

**UNIVERSIDADE DE SÃO PAULO**

Programa Interunidades de Pós-Graduação em Bioinformática

**JOÃO VICENTE DE MORAIS MALVEZZI**

**Identificação por *phage display* de candidatos vacinais para  
esquistossomose mansônica baseada na auto-cura de macacos  
rhesus (*Macaca mulatta*)**

São Paulo  
2020



JOÃO VICENTE DE MORAIS MALVEZZI

**Identificação por *phage display* de candidatos vacinais para  
esquistossomose mansônica baseada na auto-cura de macacos  
rhesus (*Macaca mulatta*)**

**Versão Corrigida Simplificada**

Dissertação apresentada ao Programa Interunidades de Pós-graduação em Bioinformática para obtenção do Título de Mestre em Ciência pelo Instituto de Matemática e Estatística da Universidade de São Paulo.

Área de concentração: Bioinformática

Orientador: Sergio Verjovski-Almeida

Co-orientadora: Júlia Maria Pavan Soler

São Paulo

2020

## FICHA CATALOGRÁFICA

M262 Malvezzi, João Vicente de Morais      Ficha Catalográfica  
Identificação por phage display de candidatos vacinais para esquistossomose mansônica baseada na auto-cura de macacos rhesus (Macaca mulatta) / João Vicente de Morais Malvezzi, orientador Sergio Verjovski-Almeida e coorientadora Julia Maria Pavan Soler. São Paulo : 2020.  
85 p.

Dissertação (Mestrado) - Universidade de São Paulo  
Orientador: Prof. Dr. Sergio Verjovski-Almeida  
Coorientadora: Profa. Dra. Julia Maria Pavan Soler  
Programa Interunidades de Pós-Graduação em Bioinformática  
Área de concentração: Bioinformática

1. Esquistossomose. 2. Macaco rhesus. 3. Auto-cura. 4. Phage-display. 5. Candidatos vacinais I. Verjovski-Almeida, Sergio, orientador. II. Soler, Julia Maria Pavan, coorientadora. III. Instituto de Matemática e Estatística. IV. Universidade de São Paulo. V.Título.

CDD: 616.96

MALVEZZI, J. V. M. **Identificação por *phage display* de candidatos vacinais para esquistossomose mansônica baseada na auto-cura de macacos rhesus (*Macaca mulatta*)**. 2020. 85 p. Dissertação (Mestrado em Bioinformática) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2020.

Aprovado em: 9 de novembro de 2020

Banca Examinadora

Prof. Dr. Leonardo Paiva Farias

Instituição: Fundação Oswaldo Cruz (FIOCRUZ) - BA

Julgamento: **Aprovado**

Prof. Dr. Ricardo De Marco

Instituição: Instituto de Física de São Carlos (IFSC) - Universidade de São Paulo

Julgamento: **Aprovado**

Prof. Dr. João Carlos Setubal

Instituição: Instituto de Química (IQ) – Universidade de São Paulo

Julgamento: **Aprovado**

## **DEDICATÓRIA**

À minha parceira de vida, Larissa Logullo, e meus pais Jorge e Gelsa, pelo carinho, paciência e apoio ao longo de todo período de desenvolvimento deste trabalho

## AGRADECIMENTOS

Ao Instituto de Matemática de Estatística pela oportunidade de realização do curso de mestrado.

Ao Instituto Butantan, por colocar à disposição a área experimental e o laboratório.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pela bolsa de estudo no primeiro ano deste trabalho.

À Fundação de Amparo à Pesquisa do Estado de São Paulo, pela concessão da bolsa de mestrado e pelo apoio financeiro para a realização desta pesquisa (FAPESP no. 2018/18117-5).

Ao Prof. Dr. Sergio Verjovski-Almeida, pela oportunidade de trabalhar neste projeto, por todos os ensinamentos, contribuindo para meu crescimento científico e intelectual durante nossos anos de convivência.

À minha co-orientadora, Profa. Dra. Júlia Maria Pavan Soler, do Instituto de Matemática e Estatística da USP, pelas explicações e discussões sobre os modelos estatísticos ZIGP e Binomial Negativa, pelas dicas na utilização do método de MCMC e pela supervisão na análise dos dados.

Agradeço a todos meus amigos do VerjoLab, que estiveram ao meu lado durante este período. Obrigado por nossas conquistas, pelos conselhos, momentos de discussão e aprendizado. De certo que todos participaram no meu desenvolvimento pessoal.

Aos meus Pais Jorge e Gelsa, por todo amor, pela dedicação em criar uma família que valoriza a educação, pelas palavras de incentivo desde meus primeiros anos na graduação.

À minha Noiva Larissa, por sempre estar ao meu lado, oferecendo toda paciência e suporte nos períodos difíceis, e por estar presente nos momentos de felicidade. Por todas conversas, aprendizado, amor e carinho.

*“O método científico é comprovado e verdadeiro. Não é perfeito, é apenas o melhor que temos. Abandoná-lo, junto com seus protocolos céticos, é o caminho para uma idade das trevas.”*

**Carl Sagan**



## RESUMO

MALVEZZI, J. V. M. **Identificação por *phage display* de candidatos vacinais para esquistossomose mansônica baseada na auto-cura de macacos rhesus (*Macaca mulatta*)**. 2020. 85 p. Dissertação (Mestrado em Bioinformática) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2020.

