Mutational signatures of different evolutionary mechanisms

Biomathematics

Nastassia Bonetti

May-August 2024

Document confidential

Univeristy : Imperial College - Department of Mathematics
Address : Exhibition Rd, South Kensington, London SW7 2BX, United Kingdom
ENSTA Paris Advisor : Laure Giovangigli
University Advisors : Prof. Nick Jones and Dr. Ferdinando Insalata

Disclosure agreement

This document is confidential. It may not be communicated outside of school in hard copy but also broadcast in electronic format.

Acknowledgements

I would like to thank my supervisor, Dr Ferdinando Insalata, and Prof Nick Jones for welcoming me to their mathematics department at Imperial College to work with their team on the study of muscular ageing in humans. During this placement, which plunged me into the world of biomathematics, I was able to understand its importance and complexity. What is more, I was able to work independently on a project with the confidence that Mr Insalata has in the members of his team.

Finally, I would like to thank all the researchers in the team who made me feel so welcome and helped me throughout my placement, and in particular Cecilia Fruet, a former trainee who worked on the same subject four years ago. She gave me support and knowledge even before the start of the course. I was able to work in excellent conditions, in a healthy and friendly environment.

Abstract

ENGLISH : Stochastic modelling of the spatio-temporal evolution of the mitochondrial genome in muscles and neurons.

Population genetics provides the framework for understanding changes in genetic composition over time. This study focuses on mitochondria, the energy producers in cells, which have their own DNA (mtDNA) that is maternally inherited and prone to frequent mutations. These mutations are associated with age-related diseases such as sarcopenia. Although several models have been proposed to explain these mutations, they often fail to match experimental results. To address this, stochastic models, which incorporate probabilistic elements, are crucial for capturing the inherent noise in biological systems and for understanding complex genetic dynamics. The key question addressed in this project is how to distinguish between different types of stochastic models. By employing the site frequency spectrum (SFS) as a statistical tool, we successfully demonstrate the ability to distinguish between three models, which are, Replicative Advantage, Stochastic Survival of the Densest and Stochastic Survival of the Slowest. This represents the main achievement of our study.

Keywords : Stochastic modeling, population genetics, mitochondrial genome, mutation, migration.

FRANÇAIS : Modélisation stochastique de l'évolution spatio-temporelle du génome mitochondrial dans les muscles et les neurones.

La génétique des populations fournit le cadre permettant de comprendre les changements dans la composition génétique au fil du temps. Cette étude se concentre sur les mitochondries, les producteurs d'énergie dans les cellules, qui ont leur propre ADN (ADNmt) hérité de la mère et sujet à de fréquentes mutations. Ces mutations sont associées à des maladies liées à l'âge telles que la sarcopénie. Bien que plusieurs modèles aient été proposés pour expliquer ces mutations, ils ne correspondent souvent pas aux résultats expérimentaux. Pour y remédier, les modèles stochastiques, qui intègrent des éléments probabilistes, sont essentiels pour capturer le bruit inhérent aux systèmes biologiques et pour comprendre les dynamiques génétiques complexes. La question clé abordée dans ce projet est de savoir comment distinguer les différents types de modèles stochastiques. En utilisant le spectre de fréquence de site (SFS) comme outil statistique, nous démontrons avec succès la capacité de distinguer trois modèles , qui sont l'avantage réplicatif, la survie stochastique du plus dense et la survie stochastique du plus lent. Il s'agit là de la principale réalisation de notre étude.

Mots-clés : Modélisation stochastique, génétique des populations, génome mitochondrial, mutation, migration.

Table of Contents

List of Figures

List of Tables

1 Introduction

Muscle ageing and mitochondrial dynamics are key areas of research due to their profound implications for cellular metabolism and health. Similar to the nucleus in cells, mitochondria also possess DNA, usually referred to as mtDNA. The birth-death process of mitochondria is independent of cellular dynamics [1], and by extension, so is the replication and degradation of mtDNA. In direct contrast to nuclear DNA, mtDNA has a high mutation rate.

A class of mtDNA mutations that occur in humans is long deletion, where a section of the mtDNA is deleted. In addition, mutations in mitochondria are associated with ageing [2]. Understanding their behavior is essential for unraveling the mechanisms underlying cellular metabolism and the development of mitochondrial diseases [3].

Mathematical and stochastic models play a crucial role in elucidating biological processes such as random genetic drift, deletions, etc. Stochastic models capture the inherent noise in biological systems, allowing for a deeper understanding of how genetic composition evolves over space and time. I needed to become proficient in these mathematical approaches to accurately model the complexities of mitochondrial dynamics.

This study investigates the interplay between normal mitochondria and mitochondrial mutants using both deterministic and stochastic mathematical models. