Tuning a predictive DNA replication programming computational model for Trypanosomatids

Bruno B. Scholl[1,2,3] **, Ligia C. Palma**[1,2] **,**
Victor S. Hariki[1,2] **, Maria Carolina Elias**[2] **,**
Marcelo S. Reis [2,3]

[1]Universidade de São Paulo (USP), São Paulo, Brazil

[2]CeTICS, Laboratório de Ciclo Celular, Instituto Butantan, São Paulo, Brazil

[3]Laboratório de Inteligência Artificial (Recod), Instituto de Computação, Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil

msreis@unicamp.br, bruno.scholl@usp.br

Abstract. *In this paper, we report the tuning of a predictive DNA replication programming computational model for both Trypanosoma brucei and Trypanosoma cruzi, unicellular protozoan endoparasites that cause African sleeping sickness and Chagas disease, respectively. This is a stochastic dynamic model for simulating the DNA replication process with concomitant constitutive transcription, enabling the analysis of the interactions between replication and transcription in these organisms. Using Optuna, an open-source hyperparameter optimizer, we explored almost 5,000 parameter combinations across both trainings of the models, each with up to 1,000 simulations averaged as the output, yielding two models that deviate by only 7.9% and 5.55% compared to experimental MFA-Seq data for T. brucei and T. cruzi, respectively, according to the SMAPE metric. The results also indicate a pattern in both models, in which the head-on collisions between replication and transcription machineries are co-localized with bases replicated earlier in the S-phase. This work paves the way for further in silico experimental exploration, aiming to unravel underlying mechanisms of the DNA replication programming in trypanosomatids.*

1. Introduction

Trypanosomatids are a family of unicellular protozoa parasites. Some of its members are the etiological agents of neglected tropical diseases (NTDs) that affect nearly 1 billion people and are more prominent in poorer countries (Feasey et al. 2010). Two members of this family with high biomedical importance are the *Trypanosoma brucei*, which causes a disease known as the African sleeping sickness, and *Trypanosoma cruzi*, the etiological agent of the Chagas disease. Trypanosomatids are eukaryotic, but their genes are organized in multigenetic units, that result in polycistronic mRNA. This, combined with the lack of transcription regulation, due to their post-transcriptional gene expression regulation, means that large DNA sections are transcribed constitutively, *i.e.*, during all cell cycle phases. Transcription, when occurring alongside replication, increases the chances of these machinery crossing paths while traveling in the same strand and potentially colliding with each other.

There are two types of collision: head-to-tail and head-on. The former is when replisome and RNA polymerase (RNAP) are moving in the same direction, at different speeds, and the latter is when they are moving toward each other. Head-on collisions are more troublesome since the replisome can stall, or even detach from the DNA, thus interrupting the ongoing replication of that DNA section and requiring another DNA replication origin firing (*i.e.*, attachment of another replisome machinery) to occur in order to complete the replication process (Tiengwe et al. 2012).

It is believed that trypanosomatids have limited replication origins (Tiengwe et al. 2012). Still, facing these challenges during replication, to maintain robustness and complete the replication in a timely manner to survive, more replication origins may need to be fired; for *T. brucei*, it has been found that more replication origins are needed than there are currently mapped (da Silva et al. 2019). Replication origins have been classified into two types: constitutive origins, which are almost always fired; and non-constitutive (facultative or dormant) origins, which fire only if the replication does not complete in time using the constitutive origins (da Silva et al. 2019).

Previously, we reported a computational model for DNA replication programming in *T. brucei* (da Silva et al. 2019). With that model, we were able to show evidence that DNA replication is robust in that organism despite the rising occurrence of replication-transcription conflicts, a theory validated through an RNA synthesis inhibition assay using α-amanitin (da Silva et al. 2019). Here, our goal was to extend that pioneer analysis through the tuning of predictive, machine learning-based models for both *T. cruzi* and *T. brucei*, aiming to predict the DNA replication programming of those organisms. This tuning incorporates species-specific genomic data, including detailed information on transcription sites and chromosome structures unique to each trypanosomatid. This species-specific approach allows for a more nuanced representation of the replication process in these parasites and can lead to significant insights.

The remainder of the paper is organized as follows: In Section 2, we present the most relevant related works found in the literature; In Section 3, we describe the biological datasets and other data used in the work, the replication model developed, and the workflow utilized to adjust the model; In Section 4, we discuss the results obtained from the model calibration, how it compares to experimental data from the real organism and what are the characteristics of the interactions observed between replication and transcription; In Section 5, we recapitulate the work, and present the future work that can be carried out to improve upon our work and answer more questions.

2. Related works

The literature review for this work has been conducted following the PRISMA statement guidelines (Page et al. 2021). We organized this review into two sections: one focusing on DNA replication computational models, in general, and one focusing on DNA replication studies in Trypanosomatids. This is to broaden the search and improve the number of related works, even if they don't fully match this line of work. For both sections, we defined the date range of publications from 2000 to 2024, as this field of DNA replication models does not have frequent publications. Also, for both sections, the ordering chosen was by the most recent publications first. The indexers used to search for literature were Scopus, ACM, and PubMed.

2.1. DNA Replication Computational Models

This section will discuss works of computational models of DNA replication in any organism. The keywords for the literature search were "DYNAMIC MODEL", "COMPUTATIONAL MODEL", and "DNA REPLICATION", organized in the following boolean expression query ("DYNAMIC MODEL" OR "COMPUTATIONAL MODEL") AND "DNA REPLICATION". This search returned 78 results from Scopus, 19 from ACM, and 44 from PubMed. Following the PRISMA, 5 resources were removed before screening as book sections, and 36 as duplicates. 109 records were screened via title and abstract. From those, 15 were selected for eligibility assessment under the criteria of describing or utilizing a computational model of the DNA replication dynamics. After the evaluation, 8 records were removed for not containing a DNA replication dynamics model. From the remaining records, we will discuss the 5 most recent articles.

In 2024, Ma *et al.* show, in a comprehensive review of cell cycle models, models of the DNA replication and replication initiation (Ma and Gurkan-Cavusoglu 2024). However, none model the transcription and its interactions during DNA replication, making them unsuitable for studying these phenomena.

In 2023, Gilbert *et al.* proposed a model of the 3D dynamics of the chromosome in genetically minimal bacteria, called JCVI-syn3A, at a high spatial resolution of 10 base pairs (bp) per particle (Gilbert et al. 2023). The model incorporates the dynamics of replication forks and could yield results in DNA replication dynamics, however, it lacks the functionality of accounting for transcription and replication-transcription interactions, making it ineligible for analyzing such characteristics.

In 2019, Yousefi *et al.* presented RepliSim, a numerical model to study the dynamics of DNA replication (Yousefi and Rowicka 2019). It simulates the movement of replication forks genome-wide, with a stochastic variance of travel speed for each replication fork. Their model has the advantage of simulating the DNA replication dynamics and the addition of stochastic fork speed, still, it is not suitable to study the interactions between replication and transcription, as it does not simulate transcription.

In 2014, Gindin *et. al.* submitted a chromatin structure-based model of the dynamic of DNA replication in human cells (Gindin et al. 2014). This model has been proven to effectively replicate the DNA replication timings of human cells, compared to experimental data. It also has the advantage of receiving a probability landscape for the probability of origin firing and the number of replication forks, making it easier to adapt to other organisms. Despite these advantages, this model does not account for transcription, rendering it unfit for studying the interactions between replication and transcription.

In 2014, Supady *et al.* proposed a deterministic model of DNA replication in *Saccharomyces cerevisiae* (Supady et al. 2013). This model simulates the dynamics of DNA replication and has a configurable variable replication fork rate, which could be adapted to other organisms. However, this model also does not account for transcription amidst the replication of the genome, and thus cannot be used to study the interactions and conflicts between replication and transcription.

2.2. Theoretical Studies on Interactions Between DNA Replication and Transcription in Trypanosomatids

In this section, works on models of DNA replication in any organism will be explored and discussed. The keywords for the literature search were "DYNAMIC MODEL", "COMPUTATIONAL MODEL", "STATISTICAL MODEL", "MATHEMATICAL MODEL", "DNA REPLICATION", "TRANSCRIPTION", "TRYPANOSOMA", and "TRYPANOSOMATID", assembled in the following boolean expression query ("DYNAMIC MODEL" OR "COMPUTATIONAL MODEL" OR "STATISTICAL MODEL" OR "MATHEMATICAL MODEL") AND "DNA REPLICATION" AND "TRANSCRIPTION" AND ("TRYPANOSOMA" OR "TRYPANOSOMATID"). This query aimed to find any works on theoretical studies of the conflicts between DNA replication and transcription in trypanosomatids. It yielded a single result joining all three indexers previously defined, so we added Springer as an additional indexer for this specific literature search. This resulted in another 21 results found with the same query. Following the procedures of the PRISMA statement, all 22 results were screened via title and abstract. From those, 6 results were selected for eligibility assessment, with the criteria of presenting a theoretical study on the DNA replication of trypanosomatids. Four of the articles did not fit under the criteria and were removed. These are the two articles that were eligible under the established criteria.

