

PROSPECÇÃO E VALIDAÇÃO *in silico* DE SONDAS DE DNA PARA A DETECÇÃO DE *Escherichia coli* E *Staphylococcus aureus*

Gerardo Manuel Delgado-Calizaya¹ (Unit); Wescley Alexandre Rabêlo Leal¹ (Unit); Esther Santos Santana² (Doutoranda/PBI); Katlin Ivon Barrios-Eguiluz³ (Coorientadora); Maria Lucila Hernández-Macedo³ (Orientador)
gerardo.manuel@souunit.com.br

¹Universidade Tiradentes/Biomedicina/Aracaju/SE.

²Universidade Tiradentes/ Pós-graduação em Biotecnologia Industrial/Aracaju/SE

³Universidade Tiradentes/ Instituto de Tecnologia e Pesquisa/Aracaju/SE.

2.00.00.00-6 - Ciências Biológicas; 2.12.00.00-9 – Microbiologia

RESUMO

Introdução: A detecção rápida e específica de patógenos bacterianos é uma necessidade constante para a saúde pública e a segurança alimentar¹. O uso de sondas de DNA específicas permite diagnósticos sensíveis por meio de métodos moleculares e em biossensores². Assim, a prospecção de sondas direcionadas a *Staphylococcus aureus* e *Escherichia coli* é fundamental, uma vez que esses microrganismos estão entre os principais agentes de contaminação alimentar e possuem baixa dose infecciosa, causando infecções mesmo em baixas concentrações^{3,4}. **Objetivos:** Identificar, com base em evidências da literatura, regiões genômicas conservadas e específicas de *E. coli* e *S. aureus*, e avaliar sua especificidade e abrangência por meio de análises de homologia e inferência filogenética. **Metodologia:** Foi realizado um levantamento bibliográfico para a seleção de genes-alvo comumente utilizados na detecção de *E. coli* e *S. aureus*, seguido da extração de sequências representativas disponíveis nas bases GenBank e RefSeq. Consideraram-se sondas entre 18 e 30 nucleotídeos, obtidas de artigos publicados em periódicos com fator de impacto igual ou superior a três. As sondas foram avaliadas por BLASTn, utilizando o algoritmo Megablast, contra os genomas de *S. aureus* ATCC 25923 (CP009361.1) e *E. coli* ATCC 25922 (CP009072), a fim de verificar a identidade, a cobertura e a especificidade das sequências, bem como possíveis reações cruzadas. A especificidade foi validada por meio de árvores filogenéticas das sequências-alvo, construídas com o software MEGA X. O alinhamento múltiplo das sequências foi realizado com o algoritmo MUSCLE, e método de agrupamento UPGMA. A inferência filogenética foi conduzida pelo método de “Maximum Likelihood”, com 1000 réplicas de bootstrap. **Resultados:** Foram identificadas 13 sondas de DNA, das quais 11 apresentaram *E-value* superior a 1e-5 e homologia com outras espécies, indicando baixa especificidade. Apenas duas sondas mostraram alta compatibilidade e cobertura completa nas cepas ATCC analisadas, direcionadas aos genes purL de *E. coli* e Sa442 de *S. aureus*. As análises entre diferentes cepas de *E. coli* dos últimos cinco anos revelaram identidade total na sequência do gene purL, confirmando sua conservação dentro da espécie. De modo semelhante, a sequência Sa442 apresentou identidade total em *S. aureus*, reforçando sua especificidade de espécie⁵. Por outro lado, as sequências do gene purL mostraram baixa identidade entre gêneros, reforçando sua utilidade como marcador molecular específico em protocolos de identificação de patógenos. **Conclusão:** A abordagem adotada foi eficaz para identificar sondas de DNA com potencial de aplicação em métodos de detecção específicos para *E. coli* e *S. aureus*. As sondas purL e Sa442 apresentaram desempenho promissor nas análises *in silico*, configurando-se como candidatas relevantes para validação experimental. Além disso, as regiões-alvo mantiveram-se conservadas dentro das espécies, mas divergentes entre gêneros, confirmando sua especificidade e potencial como marcadores moleculares.

PALAVRAS-CHAVE: Patógenos, Sondas de DNA, Validação *in silico*.

Agradecimentos: Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). À Universidade Tiradentes, ao Instituto de Tecnologia e Pesquisa, ao Laboratório de Biologia Molecular. À Universidad Católica de Santa María (Peru).

ABSTRACT

Introduction: The rapid and specific detection of bacterial pathogens is a constant need for public health and food safety¹. The use of specific DNA probes enables sensitive diagnostics through molecular methods and biosensors². Thus, the prospecting of probes targeting *Staphylococcus aureus* and *Escherichia coli* is essential, since these microorganisms are among the main agents of food contamination and have a low infectious dose, causing infections even at low concentrations^{3,4}. **Objectives:** To identify, based on evidence from the literature, conserved and species-specific genomic regions of *E. coli* and *S. aureus*, and to evaluate their specificity and coverage through homology analyses and phylogenetic inference. **Methodology:** A bibliographic review was conducted to select target genes commonly used in the detection of *E. coli* and *S. aureus*, followed by the extraction of representative sequences available in the GenBank and RefSeq databases. Probes ranging from 18 to 30 nucleotides were considered, obtained from articles published in journals with an impact factor equal to or greater than three. The probes were evaluated using BLASTn with the Megablast algorithm against the genomes of *S. aureus* ATCC 25923 (CP009361.1) and *E. coli* ATCC 25922 (CP009072), in order to verify sequence identity, coverage, and specificity, as well as possible cross-reactions. Specificity was validated through phylogenetic trees of the target sequences constructed using MEGA X software. Multiple sequence alignment was performed with the MUSCLE algorithm and clustered using the UPGMA method. Phylogenetic inference was conducted using the Maximum Likelihood method with 1000 bootstrap replicates. **Results:** Thirteen DNA probes were identified, of which eleven exhibited an E-value higher than $1e-5$ and homology with other species, indicating low specificity. Only two probes showed high compatibility and complete coverage in the analyzed ATCC strains, targeting the *purL* gene of *E. coli* and the *Sa442* gene of *S. aureus*. Analyses among different *E. coli* strains from the last five years revealed total sequence identity of the *purL* gene, confirming its conservation within the species. Similarly, the *Sa442* sequence showed complete identity in *S. aureus*, reinforcing its species specificity⁵. On the other hand, *purL* gene sequences exhibited low identity between genera, reinforcing their usefulness as specific molecular markers in pathogen identification protocols. **Conclusion:** The adopted approach proved effective in identifying DNA probes with potential application in specific detection methods for *E. coli* and *S. aureus*. The *purL* and *Sa442* probes demonstrated promising performance in *in silico* analyses, making them relevant candidates for experimental validation. Furthermore, the target regions remained conserved within species but divergent between genera, confirming their specificity and potential as molecular markers.

KEYWORDS: DNA Probes, *In Silico* Validation, Pathogens.

ACKNOWLEDGEMENTS: The Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Universidade Tiradentes, Instituto de Tecnologia e Pesquisa, Laboratório de Biologia Molecular and the Universidad Católica de Santa María (Peru).

REFERÊNCIAS/REFERENCES:

1. PANWAR, S. et al. Advanced diagnostic methods for identification of bacterial foodborne pathogens: contemporary and upcoming challenges. *Critical Reviews in Biotechnology*, 43: 982–1000, 2023.
2. AVIVI, M. Y. et al. Nucleic acid hybridization-based detection of pathogenic RNA using microscale thermophoresis. *Journal of Biological Chemistry*, 300: 3, 2024.
3. AHMAD-MANSOUR, N. et al. *Staphylococcus aureus* Toxins: An Update on Their Pathogenic Properties and Potential Treatments. *Toxins*, 13: 677, 2021.
4. PAKBIN, B. et al. A quantitative prevalence of *Escherichia coli* O157 in different food samples using real-time qPCR method. *Food Science & Nutrition*, v. 11: 228, 2022.
5. HEILMANN, F. et al. Identification of 2,600 Clinical Methicillin-Resistant *Staphylococcus aureus* Strains in The Netherlands Yielded Sporadic Cases of Strains Negative for the Species-Specific *Sa442* Gene Fragment. *Journal of Clinical Microbiology*, 42: 2350, 2004.