

SUPPLEMENTARY MATERIAL

Comparative, Pangenomic and Functional Analyses of two *Bacillus paralicheniformis* Soil-Isolated Strains from Bahia Sequenced by WGS Reveal Species Homogeneity and Bioactive Metabolites with Biotechnological Potential

Gabriel Camargos Gomes¹, Eduarda Guimarães Sousa¹, Ludmila Silva Quaresma¹, Rhayane Cristina Viegas Santos, Gabriela Munis Campos¹, Janaíne Aparecida de Paula¹, Marcus Vinícius Canário Viana¹, Rafael Junio de Oliveira¹, Giovanna Karine Viana Silva¹, Bertram Brenig², Vasco Azevedo^{1*}

¹ Cellular and Molecular Genetics Laboratory, Department of General Biology, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

² Institute of Veterinary Medicine, University of Göttingen, Göttingen, Germany

*Correspondence: Vasco Azevedo, vascoariston@gmail.com

Journal: World Journal of Microbiology and Biotechnology. Available at: <https://link.springer.com/journal/11274>

1. SUPPLEMENTARY TABLES

Table S1 Information regarding the taxonomic classification of the two assembled genomes

Isolate name	GTDB-Tk classification	ANI (%)	Closest genome reference	TYGS classification	dDDH (d0, in %)
BAC30	f__Bacillaceae;g__Bacillus;s__ <i>Bacillus paralicheniformis</i>	98.91	GCF_001042485.2	<i>Bacillus paralicheniformis</i> KJ-16	97,5
BAC220	f__Bacillaceae;g__Bacillus;s__ <i>Bacillus paralicheniformis</i>	98.8	GCF_001042485.2	<i>Bacillus paralicheniformis</i> KJ-16	92,7

Table S2 Assessment of sequencing and assembly quality, as well as genomic features of the isolates.

Isolate name	CheckM2 Completeness	CheckM2 Contamination	Genome size (pb)	% GC	Pass. GUNC	Barrnap 5S	Barrnap 16S	Barrnap 23S
BAC30	100.0	0.03	4471444	46	True	Complete	Partial	Partial
BAC220	100.0	3.06	4785386	45	True	Complete	Partial	Partial

2. SUPPLEMENTARY FIGURES

Fig. S1 ANI analysis of 30 *Bacillus paralicheniformis* genomes, performed using the pyANI software, with values ranging from orange to red, and the outgroup *Bacillus sonorensis* (ASM3405511v1) shown in blue.

