

**XU** Congresso  
Fluminense  
de Iniciação  
Científica e Tecnológica

**28<sup>o</sup>**

Encontro de  
Iniciação  
Científica  
da UENF

**20<sup>o</sup>**

Circuito de  
Iniciação  
Científica do  
IFFluminense

**16<sup>a</sup>**

Jornada de  
Iniciação  
Científica  
da UFF



**U III** Congresso  
Fluminense de  
Pós-Graduação

**23<sup>a</sup>**

Mostra de  
Pós-Graduação  
da UENF

**8<sup>a</sup>**

Mostra de  
Pós-Graduação  
do IFFluminense

**8<sup>a</sup>**

Mostra de  
Pós-Graduação  
da UFF

## Análise da diversidade genômica de bactérias do gênero *Pantoea*

*Felipe Rimes Ferreira Casais, Francisnei Pedrosa-Silva, Thiago Motta Venancio*

*Pantoea* é um gênero composto por bactérias móveis, Gram-negativas e em forma de bastonete. São amplamente encontradas em diversos ambientes, como solo e água, podendo se associar com diferentes hospedeiros, incluindo plantas e humanos. Diferentes espécies deste gênero são conhecidas pela capacidade de promover o crescimento vegetal através de mecanismos envolvidos com biocontrole de fitopatógenos, bioestimulação e biorremediação de solos contaminados por metais tóxicos. Contudo, diversas espécies de *Pantoea* são consideradas como potenciais patógenos de plantas e humanos. Com avanço das tecnologias de sequenciamento, a quantidade de genomas disponíveis do gênero *Pantoea* aumentou significativamente, permitindo a melhor compreensão dos mecanismos fisiológicos e da diversidade existente entre as espécies através da genômica comparativa. No entanto, estimar a diversidade genômica das espécies de *Pantoea* requer uma avaliação criteriosa e sistemática dos genomas disponíveis nos bancos de dados. O objetivo deste trabalho foi analisar a diversidade genômica de bactérias do gênero *Pantoea* e identificar potenciais erros de classificação taxonômicas no gênero. Foram obtidos 774 genomas de *Pantoea* banco de dados do genbank. Excluímos genomas com fragmentação superior a 500 contigs, completude inferior a 90% e contaminação superior a 10% por meio da ferramenta CheckM, excluímos genomas fragmentados (com mais de 500 contigs), incompletos (completude inferior a 90%) e contaminados. Ao todo, 130 genomas foram excluídos após análises de qualidade. A deduplicação de genomas foi realizada com a ferramenta Mash e scripts *in house*. Para identificar possíveis espécies existentes no conjunto de dados, realizamos análises de identidade média de nucleotídeos (ANI) com a ferramenta Pyani. Construímos uma árvore de distância entre os genomas utilizando a ferramenta Mashtree. Identificamos 210 genomas com valor de mash < 0,005, sugerindo que sejam genomas idênticos e possivelmente duplicados. Nosso conjunto final contém 434 genomas de *Pantoea*. Através das análises de ANI e Mash identificamos 32 grupos que compõem espécies distintas. Além disso, identificamos 25 genomas com classificação taxonômica incorreta e 75 genomas sem espécie definida (*Pantoea* spp.). Nossos resultados contribuem para o entendimento da diversidade genômica de *Pantoea* e implicações a respeito de seu potencial biotecnológico e biossegurança.

**Palavras-chaves:** Biotecnologia, bioinformática, genômica comparativa.

*Instituição do Programa de IC, IT ou PG: UENF*

*Eixo temático: Biociências*

*Fomento da bolsa (quando aplicável): CNPq*

ORGANIZAÇÃO E REALIZAÇÃO:



APOIO:



**XU** Congresso  
Fluminense  
de Iniciação  
Científica e Tecnológica

**28<sup>o</sup>**  
Encontro de  
Iniciação  
Científica  
da UENF

**20<sup>o</sup>**  
Circuito de  
Iniciação  
Científica do  
IFFluminense

**16<sup>a</sup>**  
Jornada de  
Iniciação  
Científica  
da UFF



**UIII** Congresso  
Fluminense de  
Pós-Graduação

**23<sup>a</sup>**  
Mostra de  
Pós-Graduação  
da UENF

**8<sup>a</sup>**  
Mostra de  
Pós-Graduação  
do IFFluminense

**8<sup>a</sup>**  
Mostra de  
Pós-Graduação  
da UFF

## Analysis of the genomic diversity of bacteria of the genus *Pantoea*

*Felipe Rimes Ferreira Casais, Francisnei Pedrosa-Silva Thiago Motta Venancio*

*Pantoea* is a genus composed of motile, Gram-negative, rod-shaped bacteria. They are widely found in diverse environments, such as soil and water, and can associate with different hosts, including plants and humans. Different species of this genus are known for their ability to promote plant growth through mechanisms involved in biocontrol of phytopathogens, biostimulation, and bioremediation of soils contaminated by toxic metals. However, several *Pantoea* species are considered as potential pathogens of plants and humans. With advances in sequencing technologies, the amount of available genomes of the genus *Pantoea* has increased significantly, allowing for a better understanding of the physiological mechanisms and diversity among species through comparative genomics. However, estimating the genomic diversity of *Pantoea* species requires a careful and systematic evaluation of the genomes available in databases. The aim of this work was to analyze the genomic diversity of bacteria of the genus *Pantoea* and to identify potential taxonomic misclassifications in the genus. We obtained 774 genomes of *Pantoea* from the genbank database. We excluded genomes with fragmentation greater than 500 contigs, completeness less than 90% and contamination greater than 10%. Using the CheckM tool, we excluded fragmented (more than 500 contigs), incomplete (completeness less than 90%) and contaminated genomes. In all, 130 genomes were excluded after quality analyses. Genome deduplication was performed with the Mash tool and in house scripts. To identify possible species existing in the dataset, we performed average nucleotide identity (ANI) analyses with the Pyani tool. We built a distance tree between the genomes using the Mashtree tool. We identified 210 genomes with a mash value  $< 0.005$ , suggesting that they are identical and possibly duplicate genomes. Our final set contains 434 genomes of *Pantoea*. Through ANI and Mash analyses we identified 32 clusters that make up distinct species. In addition, we identified 25 genomes with incorrect taxonomic classification and 75 genomes with no defined species (*Pantoea* spp.). Our results contribute to the understanding of *Pantoea* genomic diversity and implications for its biotechnological potential and biosafety.

**Keywords:** Biotechnology, bioinformatics, comparative genomics.

ORGANIZAÇÃO E REALIZAÇÃO:



APOIO:

