Desenho e simulação de modelos computacionais da dinâmica de replicação de DNA em kinetoplastídeos

Monografia apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
aprovação em MAC499 – Trabalho
de
Conclusão de Curso

Aluno: Gustavo Rodrigues Cayres Silva

Orientador: Marcelo da Silva Reis

Centro de Toxinas, Resposta-imune e Sinalização Celular
(CeTICS)

Laboratório Especial de Ciclo Celular, Instituto Butantan

São Paulo, 27 de Novembro de 2017

Resumo

A replicação do DNA representa a duplicação do material genético de uma célula, preparando-a para a divisão celular, um processo essencial para o desenvolvimento e reprodução de todos os organismos vivos. Através deste trabalho, discutimos a importância do estudo desse processo e fornecemos uma base de conceitos biológicos para auxiliar o leitor na compreensão dos resultados. Analisamos com profundidade os detalhes do processo de replicação, sua eficiência e a influência de outros processos intracelulares sobre o mesmo. No centro deste projeto está o desenvolvimento de um modelo computacional capaz de simular a replicação do DNA e registrar eventos relevantes ocorridos durante tal processo. Este modelo foi testado com o estudo do material genético do *Trypanosoma brucei*, através do qual fomos capazes de obter resultados relevantes sobre a posição e quantidade de origens replicativas necessárias para o sucesso do processo de replicação.

Palavras-chave: Ciclo celular, modelo computacional, replicação do DNA, transcrição do DNA, tripanosomatídeos

Conteúdo

1 Introdução **1**

2 Conceitos Fundamentais **3**
2.1 DNA . . . 3
2.2 Ciclo Celular . . . 3
2.3 Fase S . . . 4
2.4 Origens de Replicação . . . 4
2.5 RNA . . . 5
2.6 Transcrição . . . 5
2.7 Colisões . . . 5

3 Modelo **8**
3.1 Representação das Estruturas Celulares . . . 8
3.1.1 Genoma . . . 8
3.1.2 Cromossomo . . . 8
3.1.3 Maquinarias de Replicação . . . 10
3.1.4 Maquinarias de Transcrição . . . 10
3.2 Representação da Dinâmica do Ciclo Celular . . . 10

4 Implementação **14**
4.1 Linguagem . . . 14
4.2 Banco de Dados . . . 15
4.3 Orientação a Objetos . . . 15
4.4 Execução . . . 16

5 Avaliação do Trabalho **20**
5.1 Visão Geral . . . 20
5.2 Histórico . . . 20
5.3 Desafios . . . 21
5.4 Agradecimentos . . . 21

Capítulo 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". Também durante o ciclo celular acontece o processo de transcrição de RNA ao longo do cromossomo. Propriedades como a probabilidade de ativação de cada origem de replicação e o momento da fase S em que elas são ativadas não são homogêneos, tão como a atividade de transcrição em cada região do DNA. Há variações, ainda, na eficiência da maquinaria de replicação em duplicar o DNA a partir de cada uma dessas origens; o que, em muitos casos, é devido a colisões dessa maquinaria com a de transcrição. A dinâmica de colisão entre maquinarias de replicação e de transcrição deve variar entre espécies distintas, levantando a hipótese de que ela esteja por trás das diferenças na arquitetura genômica observadas entre as várias espécies de protozoários do grupo dos kinetoplastídeos (*TriTryps*) [Sil+17].

Vários esforços para investigar essa questão foram reportados: em 2012, 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* [Tie+12] e, em 2015, de espécies do gênero *Leishmania* [Mar+15], todas elas do grupo dos kinetoplastídeos. 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, foi feito 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.

Este trabalho propõe-se a abordar esse problema através da construção de um modelo computacional capaz de simular a fase S do ciclo celular, tendo em mente dados biológicos

como a distribuição das origens de replicação, as velocidades médias de replicação e de transcrição e os sítios de início de transcrição [Cal+15], permitindo a adequação o modelo aos resultados observados *in vitro*. Dessa forma esperamos obter um maior entendimento sobre a influência das colisões sobre a replicação de DNA como um todo. pelo Prof. Dr. Hugo Armelin.

A corretude do modelo foi validada através de um grande número de simulações de Monte Carlo e futuras contribuições para seu desenvolvimento foram facilitadas através do uso de controle de versão e da organização das informações biológicos em um banco de dados relacional SQL. O programa foi implementado em *Python* e utiliza bibliotecas matemáticas como o *NumPy* e o *SciPy* O modelo está escrito em *Python* e sua validação foi feita através de um grande número de simulações de Monte Carlo. O programa utiliza um banco de dados SQL e controle de versão, com o intuito de facilitar futuras contribuições ao desenvolvimento do software.

No próximo capítulo, explicaremos alguns conceitos importantes para o entendimento deste trabalho. Nos capítulos 3 e 4, percorreremos os detalhes biológicos e computacionais do desenvolvimento, respectivamente. Já no capítulo 5, analisaremos os resultados obtidos, comparando-os com resultados obtidos por experimentos *in vitro* e destacaremos a contribuição desses resultados para o avanço da área. No capítulo final, faremos uma avaliação do trabalho realizado durante este ano e das dificuldades para a criação do projeto e desta monografia.

Capítulo 2

Conceitos Fundamentais

O entendimento desta monografia envolve alguns conhecimentos essenciais sobre biologia celular; portanto, neste capítulo, abordaremos alguns desses conceitos a fim de facilitarmos a leitura dos capítulos seguintes.

2.1 DNA

O DNA (abreviação de *deoxyribonucleic acid*) é a molécula que carrega as instruções do material genético de um organismo vivo. Suas informações controlam toda a dinâmica da vida do organismo, incluindo seu crescimento e reprodução. Grande parte dos recursos e da arquitetura de uma célula são voltados a manutenção e a produção desta molécula. O DNA é formado por duas fitas, cada uma consistindo de uma sequência de nucleotídeos (moléculas pequenas que funcionam como "tijolos" na montagem da molécula de DNA), como representado na figura 2.1.

2.2 Ciclo Celular

O ciclo celular é uma série de eventos que ocorrem em cada célula, culminando na duplicação de seu material genético e em sua divisão celular. A divisão celular é realizada por todos os organismos vivos, sendo essencial tanto para seu crescimento (no caso de organismos multicelulares) quanto para sua reprodução. Em organismos eucariotos (aqueles que possuem membranas envolvendo o material genético e outras partes internas da célula), como os kinetoplastídeos estudados neste trabalho, as fases que compõem o ciclo celular (figura 2.2) são:

1. **G1**: A célula cresce e se prepara para a duplicação do DNA.
2. **S**: A célula duplica seu DNA.
3. **G2**: A célula cresce e se prepara a a divisão celular.

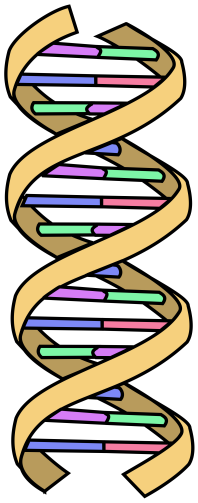


Figura 2.1: **Molécula de DNA.** Note sua estrutura de dupla-fita organizada em formato de hélice.

4. **M**: A célula se divide em duas células filhas.

5. **G0**: Cada uma das células filhas se prepara para entrar na fase G1.

Nesta pesquisa, estamos interessados em estudar os processos que ocorrem durante a fase S do ciclo celular e como esses processos interagem entre si.

2.3 Fase S

Durante a fase S do ciclo celular, ocorre a duplicação do material genético da célula, processo de grande interesse para nosso estudo. Durante esta fase, conjuntos de proteínas se ligam ao DNA, formando "maquinarias de replicação", as quais serão responsáveis por percorrer a dupla-fita de DNA, duplicando-a (figura 2.3). No fim da fase S, a célula terá o dobro do material genético inicial, quantidade necessária para a divisão celular que ocorrerá em seguida.

2.4 Origens de Replicação

Já foi mostrado que, na fase S da grande maioria dos organismos, as maquinarias proteicas de replicação não se ligam a qualquer região do DNA, mas sim a pontos específicos, cuja localização depende tanto de fatores filogenéticos (como a espécie do organismo), quanto de fatores mais complexos como o estado da molécula de DNA durante a replicação [Shi+98]. A esses pontos damos o nome de **origens de replicação**, conceito que será amplamente abordado nos próximos capítulos.

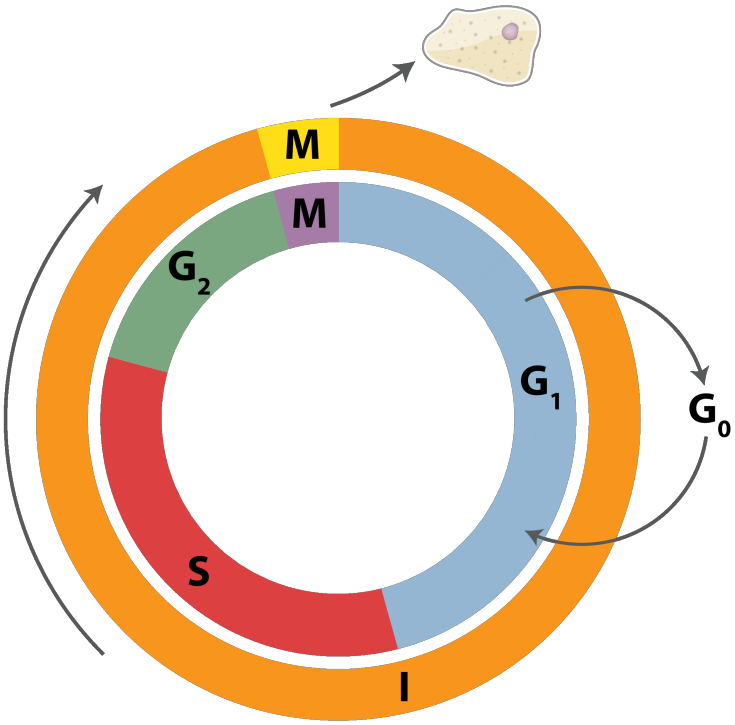


Figura 2.2: **Representação das fases do ciclo celular.** Durante o ciclo, a célula percorre as fases G1, S, G2 e M, nesta ordem. As fases G1, S e G2 são consideradas subdivisões da intérfase (segmento laranja).

2.5 RNA

Outra molécula de extrema importância para a célula, o RNA é responsável, entre outras funções, por expressar segmentos específicos do DNA, os genes. Em outras palavras, a molécula de RNA contém a informação genética do DNA em um formato que pode ser interpretado e utilizado pelos vários mecanismos das células, em destaque, para a produção de proteínas.

2.6 Transcrição

Outro processo essencial para a manutenção da vida é a transcrição do material em moléculas de RNA, ela ocorre quando proteínas se ligam a e percorrem os genes e, com as informações obtidas, produzem moléculas de RNA (figura 2.4).

2.7 Colisões

Tendo em vista os processos estudados anteriormente, fica claro que a replicação e a transcrição ocorrem de forma concomitante, levando o leitor a inferir que esses processos possam se afetar durante a fase S. De fato, quando as maquinarias de replicação e trans-

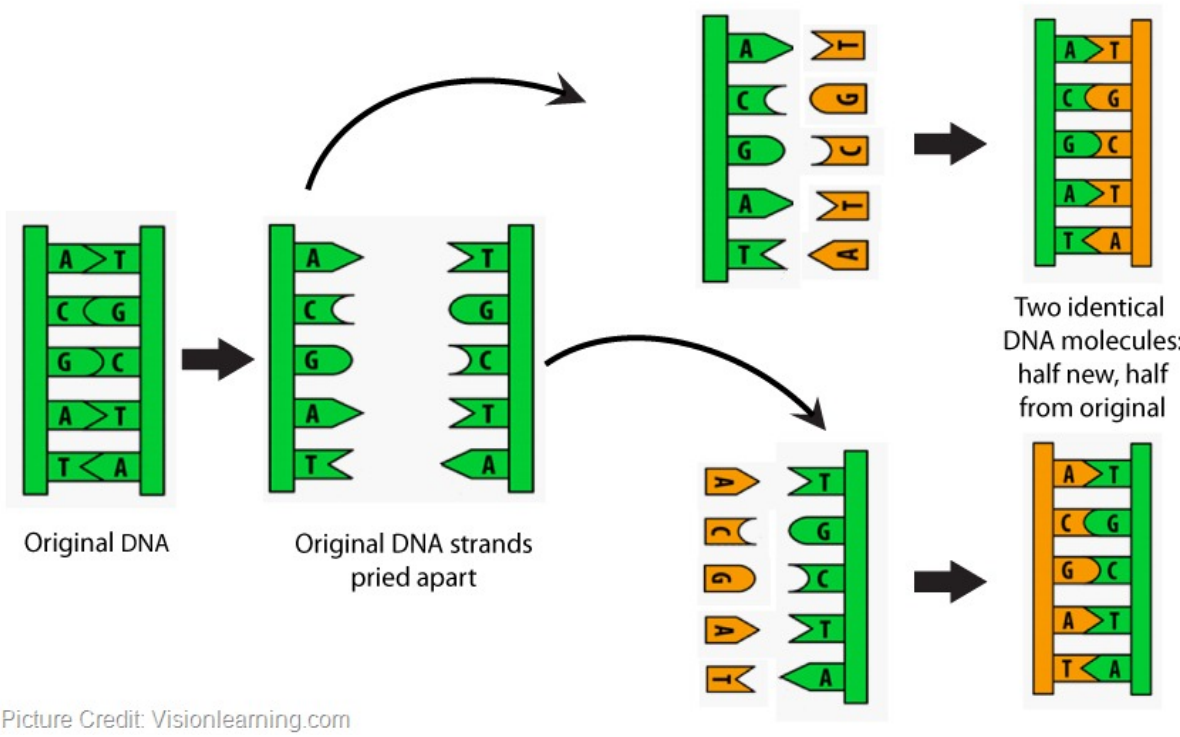


Figura 2.3: **Duplicação do material genético durante a fase S.** Proteínas se ligam à molécula de DNA e são responsáveis por separar suas duas fitas e por ligar os nucleotídeos que formarão a molécula nova.Durante o ciclo, a célula percorre as fases G1, S, G2 e M, nesta ordem. As fases G1, S e G2 são consideradas subdivisões da interfase (segmento alaranjado).

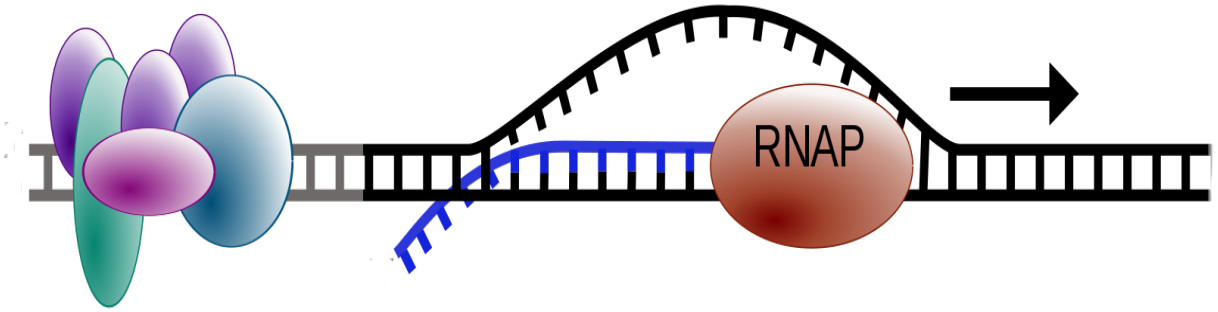


Figura 2.4: **Processo de transcrição.** RNAP (sigla para RNA polimerase, representada em laranja) é uma das proteínas responsáveis pela transcrição. Note a cadeia de RNA (azul) sendo criada *sobre* a cadeia de DNA (preto). As proteínas à esquerda estão realizando outras atividades sobre a molécula de DNA.

crição se aproximam sobre uma fita de DNA, elas se influenciam negativamente, levando a problemas (algumas vezes irreparáveis) na duplicação do material genético. No decorrer desta monografia, a essa influência negativa entre as maquinarias daremos o nome de "colisão". Em especial, se as maquinarias estiverem se movendo em sentidos opostos sobre o DNA no momento do conflito, a colisão será "frontal" (o caso em que as maquinarias se movem no mesmo sentido é de menor interesse para nosso projeto e chamaremos simplesmente de "colisão não-frontal").

Capítulo 3

Modelo

Na criação de nosso modelo, abordamos duas propriedades distintas da célula: as estruturas celulares e a dinâmica do ciclo celular. Somente através da adequação de ambas as propriedades fomos capazes de aproximar nosso modelo às observações *in vitro* e, assim, validar os resultados obtidos. Abaixo tratamos essas abordagens em detalhes.

3.1 Representação das Estruturas Celulares

Para a realização deste trabalho, focamos nossa atenção no processo de reprodução celular, especificamente nas propriedades que dizem respeito à replicação e transcrição do material genético. Portanto algumas estruturas da célula são de grande interesse para o modelo e suas características e comportamentos estão detalhados a seguir.

3.1.1 Genoma

O genoma é o conjunto de todos os cromossomos que compõem o material genético de uma dada célula (figura 3.1). Cada simulação de nosso modelo é aplicada sobre um instância do genoma do *Trypanosoma brucei* e, dessa forma, todos os resultados obtidos são relacionados ao genoma como um todo.

3.1.2 Cromossomo

O cromossomo é a unidade formadora do genoma e sobre sua estrutura ocorrem todos os processos relevantes para nosso estudo. Em nosso modelo, cada cromossomo é representado por uma fita dupla unidimensional e seu comprimento é o número de pares de bases que o formam (figura 3.2). Durante a simulação, cada cromossomo é percorrido em ambos os sentidos por várias estruturas, as maquinarias de replicação e de transcrição.

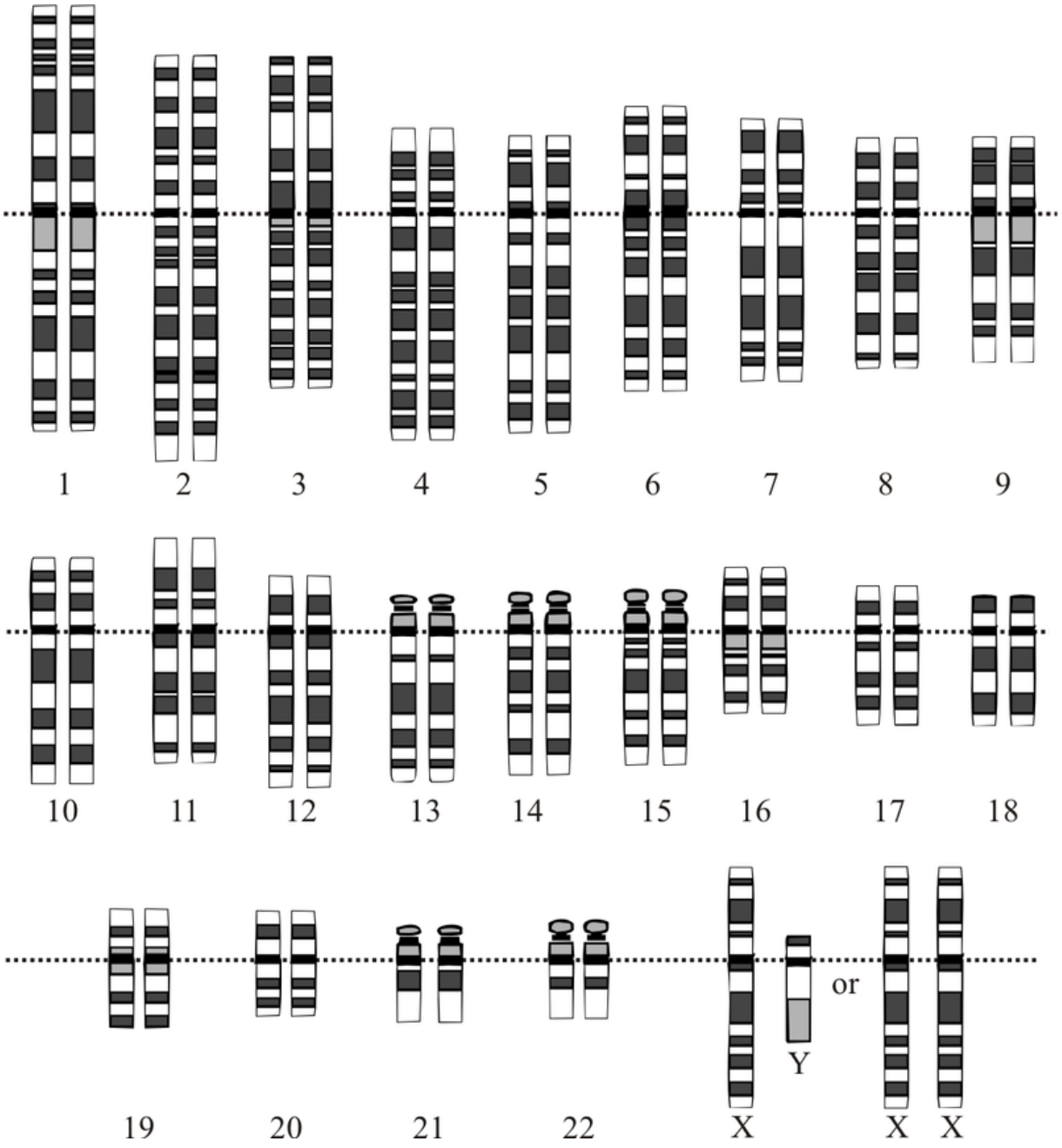


Figura 3.1: **Genoma Humano.** A coleção do material genético presenta nas células humanas. Em nosso modelo, estudamos o genoma do *Trypanosoma brucei*, o qual é consideravelmente menor (11 cromossomos).



