PhD proposal – 2025

Stochastic modelling and estimation for the distribution of lengthening and abrupt shortening of ALT cells in yeasts

Keywords: applied probability, stochastic modelling of population processes, branching processes, statistical modeling for biology, estimation, telomere length dynamics

Biological context

Telomeres are nucleoprotein structures located at the ends of chromosomes, which they protect from degradation. During the cell division, the DNA is not entirely replicated leading to a loss of telomere sequences. Without any mechanism of telomere lengthening, telomeres progressively shorten until they reach a critical length (roughly at 70 bp in yeasts). Below this critical threshold, shortened telomeres trigger a permanent cell division cycle arrest, leading to a replicative senescence. This phenomenon is known as the end replication problem.

In the yeast *Saccharomyces cerevisiae*, telomere length homeostasis is the result of a balance between the action of the enzyme *telomerase reverse transcriptase* (TERT), which adds telomere sequences on short telomeres, and losses of telomere sequences due to the replication of DNA ends at cell divisions. As a result, telomere length varies from cell to cell and from telomere to telomere within a given cell, but stay of the order of 300 bp. When TERT activity is repressed, telomeres progressively shorten following the end replication problem until the replicative senescence. However, most often, in cultures of TERT-inactivated yeasts, rare "survivors" (roughly 1 among 100 000 individuals) escape senescence thanks to other telomerase-independent telomere maintenance mechanisms (called ALT for Alternative Lengthening of Telomeres), based in particular on homologous recombination.

ALT cells are characterized by very heterogeneous distribution of telomere lengths up to 10 kb. In ALT cells, telomeres are confronted with the end replication problem and therefore shorten with each cycle of cell division. Furthermore, as with natural telomeres, replication of ALT telomeres is a challenge to the cell's replication machinery and therefore a source of stochastic replicative damage leading to abrupt shortening in the absence of telomerase. Finally, like natural telomeres, ALT telomeres are considered to be the preferential target of oxidative stress, which could be another source of abrupt telomere attrition.

Project description

The PhD project aims to develop a comprehensive model at the level of a population of lengthening and shortening of telomeres in ALT yeasts and to validate it on data of culture, giving the time evolution of the distribution of telomere lengths in the population.

Tasks

The first task of the PhD student will be to get familiar with the various types of ALT telomere data of the project. Our biologist collaborator, Marie-Noëlle Simon (CRCM,

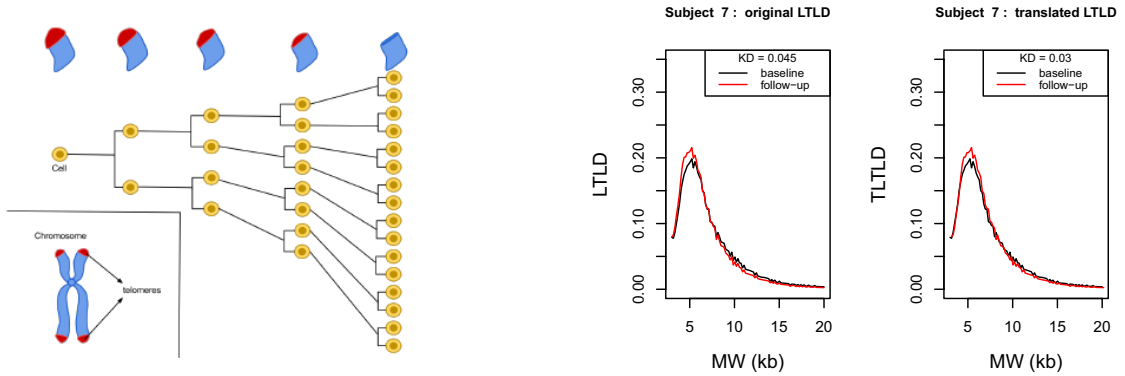

Figure 1: (a) The end replication problem. (b) Examples of distributions of telomere lengths in human, illustrating the notion of *telomere signature* [8].

Aix-Marseille University), will provide us with data of telomere length distributions at the population level, for all the telomeres or for a single telomere selected using a specific probe. These are temporal data, obtained from samples regularly collected in cultures. Due to the exponential growth of the population, subpopulations are selected on a daily or bi-daily basis either on Petri dishes with successive transplantations of a single colony, or on liquid experiments with successive dilutions. Most of the data are collected using the *Southern blot*, but we will also study data obtained with other types of methods, including the *TeSLA* method, which provides measurements of few telomeres selected in the population.

