

ANÁLISE ESTATÍSTICA ATRAVÉS DE BOXPLOT DE AMOSTRAS DE PCR

Lucca Muniz Coelho¹; Randson Davi Mendes Santos²; Bruno Corrêa Arvan³; Valmara Silveira Ponte⁴; Vinicius Pinto Rocha⁵; Milena Botelho Pereira Soares⁶; Valéria Loureiro da Silva⁷

¹ Graduando em Engenharia Elétrica; Iniciação Científica - FAPESB; lmunizcoelho@gmail.com

² Graduado em Física; Bolsista - FIOCRUZ; davissantoi@gmail.com

³ Graduando em Engenharia Elétrica; Bolsista – FIOCRUZ; arvanbruno@gmail.com

⁴ Mestre em Mecatrônica; Centro Universitário SENAI CIMATEC; Salvador-BA; valmara.ponte@fieb.org.br

⁵ Doutor em Patologia Humana; FIOCRUZ; Salvador-BA; vinicius.rocha@fieb.org.br

⁶ Doutora em Ciências Biológicas; FIOCRUZ; Salvador-BA; milena.soares@fieb.org.br

⁷ Doutora em Física; Centro Universitário SENAI CIMATEC; Salvador-BA; valeria.dasilva@fieb.org.br

RESUMO

A aplicação da fluorescência no ramo de detecção para área de saúde, é muito presente em diversos processos. Na pandemia atual, essa aplicação se faz presente na detecção de fluorescência do teste PCR para diagnóstico SARS-COV-2. Está sendo desenvolvido um fluorímetro portátil para leitura de testes usando a técnica RT-RPA, buscando viabilizar uma alternativa mais rápida e simples para os testes de RT-qPCR. Em testes de bancada buscou-se a otimização dos parâmetros de excitação e detecção do fluorímetro, utilizando quatro grupos de amostras previamente caracterizadas num equipamento de RT-qPCR. Os dados obtidos foram analisados através de gráficos Box Plots, que permitiram identificar os limites da distribuição e amostras do tipo outliers.

PALAVRAS-CHAVE: *Box Plot*; PCR; RPA; fluorímetro.

1. INTRODUÇÃO

Na pandemia de SARS-COV-2, o uso de fluorescência para identificação da presença do vírus através da reação RT-PCR (*Reverse Transcriptase - Polymerase Chain Reaction* - Reação em Cadeia da Polimerase)¹, ganhou destaque e notoriedade. O PCR em tempo real é um processo de amplificação do material genético que quando presente, reage com um fluoróforo adicionado à amostra para desempenhar a função sonda.¹ Através de um fluorímetro, equipamento com uma fonte de excitação e um sensor capaz de detectar a luz fluorescente, é possível determinar e quantificar a presença do vírus numa amostra.

A problemática desse tipo de análise está no equipamento, que requer uma infraestrutura laboratorial com pessoal treinado e possui custo alto. Além disso, utiliza um termociclador para alternar os passos de hibridização e amplificação durante o processo de PCR¹, e conseqüentemente demanda maior tempo para obtenção dos resultados.

Buscando solucionar esse problema foi proposto o desenvolvimento da técnica PCR Isotérmica ou RT-RPA (*Reverse Transcriptase - Recombinase polymerase amplification*), processo que ocorre numa temperatura constante e de maneira mais rápida², associada à de um fluorímetro portátil. Isso proporcionaria uma simplificação no teste e redução do custo, aumentando o acesso a esse tipo de teste. Num primeiro passo para otimizar as condições de operação do fluorímetro em desenvolvimento, foram coletados dados de amostras previamente medidas num equipamento comercial. Este artigo reporta a análise estatística desses dados utilizando a ferramenta *Box Plot*.³

O *Boxplot* é um tipo de gráfico usado regularmente na pesquisa científica e a sua construção é possível por meio de diversos *softwares* estatísticos.³ Ele permite a identificação de métricas de um conjunto de dados como média, mediana, percentagem de concentração (através da análise dos quartis) e dados *outliers*, que são dados que fogem ao esperado. E com isso é possível identificar as situações que geram melhores resultados e comparar os valores obtidos para tipos diferentes de amostras.

2. METODOLOGIA

As amostras utilizadas nos testes foram cedidas pelo ISI-SAS (Instituto SENAI de Sistemas Avançados de Saúde). As mesmas também foram previamente testadas por um equipamento comercial, o RT-qPCR em tempo real, modelo QuantiStudio 1 da Thermo Fisher Scientific, localizado no ISI-SAS, possibilitando um valor de referência para comparação.

Foram analisadas 120 amostras de controle e de diagnóstico entre 06/01/2021 e 10/02/2021. As amostras de controle detectam a presença do material genético humano, sendo que 47 eram positivas e 25 negativas onde o material genético foi substituído por água. Das amostras de diagnóstico, 25 testaram

positivo e 23 negativo para presença do SARS-COV-2 no equipamento de referência. Um fluorímetro de bancada utilizando um LED azul acoplado numa fibra óptica e um sensor espectral (AMS, mod. AS-7341) foi utilizado para medir a fluorescência emitida. Buscando otimizar os parâmetros de excitação e detecção do fluorímetro, foi caracterizada a resposta do sensor em ADC *counts* em função da corrente de alimentação do LED de excitação para 3 canais diferentes: o ciano, cuja banda de passagem está no pico da fluorescência produzida, o verde e o amarelo.

O resultado das medidas de fluorescência foram analisados utilizando a ferramenta estatística *Box Plot*, conforme exemplificado na Fig. 1. Essa análise foi feita para todos os grupos de amostra para as diferentes correntes de alimentação de LED. O *Box Plot* permitiu identificar os limites da distribuição e as amostras outliers. Para essa análise foi utilizado o *Power BI* como *software* de visualização dos gráficos, pois o mesmo possibilita a identificação mais dinâmica das métricas do *Box Plot*, especificamente quais amostras se comportavam como *outliers*. Para facilitar a uma comparação conjunta utilizou-se posteriormente o *Python* para plotar os mesmos gráficos e documentar os dados.

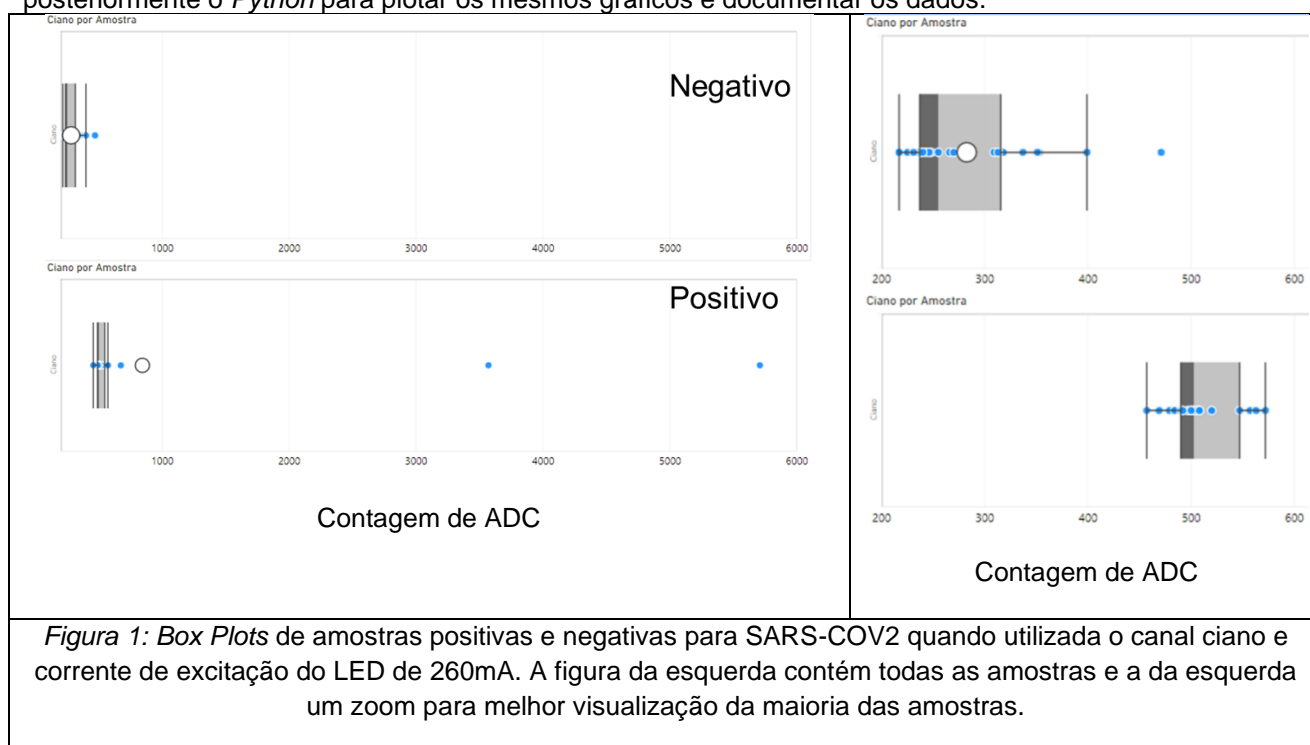


Figura 1: *Box Plots* de amostras positivas e negativas para SARS-COV2 quando utilizada o canal ciano e corrente de excitação do LED de 260mA. A figura da esquerda contém todas as amostras e a da esquerda um zoom para melhor visualização da maioria das amostras.

3. RESULTADOS E DISCUSSÃO

Para identificar a melhor configuração de excitação e detecção, foi plotada na Fig. 2 a separação entre as amostras que testaram positivo ou negativo para a presença de Sars-Cov-2 no equipamento de referência para os três canais monitorados, em função da corrente de excitação para as amostras. Os resultados são semelhantes para os canais ciano e verde e indicam que a corrente de alimentação do LED de excitação é ótima entre 0,26 e 0,39 A. As amostras de controle apresentaram resultados semelhantes.

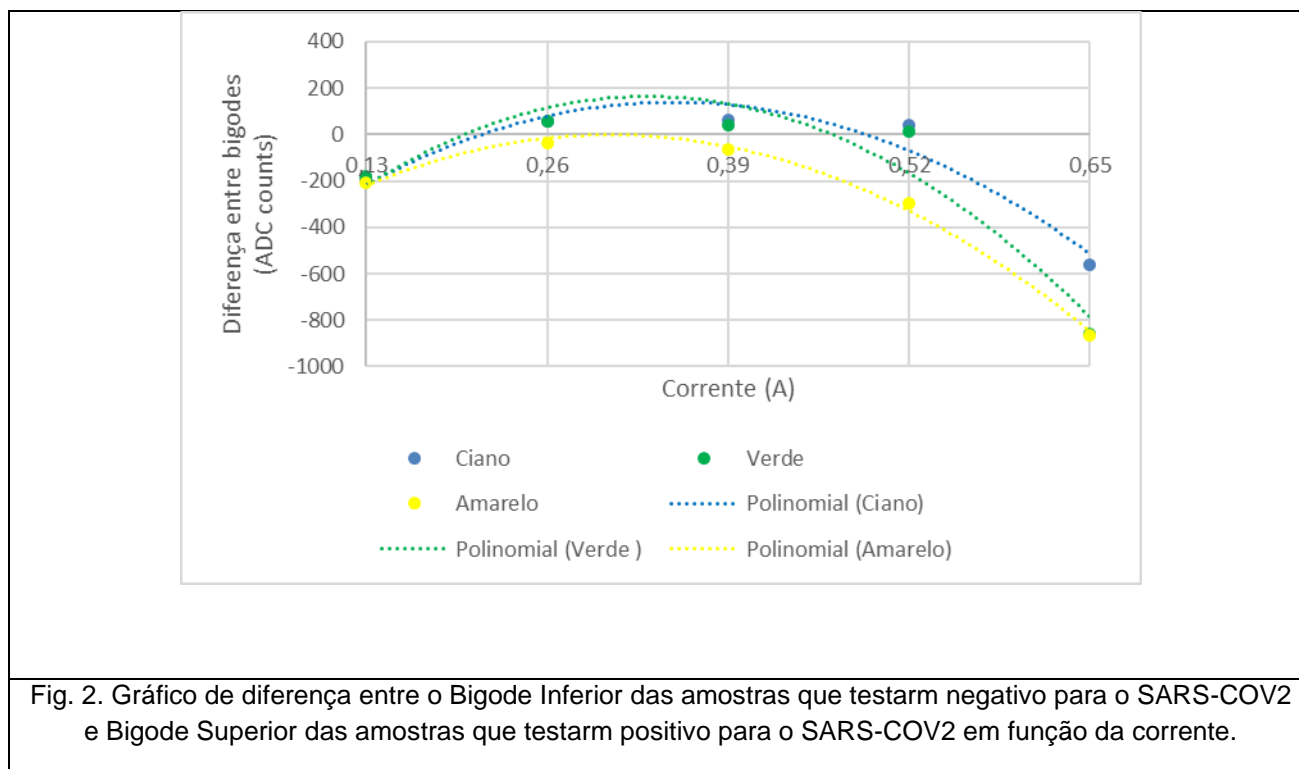


Fig. 2. Gráfico de diferença entre o Bigode Inferior das amostras que testaram negativo para o SARS-COV2 e Bigode Superior das amostras que testaram positivo para o SARS-COV2 em função da corrente.

4. CONSIDERAÇÕES FINAIS

Através da programação *Python* e do software *Power BI*, foi desenvolvida uma análise inicial das amostras do teste PCR. Através dos dados obtidos foi possível determinar a melhor intensidade luminosa a ser utilizada na excitação, ou seja, no LED, que foi com a corrente de 0,26 A, pois a mesma apresentou espaçamento interessante entre os *Boxes* e menos outliers quando comparada com correntes maiores. Dando continuidade ao trabalho de desenvolvimento, será analisada a linearidade da resposta para as condições de interesse.

Agradecimentos

Os autores agradecem à Fapesb, SENAI-CIMATEC e Fio Cruz pelo apoio financeiro

5. REFERÊNCIAS

- ¹HEID, C. A.; Stevens, J.; LIVAK, K. J. & WILLIAMS, P. M. **Real time quantitative PCR**. *Genome Res.*6, p. 986–994, 1996. <https://doi.org/10.1101/gr.6.10.986>.
- ²LOBATO, I. M.; O'SULLIVAN, C. K. **Recombinase polymerase amplification: Basics, applications and recent advances**. *TrAC Trends in Analytical Chemistry*, v. 98, p. 19–35, 1 Jan. 2018.
- ³NETO, José Valladares. **Boxplot: um recurso gráfico para a análise e interpretação de dados quantitativos**, 3rd edn. Springer, New York, 2006.