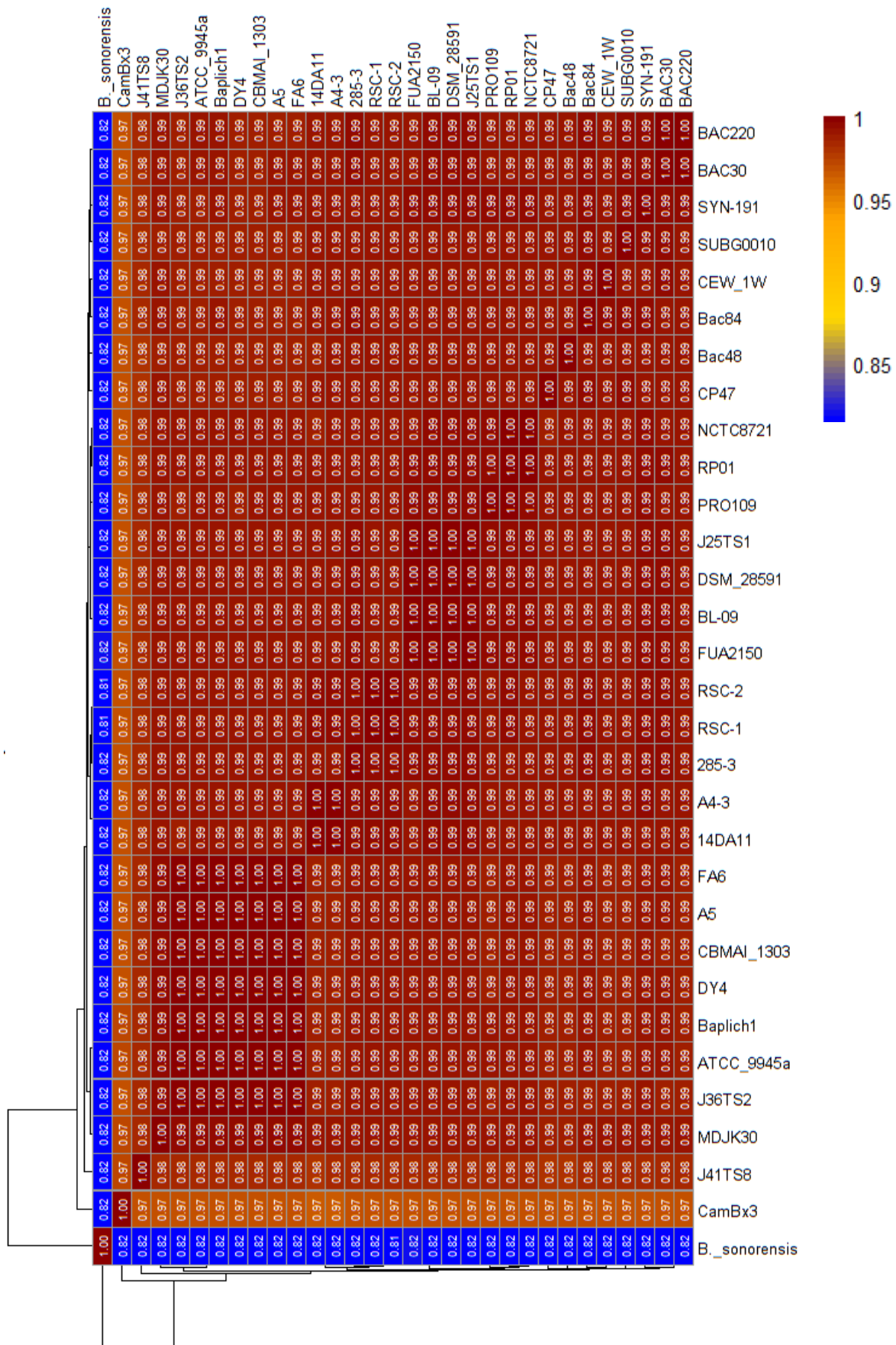


Fig. S2 Phylogenomic analysis of *B. paralicheniformis* strains using the Gegenees software, and phylogenetic tree construction with SplitsTree4 based on the Neighbor-Joining (NJ) method.