In 2019, da Silva *et al.* showed results from the stochastic dynamic model presented in this paper (da Silva et al. 2019). The model in this work is an earlier version of the ReDyMo model, that had not yet been trained for a specific organism. This model was designed to simulate the DNA replication dynamics in a stochastic fashion and its interactions with transcription machinery during replication.

In 2012, Stulemeijer *et al.* published a model of Dot1 enzymes in *T. brucei* (Stulemeijer et al. 2015). Although this model does not simulate the DNA replication or its dynamics, it models the effects of this enzyme in the replication. Despite having interactions with replication as output, this model is not suited for analyzing interactions between replication and transcription.

The scarcity of comparable studies in the literature underscores the novelty of this work, with only a handful of publications addressing similar aspects of DNA replication. This research, therefore, fills a significant gap in the field, offering unique insights into an underexplored area of trypanosomatid biology.

3. Methodology

In this section, we present the methodology of this investigation. The adopted workflow is summarized in Figure 1.

3.1. Biological datasets

The biological data used to calibrate the model consists of: MFA-Seq experimental data, which indicate, for each section of the genome, the rate of replication between two different moments in the cell cycle. For *T. brucei* the MFA-Seq data was acquired from (Tiengwe et al. 2012) and it compares the replication rate in the early-mid S phase and G2. For *T. cruzi* the MFA-Seq data is from (de Araujo et al. 2020) and compares early

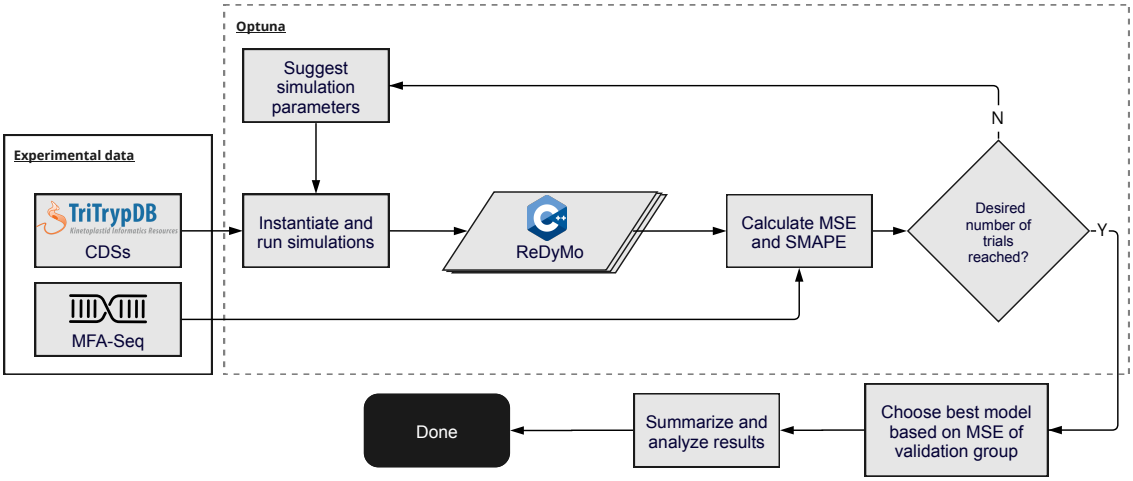

Figure 1. This diagram illustrates the workflow used for tuning the model. Optuna is used for model validation. At first, Optuna suggests a set of parameters for the model to run, and then, it instantiates the model using these parameters which loads CDS data of transcription regions; Lots of simulations run with the same set of parameters, and their results are aggregated to calculate MSE, MAE, and SMAPE errors comparing the output with experimental data. After a given number of trials, the best model is chosen according to its error metrics on the validation data set. The chosen model is then simulated and analyzed.

