

Modelagem matemática e simulação computacional da dinâmica da replicação de kinetoplastídeos

Projeto de Iniciação Científica apresentado na disciplina MAC215

Aluno: Gustavo Rodrigues Cayres da Silva

Orientador: Marcelo da Silva Reis

Instituto Butantan, São Paulo, 11 de agosto de 2016.

Resumo

Neste artigo apresentamos, no contexto da disciplina MAC215, o projeto de Iniciação Científica desenvolvido por Gustavo Rodrigues Cayres Silva no Laboratório Especial de Toxinologia Aplicada do Instituto Butantan.

Sumário

1	Introdução	1
2	Objetivos	2
3	Metodologia	2
3.1	Modelagem do sistema dinâmico estudado	2
3.2	Formalização da dinâmica do modelo utilizando cadeias de Markov	3
3.3	Implementação, simulação e testes do modelo dinâmico	3
4	Tarefas Planejadas	3
5	Forma de avaliação dos resultados	3
	Referências	3

1 Introdução

Em organismos eucariotos, a replicação de DNA consiste na duplicação dos cromossomos que compõem o material genético de uma dada célula. Tal processo ocorre essencialmente de forma paralelizada, iniciando-se diversas vezes ao longo da fase S do ciclo celular em sítios denominados origens de replicação. Todavia, a frequência e o período de disparo de cada origem de replicação não são homogêneos. Além disso, há variações na eficiência da maquinaria de replicação em duplicar DNA a partir de cada uma dessas origens, o que em muitos casos envolve colisões dessa maquinaria com a de transcrição. A dinâmica de colisão entre maquinarias de replicação e de transcrição deve variar entre diferentes espécies, levantando a hipótese de que ela está por trás das diferenças de arquitetura genômica observadas entre diferentes espécies protozoários do grupo de kinetoplastídeos (*TriTryps*).

Um esforço para investigar essa questão foi reportado em 2015: utilizando *next-generation marker frequency analysis* (MFA-Seq), uma técnica ômica de larga escala, foi mapeada a distribuição das origens de replicação nos cromossomos de *Trypanosoma brucei* e de espécies do gênero *Leishmania*, todas elas do grupo dos kinetoplastídeos [1]. Mais recentemente, sob a coordenação da Dra. Maria Carolina Elias, foi iniciado o Projeto Temático “*How do common and diverged features of the replicative stress response shape the biology of TriTryp parasites?*”, com participação do Dr. Richard McCulloch (Universidade de Glasgow, Reino Unido). Através dessa colaboração, nosso grupo fez o mapeamento com MFA-Seq das origens de replicação

em *Trypanosoma cruzi*, o kinetoplastídeo causador da Doença de Chagas. Apesar desse ensaio mostrar de forma imediata a frequência de disparo de cada origem de replicação, a dinâmica de colisão entre maquinarias de replicação e de transcrição, tanto para *T. cruzi* como para os demais kinetoplastídeos, permanece como problema em aberto.

2 Objetivos

Neste projeto, propomos a construção de um modelo computacional estocástico para descrever a dinâmica de replicação do DNA de *Trypanosoma cruzi*. Esperamos, com isso, poder estimar a dinâmica da replicação nesse organismo, incluindo como as colisões entre as maquinarias de replicação e de transcrição influenciam o processo de replicação de DNA como um todo.

O desenho do modelo deverá levar em consideração a distribuição das origens de replicação, assim como as velocidades médias de replicação e de transcrição, os sítios de início de transcrição e outras informações a priori, como as extraídas de modelos cinéticos de redes de sinalização molecular. Inicialmente propomos modelar a dinâmica desse sistema utilizando cadeias de Markov.

Já a implementação do modelo será realizada utilizando Python e seguirá boas práticas de programação, incluindo o uso de controladores de versão e de testes automatizados de unidade e de integração. Numa segunda etapa, pretendemos estender o uso desse modelo para outros kinetoplastídeos, em particular para *Trypanosoma brucei* e espécies do gênero *Leishmania*.

3 Metodologia

3.1 Modelagem do sistema dinâmico estudado

Devemos ter em mente que a duplicação de cada cromossomo se inicia a partir de uma origem de replicação durante um momento específico do ciclo de vida celular; o processo de transcrição (produção de RNA mensageiro), por outro lado, está constantemente se reiniciando. Primeiramente, podemos visualizar o objeto de estudo da pesquisa através dos esquemas na figura 1. Percebemos que, como as transcrições não cessam durante o processo de replicação, há um risco constante de choque entre as duas maquinarias.

O modelo deverá ser capaz de traduzir as características essenciais do sistema dinâmico; explicitando, por exemplo, as regiões de transcrição e as origens de replicação do cromossomo sendo analisado. Esse modelo também deverá traduzir o comportamento das maquinarias de replicação e transcrição e, num momento seguinte, possibilitará a simulação de colisões entre tais maquinarias. Analisar a influência dessas colisões no processo de replicação como um todo é o objetivo principal desta pesquisa.

