



Identificação de incongruências na atribuição taxonômica de *Paenibacillus spp.* depositadas no Genbank

Renan Zanini Porto^{1 3} Fernando Hayashi Sant'Anna^{2 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: renanzaninporto@hotmail.com

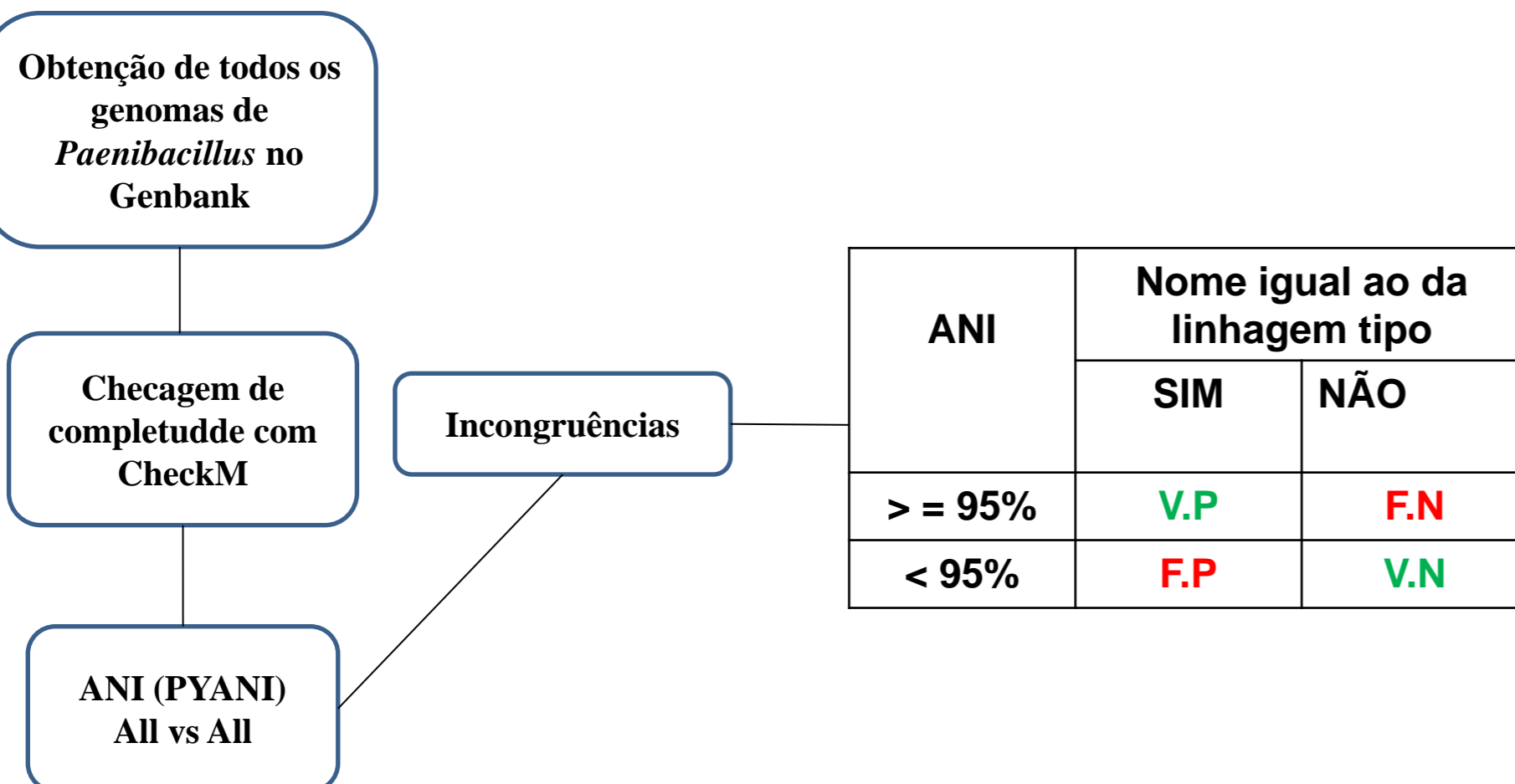
INTRODUÇÃO

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. Os marcadores taxonômicos comumente utilizados, como o rRNA 16S, em muitos casos, não discriminam isolados desse gênero em nível de espécie ocasionando, assim, classificações equivocadas que não expressam corretamente as relações filogenéticas entre os indivíduos.