S phase and G2/M; coding sequences (CDS) define sections of the genome where transcription occurs and the direction that the RNAP machinery travels to transcribe it; the number of chromosomes and their sizes. CDS and chromosome data, of both organisms, are available in the TryTripDB repository (Amos et al. 2021). For *T. cruzi*, the specific strain trained was the *Trypanosoma cruzi CL Brener Esmeraldo-like* and the CDSs used were version 53 found at TritrypDB. For *T. brucei*, the specific strain was *Trypanosoma brucei brucei TREU 927* and version 34 of the CDSs found in TritrypDB.

3.2. Development of a DNA replication simulator

ReDyMo (Replication Dynamics Model) is a stochastic dynamic model of DNA replication programming in trypanosomatids. It is written in C++ 14, using the object-oriented programming paradigm. The main difference between this and other DNA replication models is the simulation of constitutive transcription during the replication process. This allows us to analyze the impact of replication-transcription interactions.

The inputs of the model are the number of chromosomes, the number of bases in each chromosome, and the position, length, and direction of polycistronic expression regions, also referred to as transcription regions. Other model parameters are the number of available replisomes, which are shared between all chromosomes of an organism, and the period of transcription, which defines after how many iterations a new RNAP will attach to each transcription region. Other simulation parameters can be specified, configuring the simulation environment, such as the number of CPU cores to use, the timeout as the maximum number of iterations, among others.

During the model simulation, a base is selected randomly, following a uniform distribution, across the bases of all chromosomes. The selected base might activate, fol-

```
while not genome_is_replicated():
    for fork in attached_rep_forks:
        advance_fork(fork)
        chromosome[fork.base] = current_iteration
    check_replication_transcription_conflicts(attached_rep_forks)
    detach_stalled_forks()
    for i in number_free_rep_forks:
        base = random_base(chromosome_list)
        if not replicated(base):
            if will_activate_origin_on_base(base, mfa_seq):
                attach_rep_forks(base)
```

Listing 3.1. Psudocode of the simulator's main loop.

lowing a probability landscape derived from the MFA-Seq. This means that any base has a chance of activating, but bases near, or at, replication origins have a greater chance of activating. If the base is activated, a pair of replisomes is attached to the genome at the selected base, traveling in opposite directions and marking the number of the current iteration when each base was replicated. Every P (the `period` parameter) iterations, a new RNAP is attached to each transcription region of the genome, forming a carousel of evenly spaced RNAPs. This is a simplification compared to the real world, but, since the transcription is constitutive, we assume that the transcription is happening at all times and has a configurable periodicity. In the case of a head-on collision, the fork stalls and detaches from the genome, and another origin has to be fired to replicate the remaining DNA bases. The head-to-tail collisions are ignored and both the replisome and the RNAP stay attached. The listing 3.1 has a simplified pseudocode of the main loop of the model. It highlights all the actions that take place at each step of the simulation: the movement of the forks attached to the DNA, the collision calculation, the detachment of stalled forks due to head-on collisions, and, finally, the activation of new replication origins and attachment of corresponding replication forks.

The simulation output is the relative time in the simulation when each base was replicated. This output is then compressed using a custom compressor that stores the start and end of sections of uninterrupted replication, which reduces the output file by ten-fold (Hariki 2021). This model is available under the GNU GPL v3.0 at github.com/Dynamic-Systems-Biology/ReDyMo-CPP.

3.3. Model selection and tuning

To train the models, it is needed to fine-tune the parameters to the ones that make each model output the most accurate when compared with the respective real organism. This can be achieved by exploring the space of available parameter values and comparing the yielded simulated output (e.g., synthetic MFA-seq data) with the real data. The total number of combinations of all the different parameter values is quite large, especially considering they are integer values. If one were to employ a grid search, it would take a very long time and lots of computational resources. Therefore, to save time and resources, we opted to use Optuna (Akiba et al. 2019), an open-source hyperparameter optimizer, that automates the search. Optuna is very flexible, so it only requires to know what are

the parameters that configure the model and how to calculate an error from the model output. For this experiment, we only needed to optimize two parameters of the model: the number of available replisomes and the number of iterations between each transcription machinery deployment. But, even though they are only two parameters, their values are integers, making the parameter space very large. To make the parameter search space smaller, we limited the possible values for each parameter as follows:

- $2 < \texttt{replication\_forks} < 10,000$, in increments of 2;
- $0 < \texttt{transcription\_period} < 1,000,000$, where 0 means no transcription.

Three error functions were used to optimize the parameter values, comparing the synthetic MFA-Seq results from the model and real MFA-Seq data of the organism; this was accomplished by calculating the difference between the average time of replication of each base for a chromosome, compared with the MFA-Seq for that same chromosome. The metrics calculated are the mean squared error (MSE), the mean absolute error (MAE), and the symmetric mean absolute percentage error (SMAPE); this latter yields a value that is more easily interpreted. All three error functions were used by Optuna to calibrate the model in a multi-objective optimization.

To track the parameter suggestions and optimization of the parameters by Optuna, the chromosomes were separated into groups. For *T. cruzi*, chromosomes were split into training, valitation and test sets, while for *T. brucei* chromosomes were split into training and validation only, since there were just a few chromosomes (11 chromosomes in comparison to 41 ones in *T. cruzi*). Moreover, the division of the DNA content into each group was different for each organism. For *T. brucei*, the training group contained chromosomes that accounted for around 80% of the genome bases, and the validation group contained the remainder of chromosomes. For *T. cruzi*, the training group was comprised of around 70% of the genome bases, the validation group, around 20% of the bases, and the test group, close to 10% of the bases.

Each error metric was calculated for each of the chromosome groups. In each trial, the model run 500 times for the given set of parameters for *T. brucei* and 1,000 times for *T. cruzi*, and the results from these runs were then normalized and aggregated to form an average replication time per base for each chromosome. Then the resulting average values are compared, using the error functions, with MFA-Seq data that has been normalized and linearly interpolated to fill the gaps in the data. These error values are used in different steps, depending on the chromosome group. The error from the trainng group is fed into Optuna to directly guide the parameter selection of the next iterations; the error from the validation group is used to select the best model. This tries to mitigate any overfitting of the model with the training data; the error from the test group is used to assess the model accuracy in a completely independent dataset from both training and model selection. This is again to minimize biases and overfitting.

3.4. Computational resources

The model optimization was run across five servers, each with at least 60 CPU cores, and 50 GB of RAM. Each trial took, on average, 25 minutes to run, and 1,300 trials were executed for *T. brucei* and around 3,600 trials for *T. cruzi*. Optuna offers the option to use a relational database as the backend to automatically coordinate multiple trials at once and save historical data about each trial, including user defined fields, such as custom metrics. Optuna also has a dashboard for analyzing the training with graphs and visualizations.

