

ANÁLISE PANGENÔMICA COMO ESTRATÉGIA DE PROSPECÇÃO DE ALVOS VACINAIS PARA ENTERITE NECRÓTICA

JOÃO PEDRO GOMES GRECO ¹; FABRÍCIO ROCHEDO CONCEIÇÃO ²;
FREDERICO SCHMITT KREMER ³; LUCIANO DA SILVA PINTO ⁴.

¹Laboratório de Bioinformática e Proteômica, Graduação em Biotecnologia, CDTec, UFPel –
joao.greco@ufpel.edu.br

²Laboratório de Imunologia aplicada, graduação em biotecnologia, CDTEec, UFPel –
fabricio.rochedo@ufpel.edu.br

³Laboratório de Bioinformática, Graduação em Biotecnologia, CDTec, UFPel –
fred.s.kremer@gmail.com

⁴Laboratório de Bioinformática e Proteômica, Graduação em Biotecnologia, CDTec, UFPel –
ls_pinto@hotmail.com

1. INTRODUÇÃO

Várias doenças de importância veterinária são causadas por bactérias. Em aves, uma dessas doenças é causada pela bactéria *Clostridium perfringens*, uma bactéria gram-positiva causadora da doença enterite necrótica (ENA) (CASAGRANDE, 2012). Com o aumento das ocorrências de ENA em frangos de corte jovens, essa doença possui um grande impacto na produção avícola, resultando em grandes perdas aumentando a taxa de mortalidade, e o constante crescimento de animais saudáveis afetados.

O genoma deste organismo já foi desvendado e várias sequências estão disponíveis. Apesar de serem da mesma espécie e infectarem o mesmo hospedeiro, estas bactérias podem usar diferentes estratégias para causarem a doença e assim podem apresentar fatores de virulência e alvos vacinais diferentes. Para garantir uma abrangência maior na busca, é possível usar uma abordagem de vacinologia reversa (VR) para a prospecção de novos alvos. A VR busca analisar o genoma na tentativa de encontrar possíveis alvos vacinais, utilizando de diversos tipos de predição para isso. Porém, faz-se necessário determinar o grau de conservação entre as espécies e entre os genes selecionados. Um meio de realizar isto é através de análises pangenômicas, descrita primeiramente por Tettelin e colaboradores em 2005 (TETTELIN et al., 2005). O termo pangenoma tem uma definição clara de ser o conjunto total do repertório de genes de um grupo estudado e após a definição destes genes são aplicadas ferramentas de prospecção *in silico* para a busca dos alvos.

Com base no exposto, o objetivo do trabalho foi prospectar alvos vacinais para a doença ENA causada por *Clostridium perfringens* usando ferramentas de Bioinformática.

2. METODOLOGIA

O trabalho realizado consistiu na reconstrução do pangenoma e na realização de análises bioinformáticas para investigar a composição genética, as características das proteínas e a localização subcelular das proteínas presentes nos genomas selecionados, com o foco na identificação de possíveis proteínas de membrana associado à enterite necrótica. O estudo começou com a coleta dos genomas dos conjuntos de dados do NCBI no formato RefSeq, fasta e das anotações no formato GBFF. Estes foram utilizados para a montagem do pangenoma por meio da ferramenta Roary (PAGE et al., 2015), visando à análise dos genes ortólogos e à identificação dos genes comuns. As análises de bioinformática foram iniciadas

utilizando sequências de aminoácidos traduzidas a partir das sequências de nucleotídeos dos genomas, a fim de possibilitar a análise das características das proteínas.

A etapa inicial das ferramentas de filtragem foi iniciada com o uso do SignalP (TEUFEL et al., 2022), uma ferramenta responsável pela identificação dos peptídeos sinal. Em seguida, utilizou-se o PsortB (YU et al., 2010), dedicado a prever a localização subcelular, e o DeepTMHMM (KROGH et al., 2001), destinado à detecção das proteínas na região transmembrana. Após essas análises iniciais procedeu-se a uma seleção das sequências de aminoácidos que apresentaram uma probabilidade de serem de membrana, considerando a presença dos peptídeos sinais e a localização subcelular, ambos com índices superiores a 90%.

Aplicando a ferramenta Scoary (BRYNILDSRUD et al., 2016), os genomas associados à enterite necrótica foram comparados com o pangenoma, tal comparação visou analisar o genoma acessório destas cepas em relação ao genoma núcleo. Ao final dessas análises as proteínas identificadas foram submetidas para a previsão do potencial antigênico com as proteínas usando o Vaxijen (DOYTCHINOVA; FLOWER, 2007).

3. RESULTADOS E DISCUSSÃO

Após a utilização dos 604 genomas disponíveis para a construção do pangenoma, a montagem indicou a presença de 30.917 genes (Tabela 1). Porém, destes genomas apenas 16 deles continham a doença caracterizada nas suas anotações.

Tabela 1: Resultados da construção do pangenoma.

Pan-genoma	Valores de genomas obtidos
Genoma núcleo	1.566
Genoma acessório	29.351
Total de genes	30.917

O genoma núcleo quando submetido as análises de bioinformática, resultou em um total de 1566 peptídeos sinais identificados, 699 proteínas com hélice transmembrana e 378 proteínas demonstraram-se serem citoplasmáticas (Tabela 2).

Tabela 2: Resultados obtidos com as análises do genoma núcleo.

Pangenoma	Número de proteínas encontradas
Peptídeo sinal	1.566
Hélice transmembrana	699
Citoplasmática	378

As proteínas anotadas destes genes foram submetidas às diferentes abordagens para busca de alvos vacinais (Peptídeo sinal, proteínas de membrana, epítipo de ligação às células T e os epítipos das células B). Por fim, os epítipos resultantes destas abordagens foram analisados quanto a sua antigenicidade. Após essa triagem final foram encontradas 24 proteínas, mas apenas 11 se mostraram antigênicas (Tabela 3).

Tabela 3: Anotação das proteínas e antigenicidade prevista usando o software Vaxijen 2.0.

Anotação	Funções	Potencial antigênico
A	Proteína de síntese	0.4488
B	Proteína de esporulação	0.4596
C	Enzima metalo-dependente protease	0.4631
D	Ligase	0.4997
E	Transferase	0.4094
F	Proteína de esporulação IV	0.3349
G	Polimerase	0.5243
H	Fator de transcrição	0.5425
I	Quinase	0.6360
J	Quinase	0.6428
K	Fator de elongação	0.6327

4. CONCLUSÕES

Essas proteínas terão um papel fundamental na busca por novos antígenos, promovendo o desenvolvimento de medicamentos e vacinas para enfrentar as enfermidades que afetam o setor agropecuário.

5. REFERÊNCIAS BIBLIOGRÁFICAS

CASAGRANDE, Mariana Fröner. **Clostridium perfringens em ingredientes para ração de aves e controle da presença do agente utilizando tratamento químico**. 2012. 38 f. Dissertação (Mestrado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2012.

Tettelin H, Massignani V, Cieslewicz MJ, et al. **Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome."** Proc Natl Acad Sci U S A. 2005;102(39):13950-13955. doi:10.1073/pnas.0506758102. (Erratum in: Proc Natl Acad Sci U S A. 2005 Nov 8;102(45):16530.)

PAGE, A. J., CUMMINS, C. A., HUNT, M., WONG, V. K., REUTER, S., HOLDEN, M. T. G., FOOKES, M., FALUSH, D., KEANE, J. A., PARKHILL, J. **Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics**, v. 31, n. 22, p. 3691-3693, novembro de 2015. DOI: 10.1093/bioinformatics/btv421.

TEUFEL, F.; ALMAGRO ARMENTEROS, J. J.; JOHANSEN, A. R.; et al. **SignalP 6.0 predicts all five types of signal peptides using protein language models.** Nature Biotechnology, v. 40, n. 7, p. 1023-1025, 2022. DOI: 10.1038/s41587-021-01156-3.

YU, N. Y.; WAGNER, J. R.; LAIRD, M. R.; et al. **PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories**

and predictive capabilities for all prokaryotes. *Bioinformatics*, v. 26, n. 13, p. 1608-1615, 2010. DOI: 10.1093/bioinformatics/btq249.

KROGH, A.; LARSSON, B.; VON HEIJNE, G.; SONNHAMMER, E. L. **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, v. 305, n. 3, p. 567-580, 2001. DOI: 10.1006/jmbi.2000.4315.**

BRYNILDSRUD, O., BOHLIN, J., SCHEFFER, L., et al. **Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biology*, v. 17, p. 238, 2016. DOI: 10.1186/s13059-016-1108-8.**

DOYTCHINOVA, I. A.; FLOWER, D. R. **VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, v. 8, p. 4, 2007. Disponível em: <https://doi.org/10.1186/1471-2105-8-4>.**