OBJETIVOS

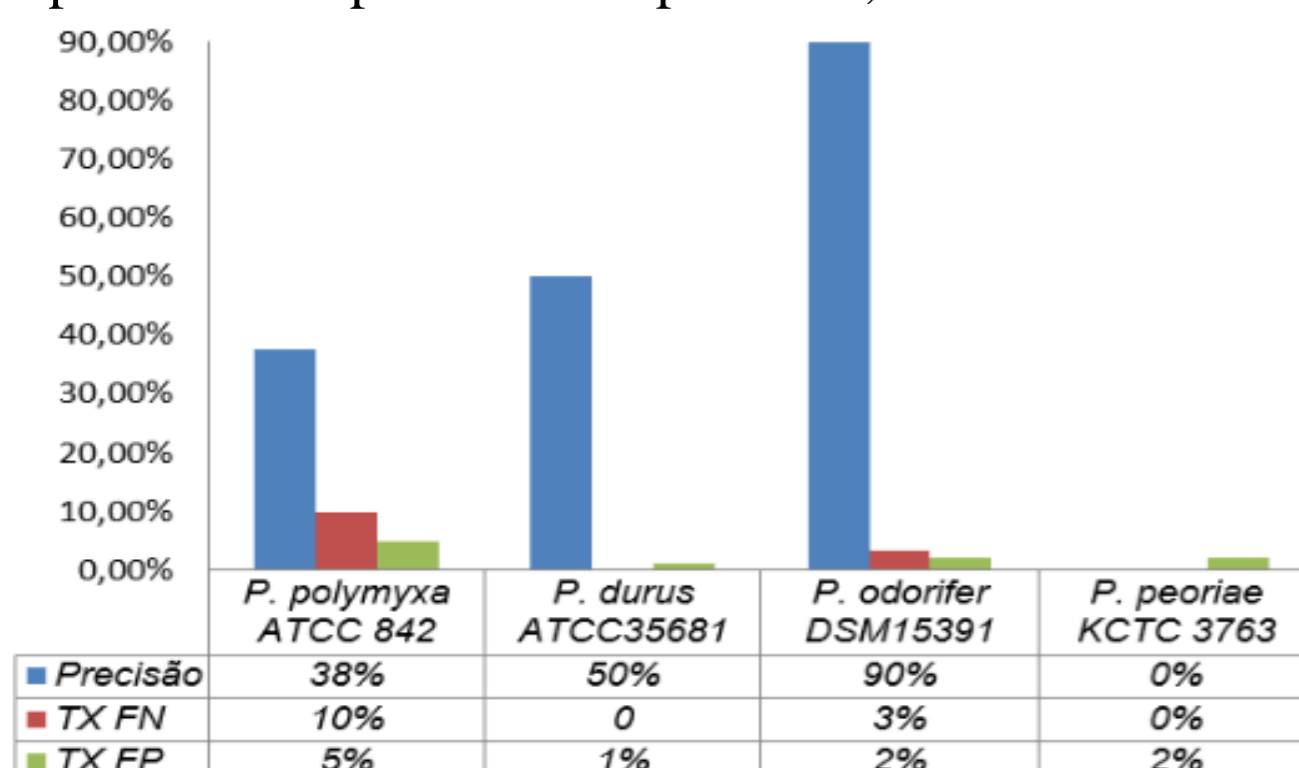
Este estudo objetiva validar as atribuições taxonômicas de genomas de *Paenibacillus spp.* depositados no GenBank, através de métricas genômicas.

METODOLOGIA



RESULTADOS

Foram identificadas incongruências na classificação em 30 (16%) dos genomas analisados, ao cruzar bidirecionalmente cada linhagem-tipo entre si e contra todos os outros genomas. Algumas espécies, como *P. polymyxa*, por exemplo, apresentaram precisão de apenas 37,5% e taxa de falso-negativos de 10%.



Foram identificados problemas tanto de identificação (Tabela 1), quanto de classificação taxonômica.

Foram identificados problemas tanto de identificação (Tabela 1), quanto de classificação taxonômica.

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

RESULTADOS

Tabela 1: Problemas de classificação: Linhagem-tipo vs. linhagem-tipo; Problemas de identificação: Linhagem qualquer não-tipo vs. linhagem-tipo.

