



UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE BIOCÊNCIAS
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

ISABELA PIMENTEL DE ALMEIDA

PIPELINE OLIGOY PARA DESENHO DE SONDAS ***OLIGOPAINT*** DO
CROMOSSOMO Y INCLUINDO SEQUÊNCIAS REPETITIVAS

Durante o desenvolvimento deste trabalho a autora recebeu auxílio financeiro
da FAPESP (Processo nº 2019/14878-4)

São Paulo/SP

2021

ISABELA PIMENTEL DE ALMEIDA

***PIPELINE* OLIGOY PARA DESENHO DE SONDAS *OLIGOPAINT* DO
CROMOSSOMO Y INCLUINDO SEQUÊNCIAS REPETITIVAS**

Versão Original

Dissertação de Mestrado apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática da Universidade de São Paulo para obtenção do título de Mestre em Ciências.

Área de Concentração: Bioinformática

Orientadora: Prof. Dra. Maria Dulcetti Vibranovski

Co-Orientador: Prof. Dr. Antonio Bernardo de Carvalho

São Paulo/SP

2021

FICHA CATALOGRÁFICA

A447 Almeida, Isabela Pimentel de
Pipeline OligoY para desenho de sondas oligopaint do cromossomo Y incluindo sequências repetitivas / Isabela Pimentel de Almeida, orientadora Maria Dulcetti Vibranovski, co-orientador Antonio Bernardo de Carvalho -- São Paulo : 2021.

142 p.

Dissertação (Mestrado) -- Universidade de São Paulo
Orientadora: Profa. Dra. Maria Dulcetti Vibranovski
Co-orientador: Prof. Dr. Antonio Bernardo de Carvalho
Programa Interunidades de Pós-Graduação em Bioinformática
Área de concentração: Bioinformática
Versão Original

1. Bioinformática. 2. FISH. 3. Citogenética. 4. Heterocromatina. I. Vibranovski, Maria Dulcetti, orientadora. II. Carvalho, Antonio Bernardo de, co-orientador. III. Universidade de São Paulo. IV. Título.

CDD - 572.8

Um dos principais obstáculos em estudos com o cromossomo Y está relacionado ao estado heterocromático e altamente repetitivo desta estrutura, levando a dificuldades de montagem de *scaffolds* e *contigs* e, assim, na falta de sequências finais montadas para o mesmo. O cromossomo Y de *Drosophila melanogaster*, organismo modelo utilizado nessa pesquisa, tem tamanho estimado em 41 Mb de sequências ricas em repetições, mas apenas 10% das mesmas estão montadas na versão mais recente do genoma. Em contrapartida, o protocolo para desenho de sondas utilizadas em experimentos de marcação fluorescente de cromossomo completo (FISH *Oligopaint*) não inclui sequências repetitivas para evitar hibridização fora da região alvo. Por esse motivo, existem atualmente menos de 1500 sondas *oligopaint* para este cromossomo Y modelo, quantidade esta ao menos dez vezes menor quando comparada com a de outros cromossomos da mesma espécie. Além disso, essa quantidade é insuficiente para realizar ensaios de FISH *Oligopaint* com eficiência. O objetivo principal desta pesquisa é desenvolver uma *pipeline* que permita desenhar sondas *oligopaint* para o cromossomo Y de qualquer espécie de interesse. A *pipeline* final inclui a utilização de ferramentas livres e existentes em Bioinformática, identificação de sequências exclusivas do cromossomo de interesse, garante ao usuário a autonomia para escolha de parâmetros e efetivamente usa sequências repetitivas exclusivas do cromossomo alvo para desenhar sondas, maximizando assim a eficiência geral dos experimentos de citogenética. Após extensos testes e validações *in silico* e *in situ*, foi constatado que a aplicação da *pipeline* desenvolvida, OligoY, permite marcar o cromossomo Y sem gerar sinal fora do alvo, apesar da utilização de sequências repetitivas para desenho das sondas *oligopaint*.

Palavras-chave: Bioinformática, FISH, citogenética, heterocromatina

Abstract

One of the main obstacles in studies with the Y chromosome is related to the heterochromatic and highly repetitive state of this structure, leading to difficulties in assembling scaffolds and contigs and, thus, in a lack of final assembled sequences for it. The Y chromosome of *Drosophila melanogaster*, the model organism used in this research, has an estimated size of 41 Mb of repeat-rich sequences, but only 10% of them are assembled in the most recent genome release. In contrast, the protocol for designing probes used in full chromosome fluorescent labeling experiments (FISH Oligopaint) does not include repetitive sequences to avoid off-target hybridization. For this reason, there are currently less than 1500 oligopaint probes for this Y chromosome model, which is a value at least ten times smaller when compared to the one observed for other chromosomes of the same species. Furthermore, this amount is insufficient to carry out FISH Oligopaint assays efficiently. The main objective of this research is to develop a pipeline that allows the design of oligopaint probes for the Y chromosome of any species of interest. The final pipeline includes the use of open-source and existing tools in Bioinformatics, identification of sequences unique to the chromosome of interest, guarantees the user the autonomy to choose parameters and effectively uses repetitive sequences unique to the target chromosome to design probes, thus maximizing overall efficiency of cytogenetic experiments. After extensive tests and validations *in silico* and *in situ*, it was verified that the application of the developed pipeline, OligoY, allows staining the Y chromosome without generating off-target signal, despite the use of repetitive sequences for oligopaint probe design.

Key-words: Bioinformatics, FISH, cytogenetics, heterochromatin