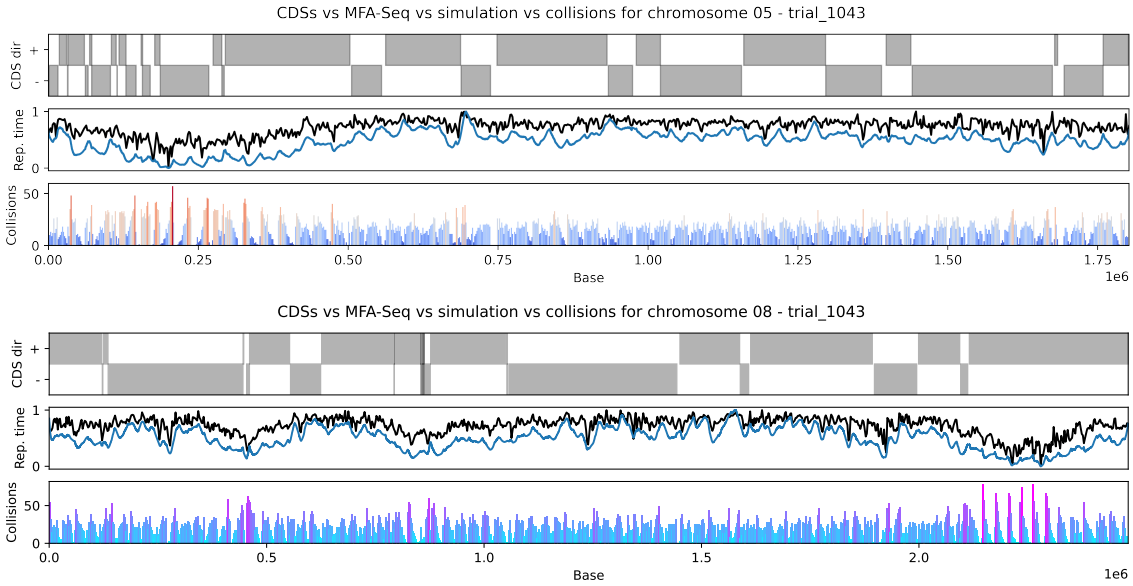

Figure 2. Graphs of CDSs, MFA-Seq, simulated replication time, and head-on collisions for chromosomes 5 and 8 of *T.* brucei versus the chromosome base in Mb, using the best parameter set. The top graph shows CDSs and their direction. The middle graph is a plot of MFA-Seq, in black, versus the simulated replication times, in blue; the lower the value, the sooner the base was replicated in the S-phase. The bottom graph is a histogram of head-on collisions across all 1,300 simulations.

4. Results

For the *T. brucei*, we conducted 1,300 trials with various parameter combinations and identified the best model, using the score for the validation chromosome group, whose parameters consist of 966 replisomes and a transcription period of 29,235 iterations. This absolute number of replisomes could be confirmed via wet lab experiments. This model reached a SMAPE of 7.45% in the training part of the dataset and 7.93% in the validation portion and, overall, Optuna did not show signs of overfitting in the model.

For *T. cruzi*, 3,600 trials were conducted, exploring the parameter space. The best model, selected by the score with the validation chromosome group, has the parameter combination of 1,034 replisomes and a transcription period of 234,250 iterations. This model has a SMAPE of 5.82% in the training group, 5.38% in the validation group and 5.55% in the test group. The test group is a brand new data set for the model, that is not related to the training of the parameters or the selection of the model, so a SMAPE of 5.55% in this group shows that the model is very well adjusted to the organism, without signs of overfitting, as all three groups have values close together.

In Figure 2, we show plots of the MFA-Seq data versus the simulated replication times (i.e., the "synthetic" MFA-Seq) for *T. brucei*, where we can see that the simulated times are close to the experimental data and, even when the lines do not match exactly, their profiles look very similar throughout the chromosome. This especially holds near the earliest replicated bases, in which both simulation and experimental data seem to agree with each other.