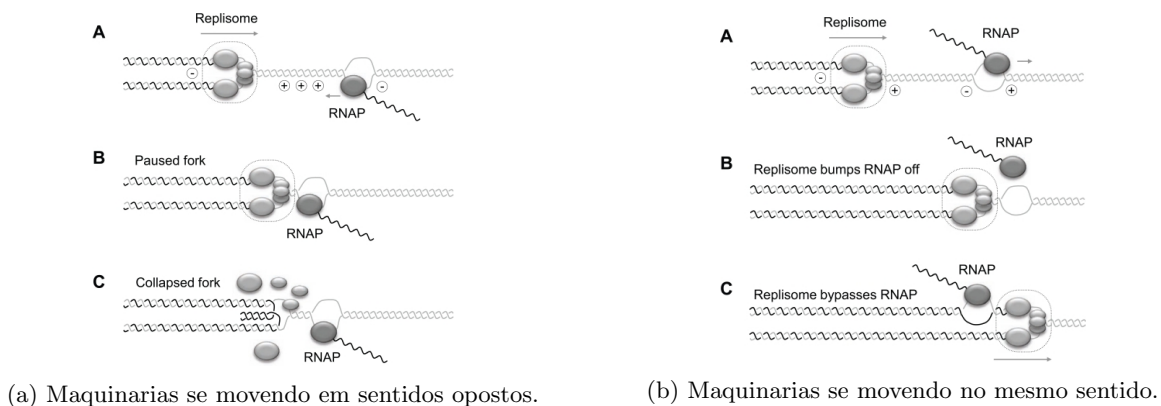


Figura 1: Colisão entre as maquinarias de replicação e a de transcrição. Figura extraída do artigo de Lin e colegas (2012) [2].

3.2 Formalização da dinâmica do modelo utilizando cadeias de Markov

Durante a modelagem, ao menos dois pontos relativos ao sistema dinâmico serão tratados através de métodos estocásticos:

- Escolha da origem de replicação que disparará em cada simulação, dentre as possíveis origens de replicação do cromossomo;
- Decisão do tempo de resolução das colisões *head-to-head* entre uma maquinaria de replicação e alguma maquinaria de transcrição. Essa resolução, que consiste em eliminar a maquinaria de transcrição e retomar o processo de replicação, possui tempo variável.

As primeiras simulações terão como objetivo especificar o comportamento das variáveis aleatórias do sistema. Como não podemos ter certeza da memória (influência de estados passados) do processo, é necessário utilizarmos cadeias de Markov para que possamos realizar as estimativas com vários graus de dependência passada. Com isso, seremos capazes de responder a perguntas como:

- A escolha da origem de replicação de um cromossomo segue qual distribuição de probabilidade?
- Essa escolha depende da origem escolhida durante a última replicação daquele cromossomo?
- Qual a distribuição do tempo de retomada da replicação após uma colisão *head-to-head*?
- A ocorrência de seguidas colisões altera esse tempo de retomada?

3.3 Implementação, simulação e testes do modelo dinâmico

Como citamos na seção 2, a implementação do modelo será realizada em *Python*, fazendo uso de Programação Orientada a Objetos (POO) com o intuito de aproximar o modelo computacional e o problema real. A simulação terá como base dados produzidos no próprio laboratório da pesquisa (Laboratório Especial de Toxinologia Aplicada do Instituto Butantan) e também informações extraídas de bancos de dados públicos tal como o TriTrypDB (<http://tritrypdb.org/tritrypdb/>) [3]. A manutenção do bom funcionamento do modelo será realizada através da implementação de testes automatizados – tanto de unidade quanto de integração – e do uso de controle de versão *git*, além da aplicação geral de boas técnicas de programação.

4 Tarefas Planejadas

Entre as tarefas que deverão ser realizadas a fim de atingirmos os objetivos estão:

- Organização em um banco de dados relacional das informações referentes à replicação de *T. cruzi* que foram produzidas em nosso laboratório e também as que se encontram em bancos públicos;
- Desenho do modelo computacional estocástico, com implementação em Python e utilização de testes automatizados dessa linguagem e da ferramenta de controle de versão *git*;
- Ajuste do modelo utilizando os dados organizados no banco de dados citado no item 1;
- Análise de sensibilidade dos parâmetros do modelo, testes de robustez e de validação. Extrapolação da metodologia para outros kinetoplastídeos e, como *benchmarking*, também para levedura.

5 Forma de avaliação dos resultados

A divulgação será feita através de relatórios semanais pelo GitHub, no repositório:

https://github.com/GustavoCayres/Relatorios_MAC0215.

Referências

- [1] Catarina A Marques, Nicholas J Dickens, Daniel Paape, Samantha J Campbell, and Richard McCulloch. Genome-wide mapping reveals single-origin chromosome replication in leishmania, a eukaryotic microbe. *Genome biology*, 16(1):1, 2015.
- [2] Yea-Lih Lin and Philippe Pasero. Interference between dna replication and transcription as a cause of genomic instability. *Current genomics*, 13(1):65–73, 2012.
- [3] Martin Aslett, Cristina Aurrecochea, Matthew Berriman, John Brestelli, Brian P Brunk, Mark Carrington, Daniel P Depledge, Steve Fischer, Bindu Gajria, Xin Gao, et al. Tritrypdb: a functional genomic resource for the trypanosomatidae. *Nucleic acids research*, 38(suppl 1):D457–D462, 2010.