Figura 3.2: **Representação do cromossomo.** Representação do cromossomo utilizada no modelo. As marcações 5' e 3' representam propriedades biológicas que influenciam a replicação, mas não são importantes para o entendimento deste trabalho.



Figura 3.3: **Representação do processo de replicação.** Note como as forquilhas replicativas (ovais azuis) partem das origens de replicação θ_1 e θ_2 e percorrem o cromossomo nos dois sentidos, duplicando-o (segmento verde).



Figura 3.4: **Representação do processo de transcrição.** As maquinarias de transcrição (círculos vermelhos) estão contidas às regiões de transcrição (setas amarelas): elas se ligam no início da região e se movem num único sentido (indicado pela seta) para seu fim. Note que múltiplas maquinarias de transcrição podem percorrer a mesma região de transcrição simultaneamente.

3.1.3 Maquinarias de Replicação

A fim de realizar a replicação do DNA, várias maquinarias proteicas se ligam ao cromossomo (o par de bases em que ocorreu a ligação é conhecido como "origem de replicação") e percorrem a fita de DNA em ambos os sentidos, formando duas forquilhas replicativas (figura 3.3).

3.1.4 Maquinarias de Transcrição

Essas maquinarias proteicas também se ligam ao cromossomo, entretanto o ponto em que ocorrerá a ligação e a região do cromossomo que será percorrida são predefinidas (as "regiões de transcrição"), o que leva a direção de movimento das maquinarias em um certa região a ser constante (figura 3.4).

Os dados necessários para a modelagem do genoma, incluindo o comprimento de cada cromossomo e as regiões de transcrição, foram obtidos do *TryTripDB*, um banco de recursos sobre kinetoplastídeos.

3.2 Representação da Dinâmica do Ciclo Celular

Devido a nosso interesse nos detalhes do processo de replicação do DNA, limitamos este estudo à fase S do ciclo celular, com as simulações capturando dados desde o momento da agregação da primeira maquinaria de replicação (em outras palavras, a ativação da primeira origem) até a duplicação completa do material genético.

A partir do momento em que uma maquinaria de replicação se liga ao DNA, duas forquilhas partem do par de base da ligação e começam a percorrer o cromossomo em

sentidos opostos, duplicando a fita numa velocidade de aproximadamente 64 pares de bases por segundo [Cal+15]. Simultaneamente maquinarias de transcrição percorrem as regiões de transcrição do cromossomo (cada região de transcrição é percorrida em um sentido específico) a uma velocidade aproximada de 40 pares de bases por segundo [Cal+15].

O disparo das origens de replicação está limitado à disponibilidade de recursos da célula: o número máximo de forquilhas percorrendo o DNA em um determinado instante é um dos parâmetros de grande importância de nosso modelo (utilizamos desde uma quantidade pequena de recursos, por volta de 5 forquilhas simultâneas, até quantidades maiores do que o esperado para o genoma humano, aproximadamente 60 forquilhas [Gin+14]). Analogamente, o processo de transcrição também está sujeito aos recursos celulares. Essa restrição foi traduzida, no modelo, através do controle da frequência com que maquinarias iniciam o processo de transcrição em cada região (variamos a frequência entre 0 e 1 disparos por minuto, de acordo com fontes de nosso laboratório).

Como discutimos nos capítulos anteriores, um dos pontos de estudo centrais de nosso modelo são os conflitos existentes entre os processos de transcrição e replicação que ocorrem durante a fase S, principalmente devido a colisões entre as várias maquinarias percorrendo a fita de DNA concomitantemente. No modelo, as possíveis naturezas dos conflitos e seus respectivos tratamentos, tendo como base dados biológicos obtidos em nosso laboratório, são:

- **Colisão Frontal.** Ocorre quando uma forquilha replicativa encontra uma maquinaria de transcrição se movendo em sentido oposto (figura 3.5). A colisão frontal leva a um grande estresse sobre a estrutura do DNA, em geral causando danos a sua estrutura. Neste caso, a maquinaria de transcrição é retirada da fita de DNA, levando à falha desta transcrição. Da mesma forma, a forquilha replicativa é incapaz de continuar a duplicação do DNA e se desliga, ficando disponível para ser dispara em outro ponto do genoma (no mesmo ou em outro cromossomo). A duplicação do DNA só será completa caso forquilhas provindas de outra origem de replicação atinjam a região do conflito.

- **Colisão Não-Frontal.** Ocorre quando uma forquilha replicativa encontra uma maquinaria de transcrição se movendo no mesmo sentido (esta colisão é possível pois a velocidade média de transcrição é menor que a de replicação, como mostrado na figura 3.6). Em conflitos desta natureza, a maquinaria de transcrição também é retirada do DNA, porém o menor dano ao DNA permite que a forquilha replicativa continue o processo de duplicação do DNA normalmente.

Embora duas forquilhas replicativas possam colidir entre si, esse processo é esperado durante a fase S: ele representa o fim da duplicação de um segmento do DNA e é seguido pelo desprendimento das forquilhas, sem dano ao DNA (figura 3.7). As regiões de

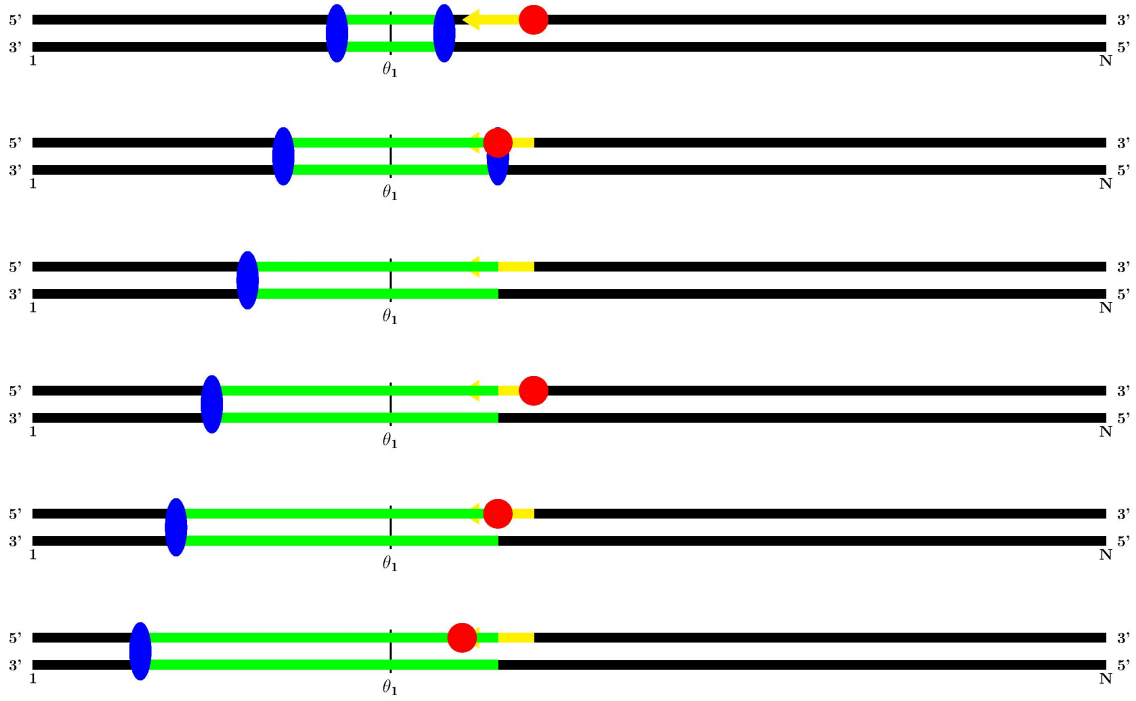


Figura 3.5: **Representação da colisão frontal entre maquinarias.** A forquilha replicativa movendo-se para a direita colide frontalmente com a maquinaria de transcrição. Isso faz com que ambas sejam retiradas da fita de DNA. Outras maquinarias de transcrição podem surgir e transcrever a região onde houve o conflito.

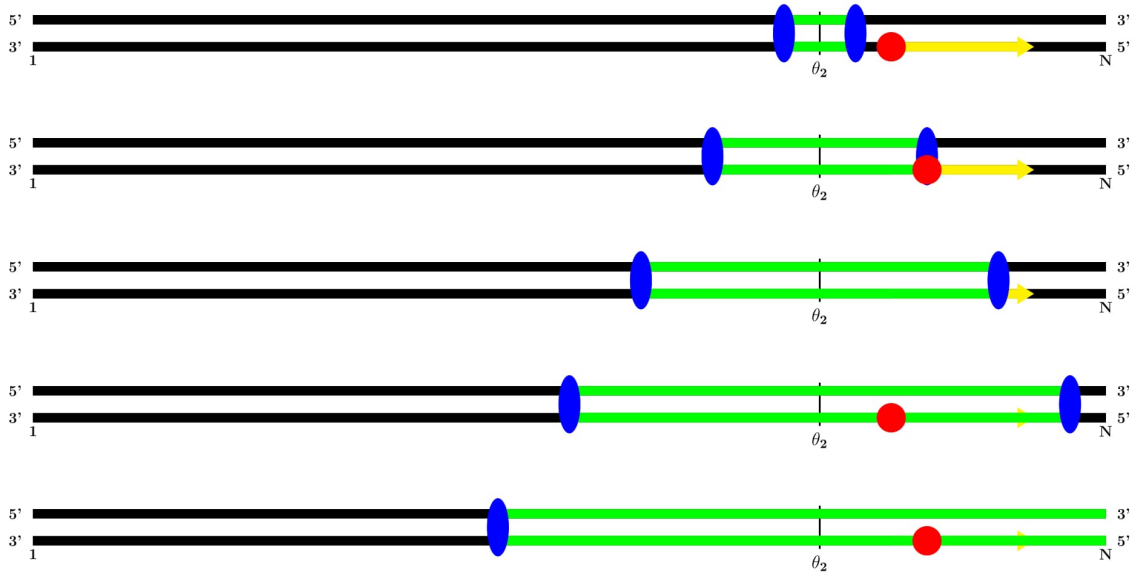


Figura 3.6: **Representação da colisão não-frontal entre maquinarias.** A forquilha replicativa movendo-se para a direita colide com a maquinaria de transcrição movendo-se no mesmo sentido. Isso faz com que a maquinaria de transcrição seja retirada da fita de DNA, mas a forquilha replicativa procede normalmente. Novas maquinarias de transcrição podem surgir sobre a região já duplicada do cromossomo.

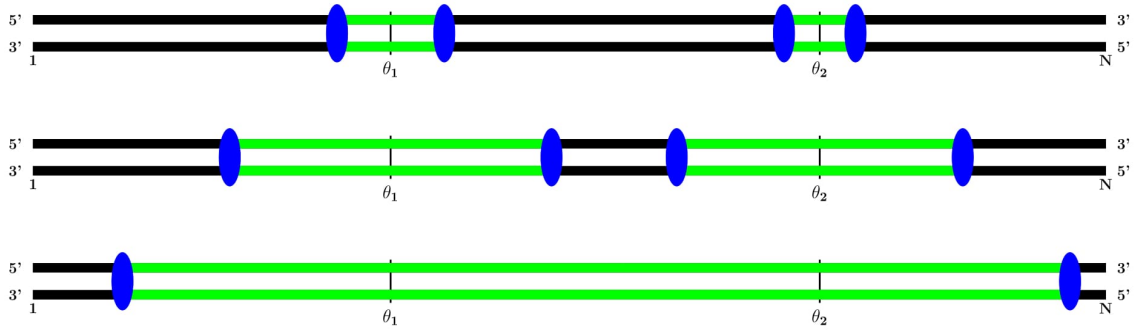


Figura 3.7: **Encontro de forquilhas replicativas.** Ao se encontrarem, as forquilhas se desligam do DNA e podem ser usadas futuramente no disparo de outras origens.

transcrição em uma fita do DNA são disjuntas, portanto todas maquinarias de transcrição percorrendo a mesma região se movem na mesma direção, garantindo que não haja colisão entre essas maquinarias.

Capítulo 4

Implementação

As decisões de implementação foram pontos continuamente revisados durante o desenvolvimento deste trabalho. Após o estudo extensivo de diversos outros trabalhos na área, organizamos o modelo computacional da seguinte forma.

4.1 Linguagem

Ao escrevermos o código do projeto, utilizamos quase que exclusivamente a linguagem Python. Através dessa linguagem, fomos capazes de:

- Desenvolver *scripts* para a análise e integração dos dados obtidos na literatura a nossos próprios. A variedade de formatos em que essas informações são encontradas requereu uma linguagem capaz de fazer o tratamentos de *strings* de forma rápida e natural.

- Organizar as estruturas biológicas com orientação a objetos, de forma a isolar as características de cada entidade e limitar a interação entre diferentes estruturas, permitindo uma melhor adequação dessas interações ao comportamento observado experimentalmente.

- Aproveitar a eficiência de poderosas bibliotecas matemáticas, como o *NumPy* e o *SciPy*, permitindo cálculos mais precisos e eficientes, principalmente no tratamento de colisões.

- Paralelizar a execução das simulações, aumentando a eficiência e a escalabilidade do modelo.

Em alguns raros momentos do desenvolvimento, utilizamos as linguagens Perl, para analisar os resultados das simulações, e SQL, para o gerenciamento do banco de dados.

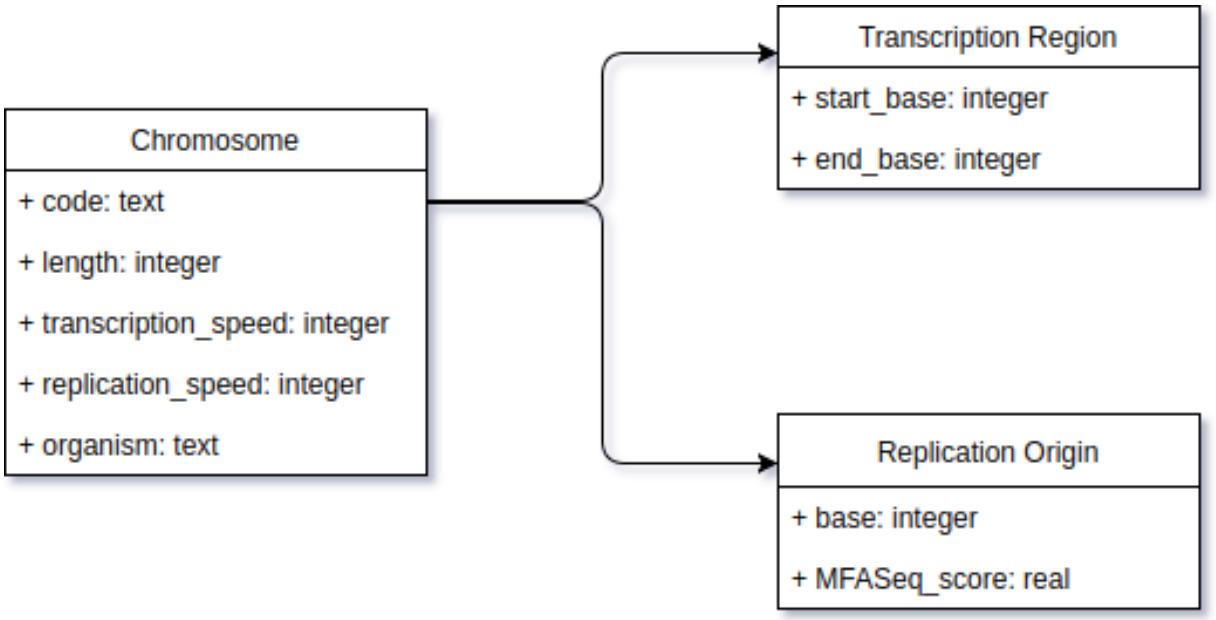


Figura 4.1: **Diagrama das Entidades do Banco.** Note que não é necessário modelar o genoma: ele é o conjunto de todos os cromossomos com o mesmo organismo.

4.2 Banco de Dados

Para tornar o acesso das informações eficientes e, mais importante, facilitar o compartilhamento dos dados obtidos com outros modelos, optamos por utilizar um banco de dados SQLite (figura 4.1), cujo ponto forte é a facilidade de acesso pela linguagem Python.

4.3 Orientação a Objetos

Na modelagem de processos biológicos, é natural o uso de orientação a objetos a fim de estudarmos cada estrutura isoladamente e, num momento seguinte, estudar o comportamento do sistema quando essas estruturas interagem entre si (figura 4.2). A classe *Genome* agrupa os cromossomos é responsável por verificar o fim da simulação, que ocorre quando todos os cromossomos foram replicados. A classe *Cromosome* é a responsável por armazenar e atualizar o estado dos pares de bases: quando uma forquilha replicativa percorre uma região de um cromossomo, este a duplica, marcando as bases que a compõem como duplicadas.

O cromossomo é percorrido por objetos da classe *Fork*, podendo ser instâncias da classe *Transcription Fork* ou *Replication Fork*. Como dito anteriormente, o movimento das forquilhas de replicação é registrado através da duplicação dos pares de bases, já o movimento da forquilha de transcrição não fica armazenado, contudo ele é imperativo para podermos avaliar as colisões entre as duas maquinarias.

Por fim, ambas as forquilhas possuem classes de gerenciamento, responsáveis por ligá-las ao cromossomo no momento oportuno. A classe *Transcription Region* representa cada

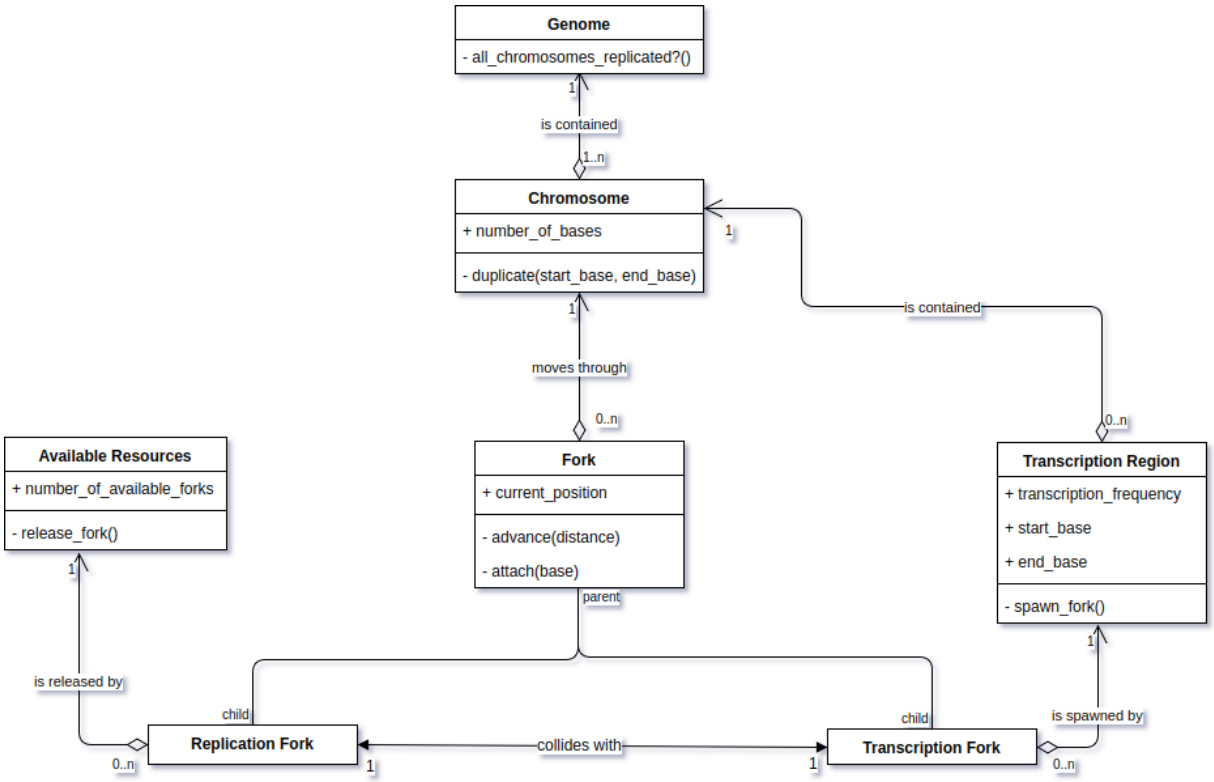


Figura 4.2: **Diagrama UML dos Classes Relevantes do Sistema.**

região de transcrição de um cromossomo e, com uma certa frequência pré-definida, cria novas maquinarias de transcrição que percorrem a região (desde o par de bases inicial da região até o final). Analogamente, a classe *Available Resources* libera forquilhas para se ligarem aos cromossomos. Note que cada região de transcrição é própria de um cromossomo (de fato, ela representa um segmento de bases daquele cromossomo); os recursos para forquilhas replicativas, por outro lado, são compartilhados por todos os cromossomos de um genoma. Dessa forma traduzimos mais precisamente a "competição" por material proteico característica da fase S do ciclo celular [Gin+14].

4.4 Execução

A execução do modelo envolve repetidas iterações sobre os processos de replicação e de transcrição, cada uma dessas iterações representando uma fração do tempo da fase S. Após a preparação dos cromossomos, o modelo imediatamente entre na fase S do ciclo celular, no qual dois processos principais acontecem: o avanço das maquinarias ligadas aos cromossomos e a ligação de novas maquinarias a esses.

