

Identificação de erros na atribuição taxonômica de sequências genômicas de *Paenibacillus spp.* depositadas no Genbank

Renan Zanini Porto^{1,3}, Luciane M. P. Passaglia^{2,3}

¹Graduando em Ciências Biológicas, Universidade Federal do Rio Grande do Sul – UFRGS

²Orientador, Universidade Federal do Rio Grande do Sul – UFRGS

³Núcleo de Microbiologia Agrícola, Departamento de Genética, Instituto de Biociências- IB, UFRGS

Contato: renanzaniniporto@hotmail.com

INTRODUÇÃO

A identificação de isolados de *Paenibacillus spp.* em nível de espécie é fundamental para direcionar seu apropriado estudo e utilização. O gênero se caracteriza como um grupo parafilético e extremamente diverso, com bactérias isoladas de locais variados como solo, água e rizosfera. Apresentam um grande potencial biotecnológico, podendo ser utilizadas na promoção do crescimento vegetal, na proteção contra ataques de fitopatógenos, na produção de antibióticos e no controle de insetos e pragas. O gene do rRNA 16S é um marcador taxonômico que não apresenta resolução suficiente para discriminar as linhagens de *Paenibacillus* em nível de espécie. Com o advento das tecnologias de sequenciamento de alto rendimento, surgiu a possibilidade de identificar e classificar organismos por meio de métricas genômicas, como a identidade nucleotídica média (ANI – *Average Nucleotide Identity*).

OBJETIVOS

Este estudo objetiva validar as atribuições taxonômicas de genomas de *Paenibacillus spp.* depositados no GenBank, através de métricas genômicas.

METODOLOGIA

Obtenção de todos os 439 genomas de *Paenibacillus* no Genbank

Checkagem de completude com CheckM

Incongruências baseadas em 129 linhagens tipo

ANI (PYANI) All vs All

ANI	Nome igual ao da linhagem tipo	
	SIM	NÃO
> = 95%	V.P	F.N
< 95%	F.P	V.N

RESULTADOS

Cerca de quatro espécies apresentaram uma taxa de falso-negativos e/ou taxa de precisão relevantes. A espécie *Paenibacillus polymyxa*, por exemplo, apresentou precisão de 31% e taxa de falsos-negativos de 33%, na atribuição taxonômica de suas linhagens. A espécie *Paenibacillus pasadenensis* também apresentou taxas de precisão muito baixas, visto que apenas uma de suas linhagens obteve identidade suficiente com a linhagem-tipo. A atribuição à espécie *P. peoriae* apresentou baixa precisão, muitas linhagens apresentaram identidade maior que o limiar de circunscrição específico com a linhagem-tipo da espécie *P. polymyxa*, sugerindo um grupo diferente de genoespécies a ser investigado.

RESULTADOS

Figura 1: Gráfico das taxas de precisão, falso-positivos e falsos negativos das espécies que apresentaram as maiores incongruências de atribuições taxonômicas.