For *T. cruzi*, the same graphs are shown in Figure 3. All chromosomes in the

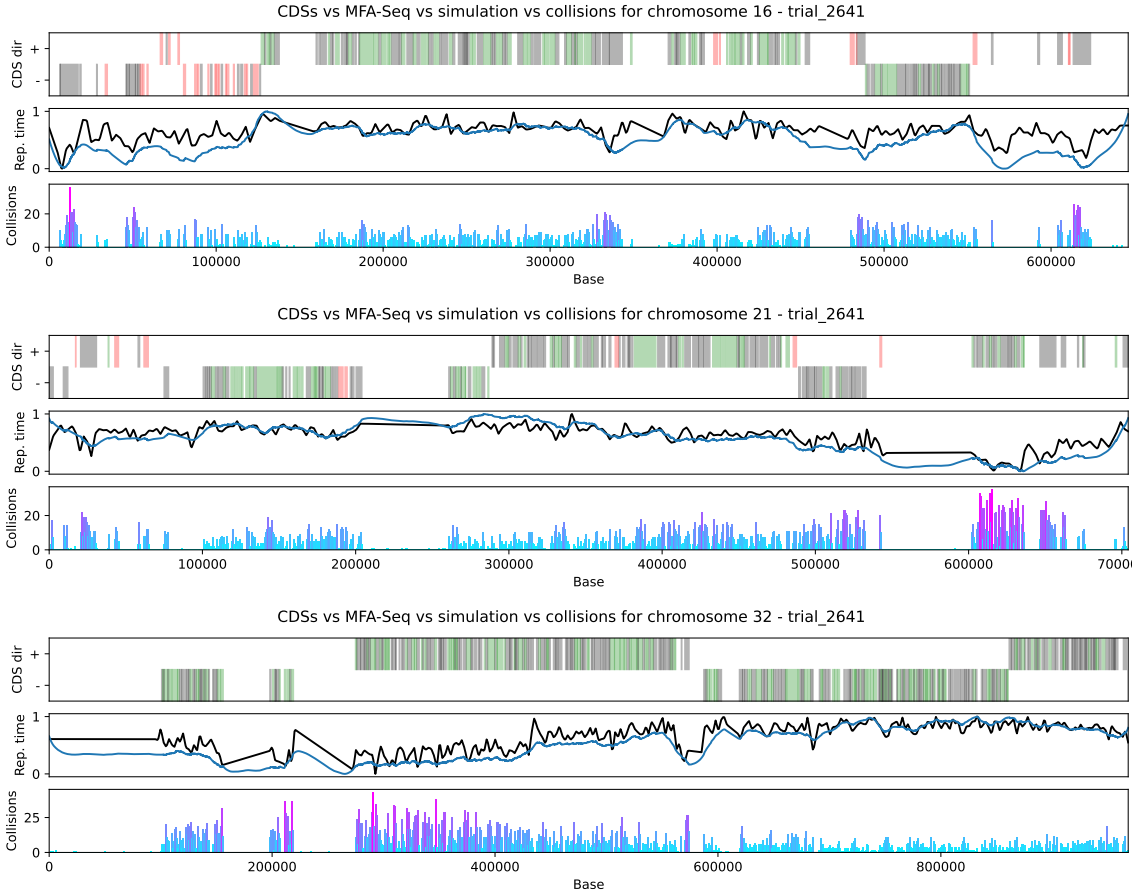

Figure 3. Graphs of CDSs, MFA-Seq, simulated replication time, and head-on collisions for chromosomes 16, 21, and 32 of *T. cruzi*, of the best model. The top graph shows CDSs, their direction, and classification as green for conserved genes, red for what is called disruptive compartment, composed of MASPs, Mucins, and trans-sialidases (Berná et al. 2018). The middle graph is a plot of MFA-Seq, in black, versus the simulated replication times, in blue; the lower the value, the earlier the base was replicated in the S-phase. The bottom graph is a histogram of head-on collisions across all 3,600 simulations.

image are from the test group, which shows how well the model has calibrated for this organism. The same proximity between MFA-Seq and simulated times can be observed with an even greater effect. Not only the profile follows the experimental data, but the actual values are much closer to the experimental data across the whole chromosome.

This difference of similarity between simulated times and MFA-Seq across the two models agrees with the difference in the SMAPE values of the models. *T. brucei* has a higher SMAPE, compared to *T. cruzi*, which is enough to be visible in the graphs.

Looking more closely at the collision histogram, one pattern that stands out is that areas surrounding bases that were replicated earlier have increased collision counts. As seen in Figure 2, close to the 0.5 Mb mark, close to the 1.0 Mb mark and right past the 2.0 Mb mark. This pattern is observed in all *T. brucei* chromosomes, except in chromosome 9, which lacks polycistronic regions near the bases with an earlier replication time. This correlation can also be observed for *T. cruzi* in Figure 3. Regions in which the bases

replicate earlier in the S-Phase have a greater incidence of head-on collisions between replisomes and RNAP.

Moreover, other locations hypothesized to have a greater replication-transcription conflict incidence did not show such evidence. Locations such as where multigenic regions, with opposing coding strands, are close together, or the starts and ends of polycistronic coding regions, did not present more head-on collisions than other locations of the chromosomes in both *T. brucei* and *T. cruzi*.

Regarding model training, both models had good results approximating the replication times of the organisms. The differences seen in SMAPE values and graphs could be attributed to the differences in the training procedures. For *T. brucei*, only 500 simulations for each parameter combination had been run, compared to 1,000 simulations run for each parameter combination for *T. cruzi*. Also, the model for *T. cruzi* tested 3,600 parameter combinations, compared to *T. brucei*'s model which trained for 1,600 iterations.

5. Final remarks

In this work, we used genomic data as input to a model that was tuned by the adjustment of its parameters, obtaining a predictive model that simulates the replication time of chromosome bases close to MFA-Seq experimental data for two trypanosomatids: *Trypanosoma brucei* and *Trypanosoma cruzi*. We were able to get models that diverged, according to the SMAPE metric, 7.9% from the observed data for the *T. brucei* model (in the validation set), and 5.55% for *T. cruzi* model (in the test set).

Moreover, we analyzed the interaction between transcription and replication machinery in the models, specifically the head-on collisions; we found that regions at which the bases are replicated earlier in the S-phase tend to have a greater incidence of head-on collisions. Nearly all chromosomes of both organisms presented this pattern, except *T. brucei*'s chromosome 9, which doesn't have transcription near the bases replicated at the beginning. Additionally, contrary to initially believed, certain regions of the genome, such as areas with changes in transcription direction, have not been found to have an increased incidence of head-on collisions.