Indivíduos do gênero *Schistosoma* são hemoparasitas causadores da esquistossomose, uma doença distribuída mundialmente que afeta mais de cem milhões de pessoas. No Brasil, o agente etiológico da esquistossomose é o *Schistosoma mansoni*. Desde a publicação do transcriptoma da espécie em 2003, foi obtido pouco sucesso nos testes de genes candidatos vacinais, sendo necessários novos métodos para reconhecer novos alvos candidatos. Uma abordagem é a utilização do macaco rhesus como modelo de estudo, que é capaz de curar-se espontaneamente da infecção por *Schistosoma*. Este trabalho envolve o desenvolvimento de um *pipeline* de bioinformática essencial para explorar dados obtidos com uma nova tecnologia de *phage display* com oligonucleotídeos sintéticos, capaz de fazer um *screening* dos anticorpos produzidos por macacos rhesus infectados e auto-curados; utilizamos pela primeira vez o macaco rhesus como modelo animal em conjunto com esta nova tecnologia. Aproximadamente 120 mil oligonucleotídeos sintéticos foram utilizados para construir uma biblioteca de *phage display* e representam, com mesma abundância, as sequências codificadoras que cobrem todos os fragmentos de 58 aminoácidos ao longo de todas as proteínas do parasita que possuem sequência conhecida, permitindo um *screening* não-enviesado de todos os anticorpos dos macacos rhesus que medeiam a auto-cura. No ensaio de *phage display* foi usado sequenciamento em larga-escala dos fagos capturados pelas amostras de soro de cada um de 12 macacos infectados, colhidas nos tempos 0 e 8 semanas após infecção. O *pipeline* de bioinformática aqui desenvolvido identificou que 98.3% das sequências de oligonucleotídeos da biblioteca foram detectadas pelo menos uma vez entre as 63 amostras sequenciadas. Neste pipeline, são utilizadas a distribuição de Poisson generalizada com zeros inflacionados e a distribuição binominal Negativa como abordagens estatísticas para calcular a significância de enriquecimento dos peptídeos capturados pelo soro de cada macaco em cada tempo. 421 peptídeos foram identificados como significativamente enriquecidos por ambos os métodos estatísticos, e deste resultado 61 peptídeos estavam enriquecidos em pelo menos quatro amostras de macacos da semana 8 após a infecção, sendo 9 peptídeos pertencentes a 8 genes codificadores de proteínas (e anotados como SmPs). Os restantes 52 peptídeos representam diferentes isoformas de *splicing* alternativo de 2 genes de micro-exons (anotados como micro-

exon genes, MEGs). Entre as proteínas Smgs, apenas uma está descrita como intracelular, sendo as demais descritas como expressas na interface parasita-hospedeiro. Os anticorpos desenvolvidos pelos macacos contra estas proteínas podem estar envolvidos no processo de morte do parasita. Com a aplicação do nosso pipeline de análise de enriquecimento nos dados de sequenciamento de fagos capturados pelas amostras de soro dos macacos colhidas entre as semanas 10 e 62 após a infecção, que ainda estão em processamento, esperamos continuar identificando estes peptídeos já reconhecidos na semana 8, além de identificar novos peptídeos antigênicos do parasita, que possam ser testados em estudos futuros para o desenvolvimento de uma vacina contra a esquistossomose.

**Palavras-chave:** Esquistossomose. Macaco rhesus. Auto-cura. *Phage display*. Candidatos vacinais.

## ABSTRACT

MALVEZZI, J. V. M. **Phage display identification of vaccine candidates for schistosomiasis mansonii based on self-healing of rhesus monkeys (*Macaca mulata*)**. 2020. 85 p. Masters Thesis (Masters in Bioinformatics) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2020.

Individuals of the genus *Schistosoma* are hemoparasites that cause schistosomiasis, a worldwide distributed disease that affects more than one hundred million people. In Brazil, the etiologic agent of schistosomiasis is *Schistosoma mansoni*. Since the publication of the species transcriptome in 2003, there has been little success in testing vaccine candidates, pointing to the need of new methods to recognize new candidate targets. One approach is to use rhesus monkeys as a study model, which is able to heal spontaneously from *Schistosoma* infection. This work involves the development of an essential bioinformatics pipeline to explore the data obtained with a new phage display technology with synthetic oligonucleotides, capable of screening the antibodies produced by infected and self-cured rhesus monkeys; we used the rhesus monkey as an animal model for the first time in conjunction with this new technology. The approximately 120,000 synthetic oligonucleotides used to build the phage display library represent, with equal abundance, the coding sequences that cover all 58-amino-acid fragments across all parasite proteins that have a known sequence, allowing for unbiased screening of all antibodies of rhesus monkeys that mediate self-healing. For the phage display assay, large-scale sequencing of the phages captured by serum samples from each of 12 infected monkeys, collected at times 0 and 8 weeks after infection, was used. The bioinformatics pipeline developed here identified that 98.3% of the oligonucleotide sequences in the library were detected at least once among the 63 sequenced samples. In this pipeline, we use the zero-inflated generalized Poisson distribution and the Negative binomial distribution as two statistical approaches to calculate the significance of enrichment of the peptides captured by the serum of each monkey at each time point. 421 peptides were identified as significantly enriched by both statistical methods, and from this result 61 peptides were enriched in at least four serum samples from week 8 after infection, with 9 peptides belonging to 8 protein coding genes (annotated as Smps). The remaining 52 peptides represent different alternative splicing isoforms of 2 micro-exon genes (annotated as micro-exon genes, MEGs). Among Smp proteins, only one protein is described as intracellular, while the others are described as expressed at the host-parasite interface. The antibodies developed by monkeys against these proteins may be involved in the parasite's death process. With the application of the enrichment analysis pipeline to the

sequencing data of phages captured by monkeys' serum collected between weeks 10 and 62 after infection, which are still being processed, we expect to continue identifying the same peptides already recognized in week 8, in addition to identifying new antigenic peptides from the parasite, which can be tested in future studies for the development of a vaccine against schistosomiasis.

**Keywords:** Schistosomiasis. Rhesus monkey. Self-healing. Phage display. Vaccine candidates.