Then, the PhD student will design a stochastic model of telomere evolution at the level of a single lineage, including mechanisms of abrupt telomere shortening and lengthening. The parameters of this model will be estimated from the data of evolution of the length of single telomeres. We can take our inspiration from the work [1,2] for the modeling part and from [3] for the estimation part.

The third step of the project is to construct a stochastic model of telomere evolution at the level of the population, including the individual mechanisms developed earlier. This will take the form of a branching process, where the progeny of different individuals follow independent processes, structured by the vector of individual telomere lengths. Such infinite-dimensional branching processes will be represented as measure-valued Markov processes following the formalism of [4,5]. In particular, the student will prove that the many-to-one formula holds true for this process, as in [6,7] where the authors study telomere length dynamics, and provide numerical simulations in order to illustrate its main property.

The fourth step of the project will consist in building from the previous model another model for the successive transplantations or dilutions experiments described above, able to account for the collected data of telomere length distribution at the level of the population. This model will include resources dynamics, hence breaking the branching property. The goal is here to estimate the main parameters of the model, namely the rates of abrupt attrition and lengthening of telomeres and their range and the rate of progressive telomere shortening due to the end replication problem, from the data. Due to the complexity of the model, it is particularly difficult to design model-specific methods and we will thus use classical optimisation methods, such as least-square estimates, combined with expert opinion on the biological relevance of the obtained parameters values. One of the biological questions we would like to answer is the following: is the rate of progressive telomere shortening due to the end replication problem the same for ALT yeasts than for normal

yeasts (with active telomerase)?

Other questions can also be considered during the PhD thesis, depending on the progress and the interest of the student, such as modeling questions regarding the proliferative capacity and the competitivity of ALT cells in populations with multiple non-attrition strategies.

PhD thesis context

The PhD thesis will take place in the Probability and Statistics team of the Institut Élie Cartan de Lorraine (IECL) and in the SIMBA (Statistical Inference and Modeling for Biological Applications) team of Inria Nancy. The PhD student will be involved in discussions with Marie-Noëlle Simon (CRCM, Aix-Marseille University) on the biological and data aspects of the project. During the PhD thesis, the student will have the opportunity to discover the world of mathematical research through the life of a dynamic mathematics laboratory, and to attend seminars and working groups in probability and statistics.

Funding

The PhD thesis is funded by PEPR Maths VivES.

Skills

The candidate should have skills in statistics and/or stochastic modeling. R, Python or Matlab programming skills are also required. An affinity or experience with biological applications will be highly appreciated.

Supervision

The PhD thesis will be supervised by Nicolas Champagnat, Coralie Fritsch (IECL and INRIA Nancy) and Denis Villemonais (University of Strasbourg).

Contacts

nicolas.champagnat@inria.fr, coralie.fritsch@inria.fr, denis.villemonais@unistra.fr

Application

The application file should be submitted on the Inria job offer web site.

Bibliography

[1] Thibault Bourgeron, Zhou Xu, Marie Doumic, Maria Teresa Teixeira, The asymmetry of telomere replication contributes to replicative senescence heterogeneity, *Scientific Reports*, vol 5, 2015.

[2] Hugo Martin, Marie Doumic, Maria Teresa Teixeira, Zhou Xu, Telomere shortening causes distinct cell division regimes during replicative senescence in Saccharomyces cerevisiae, *Cell & Bioscience*, vol. 11, 2021.

[3] Christoph Zimmer and Sven Sahle. Comparison of approaches for parameter estimation on stochastic models: Generic least squares versus specialized approaches. *Computational Biology and Chemistry*, vol. 61, 2016.

[4] Vincent Bansaye and Sylvie Méléard. Stochastic Models for Structured Populations, Scaling Limits and Long Time Behavior, Springer, 2015.

[5] Emma Horton and Andreas Kyprianou. Stochastic Neutron Transport And Non-Local Branching Markov Processes, Springer, 2023.

[6] Athanase Benetos, Coralie Fritsch, Emma Horton, Lionel Lenotre, Simon Toupance, Denis Villemonais. *Stochastic branching models for the telomeres dynamics in a model including telomerase activity*, arXiv:2407.11453, 2024.

[7] Jules Olayé and Milica Tomasevic. *Long-time behaviour of a multidimensional age-dependent branching process with a singular jump kernel*, arXiv:2408.02476, 2024.

[8] Simon Toupance, Denis Villemonais, Daphné Germain, Anne Gegout-Petit, Anne and Eliane Albuisson, Athanase Benetos. The individual's signature of telomere length distribution, *Scientific reports*, vol 9(1), 2019.