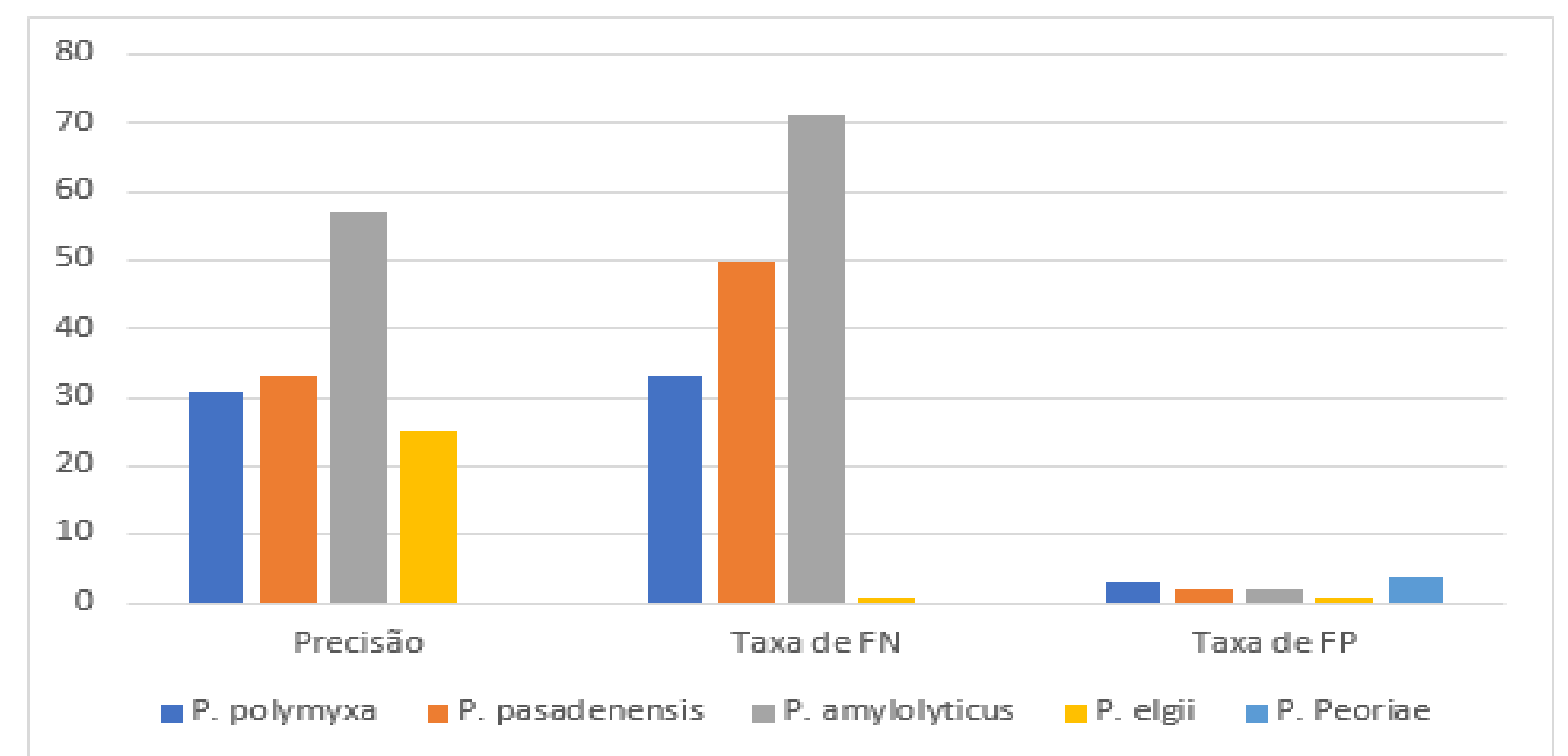


Tabela 1: Grupos de genoespécies formados por proximidade de identidade nucleotídica média, linhagens outrora mal classificadas.

Classificação	Linhagem <i>Polymyxa</i>	Linhagem <i>Peoriae</i>	ANI
GENOESPÉCIES	<i>Paenibacillus polymyxa</i> CCI-25	<i>Paenibacillus peoriae</i> FSL A5-0030	~96-98%
	<i>Paenibacillus polymyxa</i> WLY78		
	<i>Paenibacillus polymyxa</i> DSM 365		
	<i>Paenibacillus polymyxa</i> J		
	<i>Paenibacillus polymyxa</i> CFSAN 34343		
	<i>Paenibacillus polymyxa</i> CFSAN 34341		
	<i>Paenibacillus polymyxa</i> CFSAN 34342		
	<i>Paenibacillus polymyxa</i> ATCC 15970		
	<i>Paenibacillus polymyxa</i> CR1		
	<i>Paenibacillus peoriae</i> FSL R7-0131		
	<i>Paenibacillus peoriae</i> FSL J3-0120		
	<i>Paenibacillus peoriae</i> FSL H8-0551		