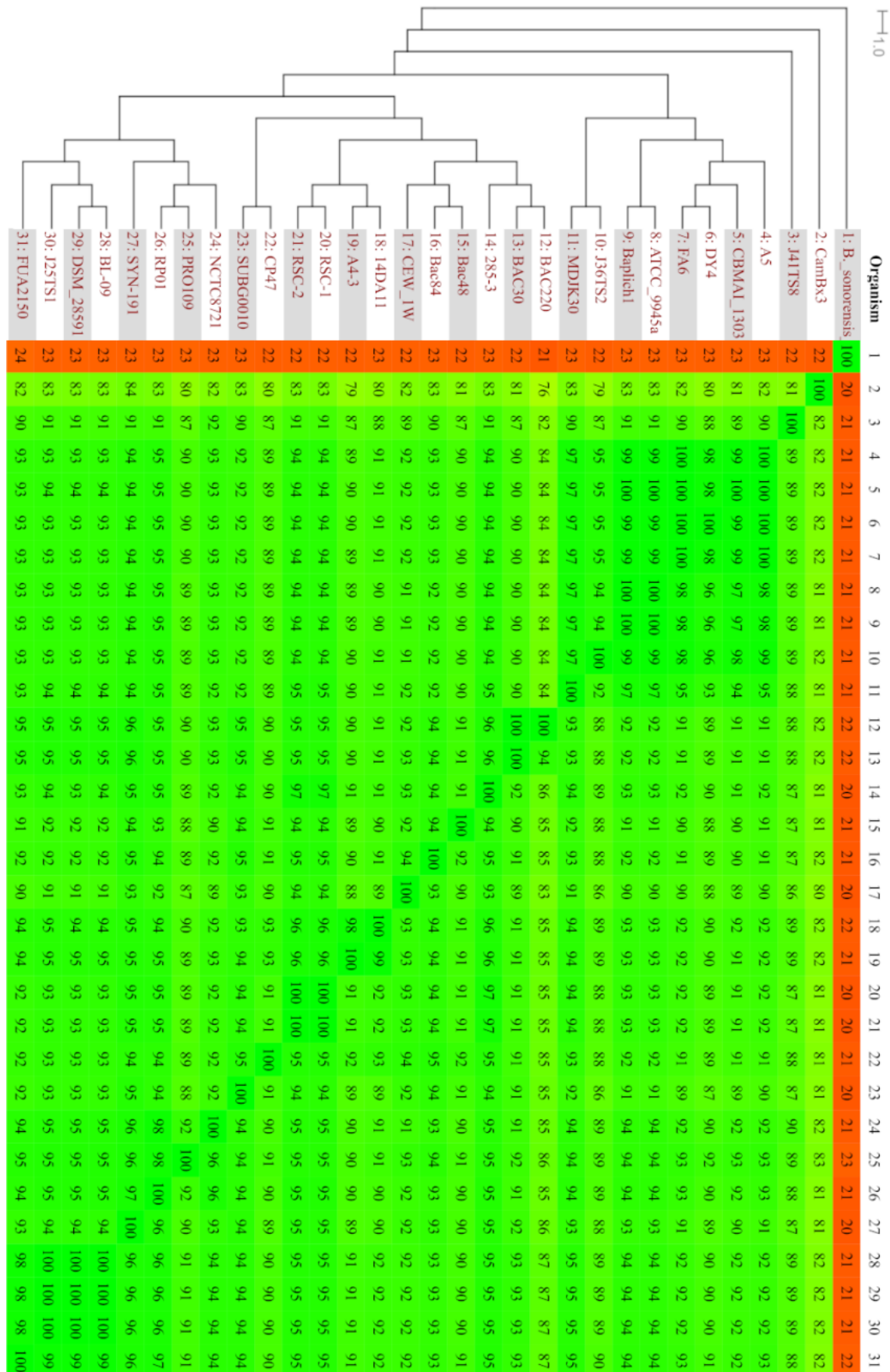


Fig. S3 Distribution of gene families within the pangenome, obtained using the PPanGGolin software. Core (Persistent) genes are shown in blue; Shell (shared) genes in orange; and Cloud (unique) genes in green.

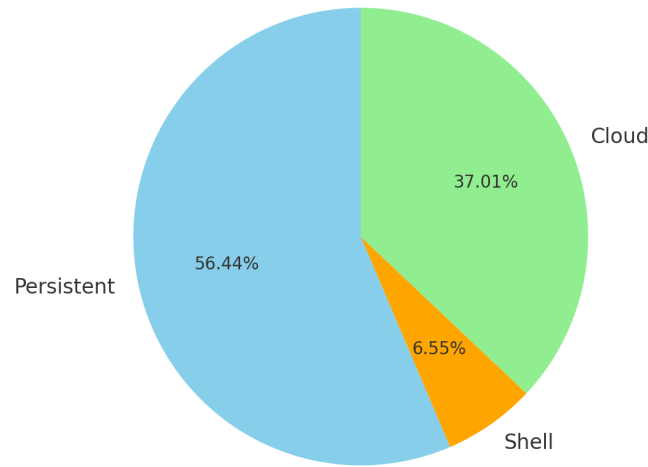


Fig. S4 Functional classification of *Bacillus paralicheniformis* species and the analysis of the isolates impact in the species genetic repertoire using the COG Classifier software **(a)** The core genome of the isolates BAC30 and BAC220 **(b)** Comparison of core genes from the 28 *Bacillus paralicheniformis* strains excluding the isolates; **(c)** Comparison of core genes from the 28 *Bacillus paralicheniformis* strains including the isolates.