As the next steps of this study, the model's accuracy could be enhanced by inputting more data, such as ChIP-Seq experimental data, that yields DNA replication origins. In addition, with the objective of achieving an even lower SMAPE, the models could be trained for more trials and the parameters could be explored with values outside of the ranges used in this study. Finally, adaptation of the model to other members of the Trypanosomatida family could also be attempted.

Acknowledgements

This work was supported by CNPq and by grants 13/07467-1, 19/21619-5, and 20/00694-6, São Paulo Research Foundation (FAPESP).

References

[Akiba et al. 2019] Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

[Amos et al. 2021] Amos, B. et al. (2021). VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Research*, 50(D1):D898–D911.

[Berná et al. 2018] Berná, L., Rodriguez, M., Chiribao, M. L., Parodi-Talice, A., Pita, S., Rijo, G., Alvarez-Valin, F., and Robello, C. (2018). Expanding an expanded genome: long-read sequencing of trypanosoma cruzi. *Microbial Genomics*, 4(5).

[da Silva et al. 2019] da Silva, M. S. et al. (2019). Transcription activity contributes to the firing of non-constitutive origins in african trypanosomes helping to maintain robustness in s-phase duration. *Scientific Reports*, 9(1):18512.

[de Araujo et al. 2020] de Araujo, C. B., da Cunha, J. P. C., Inada, D. T., Damasceno, J., Lima, A. R. J., Hiraiwa, P., Marques, C., Gonçalves, E., Nishiyama-Junior, M. Y., McCulloch, R., and Elias, M. C. (2020). Replication origin location might contribute to genetic variability in trypanosoma cruzi. *BMC Genomics*, 21(1):414.

[Feasey et al. 2010] Feasey, N., Wansbrough-Jones, M., Mabey, D. C., and Solomon, A. W. (2010). Neglected tropical diseases. *British medical bulletin*, 93(1):179–200.

[Gilbert et al. 2023] Gilbert, B. R., Thornburg, Z. R., Brier, T. A., Stevens, J. A., Grünewald, F., Stone, J. E., Marrink, S. J., and Luthey-Schulten, Z. (2023). Dynamics of chromosome organization in a minimal bacterial cell. *Frontiers in Cell and Developmental Biology*, 11:1214962.

[Gindin et al. 2014] Gindin, Y., Valenzuela, M. S., Aladjem, M. I., Meltzer, P. S., and Bilke, S. (2014). A chromatin structure-based model accurately predicts dna replication timing in human cells. *Molecular Systems Biology*, 10(3):722.

[Hariki 2021] Hariki, V. S. (2021). Decyphering the dna replication dynamics of trypanosoma cruzi through computational modeling.

[Ma and Gurkan-Cavusoglu 2024] Ma, C. and Gurkan-Cavusoglu, E. (2024). A comprehensive review of computational cell cycle models in guiding cancer treatment strategies. *npj Systems Biology and Applications*, 10(1):71.

[Page et al. 2021] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., and Moher, D. (2021). The prisma 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372.

[Stulemeijer et al. 2015] Stulemeijer, I. J. E., De Vos, D., van Harten, K., Joshi, O. K., Blomberg, O., van Welsem, T., Terweij, M., Vlaming, H., de Graaf, E. L., Altelaar, A. F. M., Bakker, B. M., and van Leeuwen, F. (2015). Dot1 histone methyltransferases share a distributive mechanism but have highly diverged catalytic properties. *Scientific Reports*, 5(1):9824.

[Supady et al. 2013] Supady, A., Klipp, E., and Barberis, M. (2013). A variable fork rate affects timing of origin firing and s phase dynamics in saccharomyces cerevisiae. *Journal of Biotechnology*, 168(2):174–184.

[Tiengwe et al. 2012] Tiengwe, C. et al. (2012). Genome-wide analysis reveals extensive functional interaction between dna replication initiation and transcription in the genome of trypanosoma brucei. *Cell reports*, 2(1):185–197.

[Yousefi and Rowicka 2019] Yousefi, R. and Rowicka, M. (2019). Stochasticity of replication forks' speeds plays a key role in the dynamics of dna replication. *PLOS Computational Biology*, 15(12):e1007519.