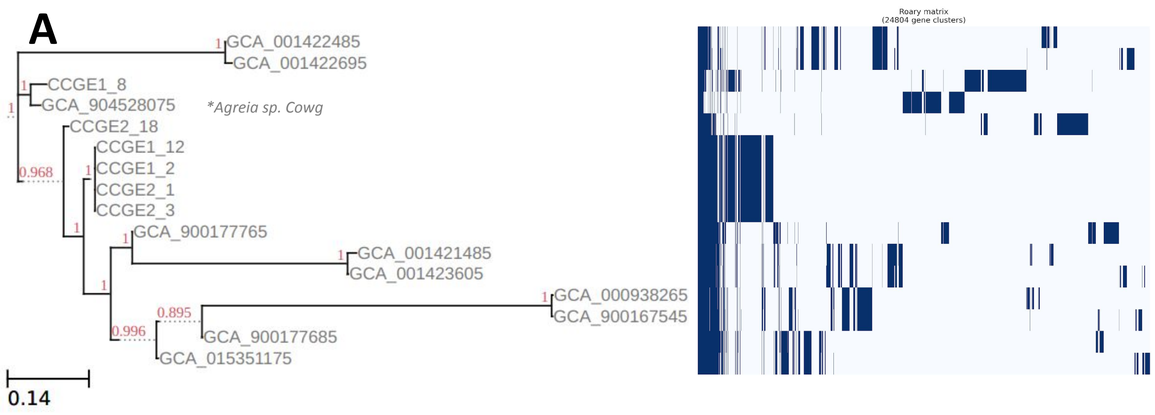

Figure2: Pangenome. A. Heatmap and dendrogram showing gene presence/absence, B. Gene distribution in the pangenome with respect to gene presence among 16 strains, C. Venn diagram representing the number of genes and the percentage shared between four representative genomes and the reference genome (*Agreia sp. COWG)*

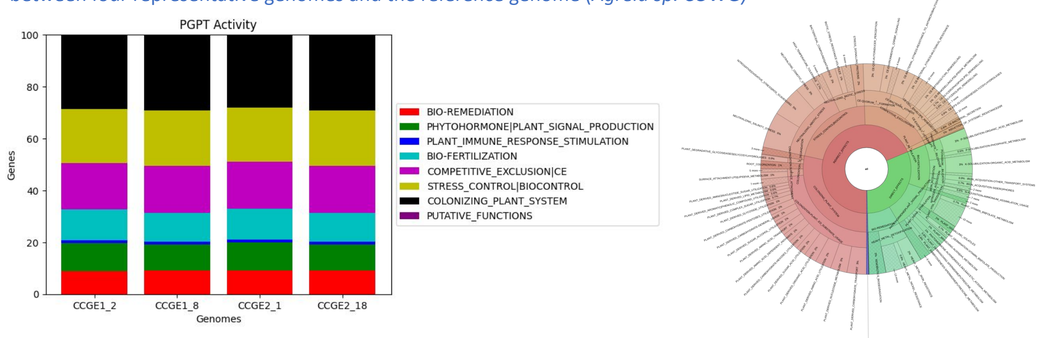

Figure3: Genome mining analysis: A. Illustrates the genomic distribution of functional categories in four different strains . Each strain's genome is categorized into eight functional classes, and the percentages of genes within these classes. B. Krona visualization

CONCLUSIONS

This study unearths a wealth of potential PGPT-associated genes within the genomes of endophytic bacteria. By constructing the *Agreia* pangenome and pinpointing genes linked to beneficial activities, we have begun to shed light on the mechanisms that could enhance the growth of the indigenous Antarctic plant species involved in soil bioremediation efforts. The combination of culture-dependent methods with genomic studies has become crucial for the strategy of defining plant-microorganism combinations in applications within the field of environmental biotechnology. These findings hold significance for the success of the bioremediation project and provide a stepping-stone for further research into the genetic factors driving plant-microbe interactions in extreme environments.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to the following individuals and organizations for their valuable contributions and support:
CONICET: For their continued support and funding, which made this research possible.
Instituto NANOBIOTEC: For their collaboration and access to cutting-edge nanobiotechnology facilities.
IAA (Instituto Antártico Argentino): For their support and resources that facilitated research in the challenging Antarctic environment.
HASSELT UNIVERSITY: For providing us the sequencing facilities
The dedicated Argentine team at Base Carlini in Antarctica: For their extraordinary efforts, assistance, and insights during our research in the harsh conditions of the Argentine Antarctic base.
To all our colleagues and collaborators, both near and far, who contributed to the success of this project.
To our families and friends: For their unwavering patience and support throughout this endeavor.

REFERENCES

[1]Massot, F., Bernard, N., Alvarez, L. M. M., Martorell, M. M., Mac Cormack, W. P., & Ruberto, L. A. M. (2022). Microbial associations for bioremediation. What does "microbial consortia" mean?. Applied microbiology and biotechnology, 106(7), 2283–2297. https://doi.org/10.1007/s00253-022-11864-8
[2]Lucas Ruberto, Lucas Martínez Álvarez, Francisco Massot and Walter Mac Cormack. Management and Bioremediation of Hydrocarbon-polluted Soils in Antarctica. Antarctic Affairs. 2020, VII: 53-64