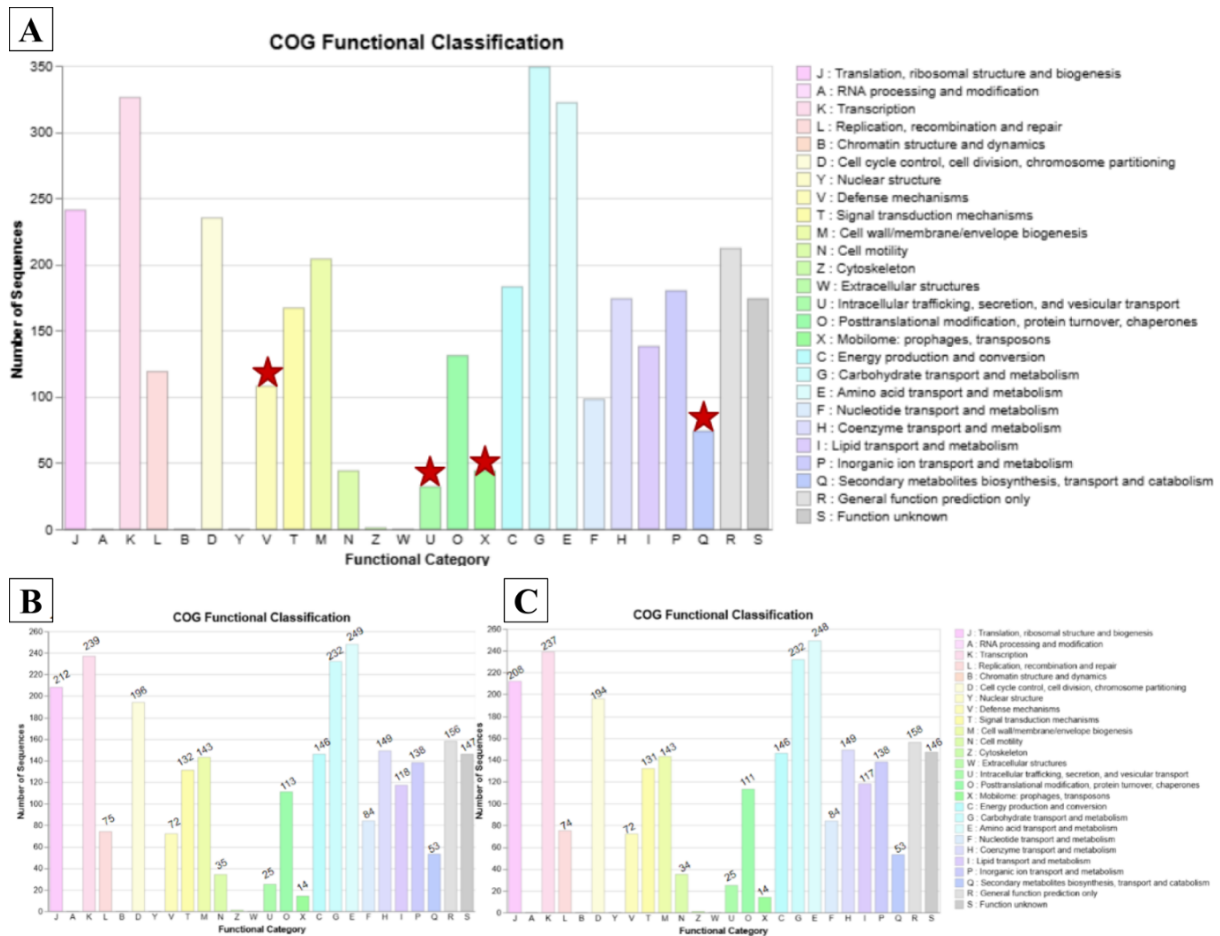


Fig. S5 Analysis of the isolates BAC30 and BAC220 impact on the gene repertoire of the species, based on alpha values observed in PPanGGolin **(a)** Dataset including the isolates **(b)** Dataset excluding the isolates

