

# A Glimpse into an Uncharted Viral Territory: Exploring Prophage-like Regions in a Brazilian *Paracidovorax citrulli* Isolate

Marcelo Henrique Oliveira Gonçalves<sup>1\*</sup>, Fábio Junior Nogueira Fernandes<sup>2</sup>, Elineide Barbosa de Souza<sup>3</sup>

<sup>1</sup> Department of Agronomy, Universidade Federal Rural de Pernambuco, Recife – PE, Brazil.

<sup>2</sup> Department of Microbiology, Universidade Federal de Viçosa, Viçosa – MG, Brazil.

<sup>3</sup> Department of Biology, Universidade Federal Rural de Pernambuco, Recife – PE, Brazil.

\* Electronic correspondence: marcelo.h.goncalves@ufrpe.br

The bacterium *Paracidovorax citrulli* (syn. *Acidovorax citrulli*) is the causal agent of bacterial fruit blotch in cucurbits, and there is limited knowledge regarding the diversity of viruses that infect it, with the few reported instances primarily involving lytic viruses with potential applications in biocontrol. Nonetheless, certain bacteriophages can integrate into the genome of their hosts, directly influencing their evolutionary trajectory. Recently, the genome of a Brazilian isolate obtained from “Melão-pepino” (*Cucumis melo* var. *cantalupensis*) in Rio Grande do Norte, designated as CCRMAcMP2, was sequenced. Automated annotation of this genome revealed the presence of three open reading frames (ORFs) that encode viral proteins within its genome. In this context, the objective of the study was to identify prophage-like regions in the genome of this isolate. Subsequently, the investigation of putative phage-related regions within the genome was undertaken utilizing the PHASTER platform. Employing Geneious software, these identified regions were meticulously extracted, and their respective ORFs were reidentified and annotated. A Megablast analysis was executed against a nucleotide database sourced from GenBank's RefSeq, and subsequently, sequences classified as intact by the PHASTER platform were subjected to VirFam for the precise identification of structural proteins. Four phage-like regions were discerned as being integrated within the genome of the bacterium, specifically within Contig01, Contig06, and Contig34. Among these, two were categorized as intact, while the remaining two were classified as incomplete. The sizes of these regions ranged from 14 Kb to 48.5 Kb, collectively representing approximately 2.32% of the bacterium's total genome. On average, these regions displayed a GC content of 66.2% ( $\pm 0.7$ ), which differed from the genome of the species. It was not possible to find any closely related sequences within the dataset containing 14,289 viral RefSeqs for three of the identified regions. Furthermore, although one of the regions, designated as CCRMAcMP2 prophage 01, exhibited 71% identity with the *Hollowayvirus H66* (NC\_042342), which infects *Pseudomonas*, there was only a 7% coverage with our sequence. Considering the typically host-specific characteristics of viruses and the limited availability of documented bacteriophages capable of infecting species within the *Paracidovorax* genus, it is highly probable that these identified regions represent novel lysogenic viruses that remain undescribed.

**Key words:** PHASTER; Bacterial fruit blotch; *Hollowayvirus*; *Acidovorax citrulli*.

## Um vislumbre de um território viral desconhecido: Explorando regiões semelhantes a profagos em um isolado brasileiro de *Paracidovorax citrulli*

A diversidade de vírus que infectam *Paracidovorax citrulli* é pouco conhecida. E tendo em vista o recente sequenciamento do isolado CCRMAcMP2 de *P. citrulli* e a anotação de ORFs que codificam sequências virais, o propósito deste estudo consistiu em investigar a existência de regiões análogas a profagos em seu genoma. Foi possível identificar quatro regiões que não exibem semelhança com as sequências de vírus de referência previamente depositadas no GenBank, o que sugere a possível presença de potenciais vírus lisogênicos até então não relatados.

**Palavras-chave:** PHASTER; Mancha aquosa; *Hollowayvirus*; *Acidovorax citrulli*.

Acknowledge: This work was developed having support from FACEPE - Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco and CNPq (Proc. 311365/2020-5).