Classificação	Linhagem-tipo	Linhagem	ANI
FALSO POSITIVO	<i>Paenibacillus alvei</i> DSM-29 ^T	<i>Paenibacillus alvei</i> TS-15	84%
		<i>Paenibacillus alvei</i> A6-6i	84%
	<i>Paenibacillus odorifer</i> DSM 15391 ^T	<i>Paenibacillus odorifer</i> FSL H7-0710	85%
		<i>Paenibacillus odorifer</i> FSL H7-0443	86%
		<i>Paenibacillus odorifer</i> FSL R5-0923	87%
		<i>Paenibacillus peoriae</i> FSL A5-0030	86%
		<i>Paenibacillus peoriae</i> FSL R7-0321	87%
		<i>Paenibacillus peoriae</i> FSL R7-0131	86%
		<i>Paenibacillus peoriae</i> FSL J3-0120	87%
	<i>Paenibacillus peoriae</i> KCTC 3763 ^T	<i>Paenibacillus peoriae</i> FSL H8-0551	86%
<i>Paenibacillus peoriae</i> HS311		92%	
FALSO NEGATIVO	<i>Paenibacillus durus</i> DSM 1735 ^T	<i>Paenibacillus durus</i> ATCC 35681 (azotofixans ¹)	97%
		<i>Paenibacillus dauci</i> H9 ^T	95%
	<i>Paenibacillus massiliensis</i> 230106 ^T	<i>Paenibacillus panacisoli</i> DSM 21345	98%
		<i>Paenibacillus polymyxa</i> ATCC 842 ^T	<i>Paenibacillus</i> sp. 181mfcol5.1 181MFcol5.1
	<i>Paenibacillus alucanolyticus</i> 5162 ^T	<i>Paenibacillus</i> sp. FSL R5-808 FSL R5-808	99%

Tabela 2: Grupos de genospecies formados por proximidade de identidade nucleotídica média, linhagens outrora mal classificadas.

Classificação	Linhagem <i>Polymyxa</i>	Linhagem <i>Peoriae</i>	ANI
GENOSPECIES	<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 polymyxa</i> CR1		

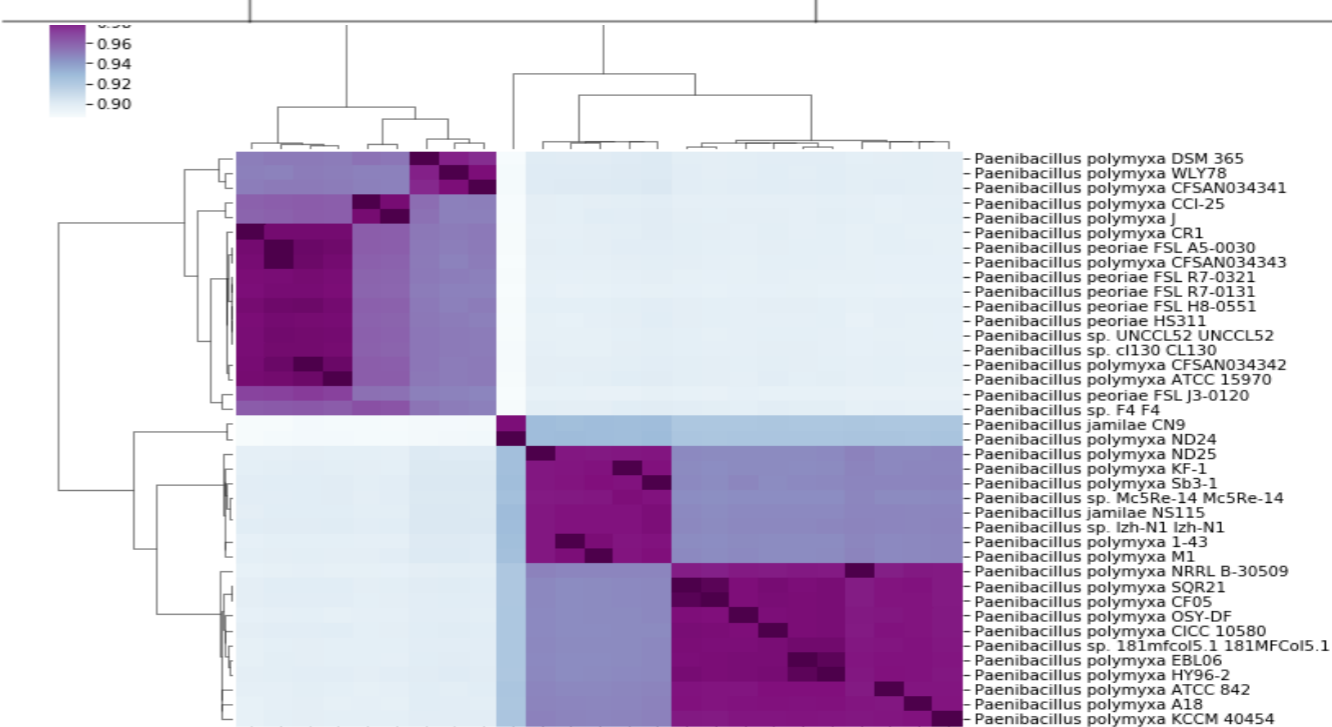


Figura 2: Heatmap representando os clusters formados pela correlação das genospecies classificadas equivocadamente como pertencentes a *P. polymyxa* e *P. peoriae*.

CONCLUSÃO

Conclui-se que existem muitos erros no GenBank que poderiam ser evitados se a análise de métricas genômicas de classificação fossem compulsórias antes da submissão de sequências.

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).