O avanço das maquinarias, como exposto anteriormente, está sujeito a colisões de diversos tipos. Nosso sistema utiliza *Discrete Event Simulation*, ou seja, seu avanço ocorre em intervalos de tempo suficientemente pequenos, de forma que seja mínima a chance de ocorrência de múltiplos conflitos simultâneos. Assim, as colisões são tratadas

logo após sua ocorrência, seguindo o comportamento estipulado para colisões frontais e não-frontais discutido no capítulo anterior (figura 4.3). Devemos considerar, ainda, que cada maquinaria avança somente até seus limites. Forquilhas replicativas movem-se até encontrarem uma das pontas do cromossomo, outra forquilha seguindo em direção oposta ou uma base já duplicada do cromossomo. Já maquinarias de transcrição estão limitadas às suas respectivas regiões de transcrição.

A ligação de novas forquilhas replicativas depende de diversas informações obtidas *a priori* na literatura e foi um dos pontos de maior discussão deste trabalho (figura 4.4). A cada intervalo de tempo pré-definido, o programa escolha uma base dentre todas as do genoma para ser disparada (em outras palavras, para receber forquilhas replicativas e se tornar uma origem de replicação). Para toda base do genoma está associada uma probabilidade de ativação que será testada quando aquela base for escolhida para o disparo. Ser escolhida e passar no teste de disparo, entretanto, não são suficientes para escolher a base como uma origem: somente bases ainda não duplicadas podem ser escolhidas como origens (caso contrário, forquilhas replicativas seriam ligadas a regiões já duplicadas, o que é biologicamente impossível).

O ligação de maquinarias de transcrição é mais simples: a cada intervalo de tempo pré-definido, cada região de transcrição dispara uma maquinaria em sua posição inicial. Essa maquinaria percorre a região e é retirada ao atingir o fim daquela (ou possivelmente antes, no caso de algum conflito).

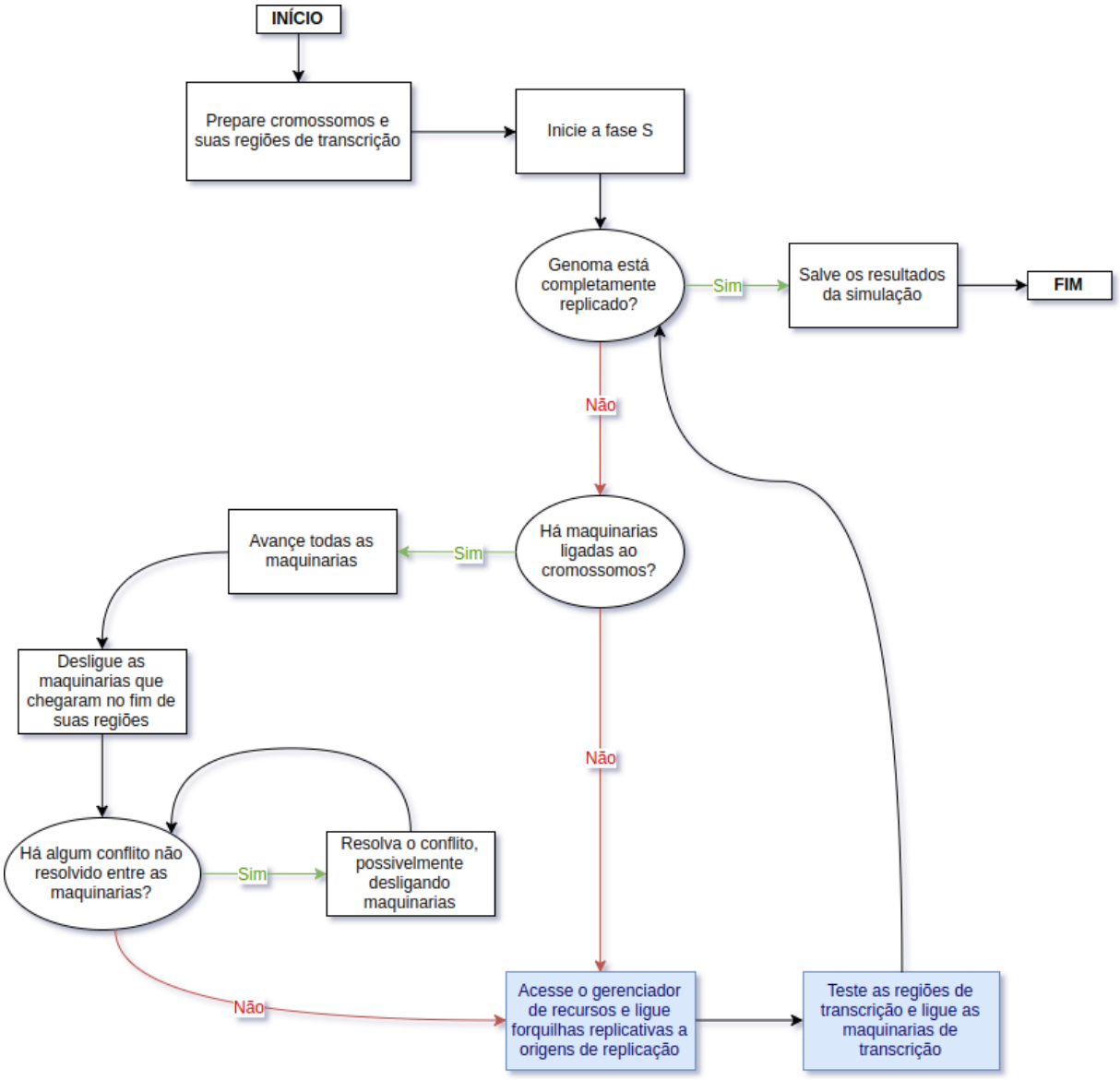


Figura 4.3: **Fluxograma Principal da Execução do Programa.** As caixas retangulares representam processos da execução, enquanto as ovais mostram pontos de testes. As caixas em azul são explicadas em mais detalhes nesta seção.