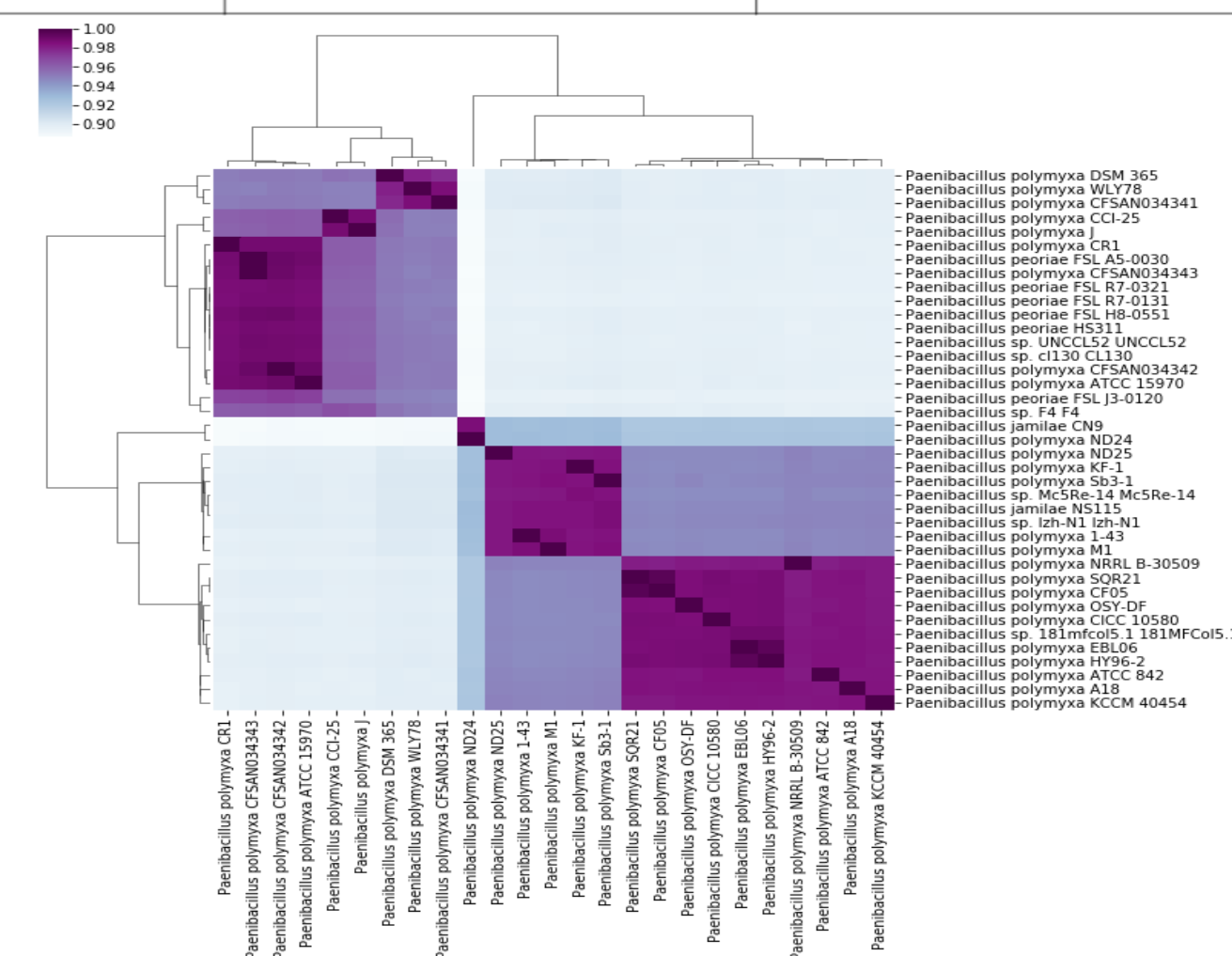


Figura 2: Heatmap representando os clusters formados pela correlação das genoespécies classificadas equivocadamente como pertencentes a *P. polymyxa* e *P. peoriae*.

CONCLUSÃO

Muitos erros de atribuição foram encontrados no GenBank, portanto recomenda-se o uso de métricas genômicas para a correta classificação antes de se submeter sequências genômicas a este banco de dados.

PERSPECTIVAS

Validar os resultados obtidos pelas métricas utilizadas através de métodos filogenéticos.

REFERÊNCIAS

- Bach, E. *et al.* Detection of misidentifications of species from the *Burkholderia cepacia* complex and description of a new member, the soil bacterium *Burkholderia catarinensis* sp. nov. *Pathogens and Disease*, volume 75. (2017)
- Varghese, N. J. *et al.* Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* (2015).
- Goris, J. *et al.* DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* (2007).