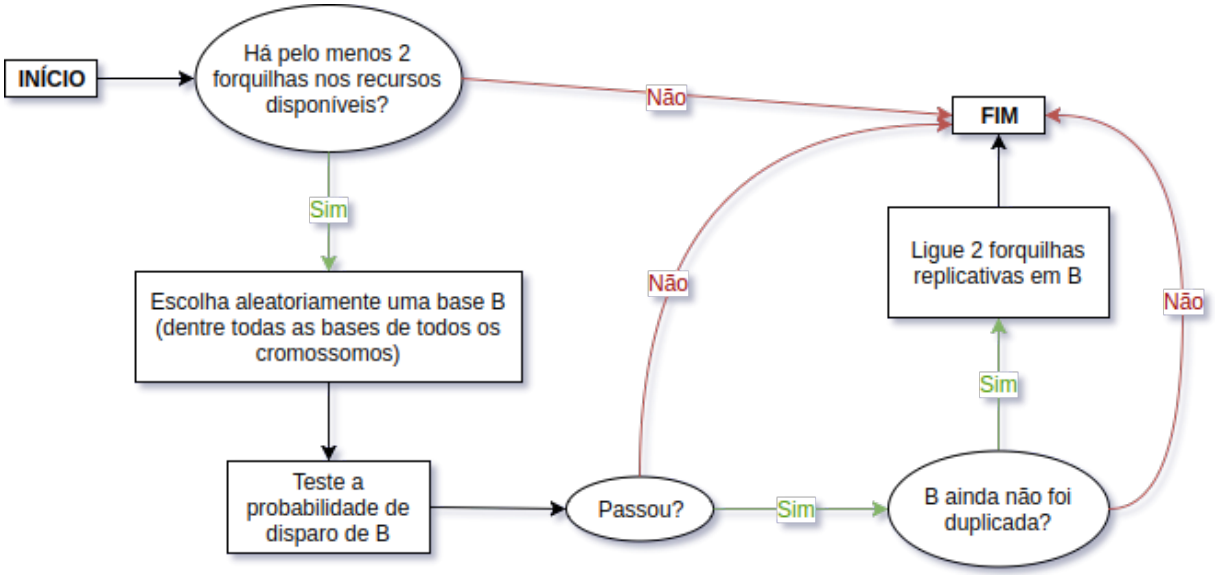


Figura 4.4: **Fluxograma do Processo de Ligação de Forquilhas Replicativas.**

Capítulo 5

Avaliação do Trabalho

5.1 Visão Geral

O desenvolvimento deste trabalho foi, sem dúvida, minha melhor realização acadêmica. Através dele tive a oportunidade de estudar conteúdos em áreas distantes das que eu esperava abordar quando estava iniciando este curso de computação, além de poder me inserir no âmbito da pesquisa, experienciando a dinâmica das carreiras fora do mercado de trabalho. Este trabalho também serviu como uma ótima oportunidade para aprimorar minhas habilidades em desenvolvimento de software, apresentação e divulgação de projetos, organização e realização de atividades interdisciplinares.

5.2 Histórico

A preparação para o desenvolvimento deste projeto começou em agosto de 2016, quando, por indicação de um colega do curso, conheci o Dr. Marcelo da Silva Reis e começamos a estudar o tema da replicação celular. Abordamos desde os pontos mais gerais até as especificidades que seriam importantes para o desenvolvimento de nosso modelo.

Começamos, então, o trabalho árduo de desenvolvimento deste projeto, intercalado frequentemente por momentos de estudo e aprofundamento na literatura, com enfoque tanto nos conceitos biológicos quanto na resolução dos problemas computacionais que surgiram ao longo do desenvolvimento.

No final de 2016, o trabalho foi apresentando com um pôster no IME-USP, como finalização da disciplina Apoio à Pesquisa. Já em outubro de 2017, outro pôster foi apresentado na *X-Meeting*, conferência internacional de bioinformática e, em novembro de 2017, mais uma apresentação foi feita na Reunião Científica Anual do Instituto Butantan. Todas essas apresentações reforçam a importância do trabalho e permitiram meu crescimento como aluno e pesquisador.

5.3 Desafios

Ao longo da criação do modelo surgiram inúmeros desafios, principalmente sobre conhecimentos biológicos dos quais eu tinha pouca ou nenhuma base no início do desenvolvimento. As várias exposições do trabalho ao público científico, incluindo toda a preparação até a eloquência durante as apresentações, também foram dificuldades vencidas durante esses últimos meses.

5.4 Agradecimentos

Agradeço às seguintes pessoas, cujo auxílio e suporte foram indispensáveis para o desenvolvimento deste projeto:

- Marcelo da Silva Reis, meu orientador e colega durante esses vários meses, responsável por me guiar, auxiliar e, claro, cobrar resultados dos quais ele sabia que eu era capaz, tendo em vista sempre o meu melhor.
- Marcelo Santos da Silva, colega do Instituto Butantan que contribuiu imensamente com as informações necessárias para o avanço e desenvolvimento do projeto.
- Gustavo Estrela de Matos, colega do IME-USP responsável por me indicar a este projeto e a quem estarei sempre agradecido.
- Duílio Felix Oliveira da Silva, ex-administrador da Rede Linux e estimado amigo que me ajudou repetidas vezes com várias dúvidas de programação.
- Todos meus colegados do Instituto Butantan e do IME-USP, cujo suporte foi essencial para meu sucesso acadêmico.
- Meus pais, Alcíone de Oliveira Santos Cayres e Miltom Rodrigues da Silva, cujo suporte emocional e financeiro me permitiu focar exclusivamente nos estudos.
- Vinícius Carlos Brando, meu parceiro que esteve do meu lado com grande atenção, mesmo quando era impossível retribuí-la.

Muito obrigado e espero que meus esforços neste trabalho só tenham nos aproximado e aumentado o apreço que temos uns pelos outros.

Bibliografia

[Asl+10] Martin Aslett, Cristina Aurrecoechea, 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*». Em: *Nucleic acids research* 38.suppl 1 (2010), pp. D457–D462.

[Cal+15] Simone Guedes Calderano, William C Drosopoulos, Marina Mônaco Quaresma, Catarina A Marques, Settapong Kosiyatrakul, Richard McCulloch, Carl L Schildkraut e Maria Carolina Elias. «Single molecule analysis of *Trypanosoma brucei* DNA replication dynamics». Em: *Nucleic acids research* (2015), gku1389.

[Gin+14] Yevgeniy Gindin, Manuel S Valenzuela, Mirit I Aladjem, Paul S Meltzer e Sven Bilke. «A chromatin structure-based model accurately predicts DNA replication timing in human cells». Em: *Molecular systems biology* 10.3 (2014), p. 722.

[HF+04] Christiane Hertz-Fowler, Chris S Peacock, Valerie Wood, Martin Aslett, Arnaud Kerhornou, Paul Mooney, Adrian Tivey, Matthew Berriman, Neil Hall, Kim Rutherford et al. «GeneDB: a resource for prokaryotic and eukaryotic organisms». Em: *Nucleic acids research* 32.suppl 1 (2004), pp. D339–D343.

[LP12] Yea-Lih Lin e Philippe Pasero. «Interference between DNA replication and transcription as a cause of genomic instability». Em: *Current genomics* 13.1 (2012), pp. 65–73.

[Mar+15] Catarina A Marques, Nicholas J Dickens, Daniel Paape, Samantha J Campbell e Richard McCulloch. «Genome-wide mapping reveals single-origin chromosome replication in *Leishmania*, a eukaryotic microbe». Em: *Genome biology* 16.1 (2015), p. 1.

[Shi+98] Katsuhiko Shirahige, Yuji Hori, Katsuya Shiraishi, Minoru Yamashita, Keiko Takahashi, Chikashi Obuse, Toshiki Tsurimoto e Hiroshi Yoshikawa. «Regulation of DNA-replication origins during cell-cycle progression». Em: *Nature* 395.6702 (1998), pp. 618–621.

[Sil+17] Marcelo Santos Silva, Paula Andrea Marin Muñoz, Hugo Aguirre Armelin e Maria Carolina Elias. «Differences in the Detection of BrdU/EdU Incorporation Assays Alter the Calculation for G1, S, and G2 Phases of the Cell Cycle in Trypanosomatids». Em: *Journal of Eukaryotic Microbiology* (2017).

[Tie+12] Calvin Tiengwe, Lucio Marcello, Helen Farr, Nicholas Dickens, Steven Kelly, Michal Swiderski, Diane Vaughan, Keith Gull, J David Barry, Stephen D Bell et al. «Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of Trypanosoma brucei». Em: *Cell reports* 2.1 (2012), pp. 185–197.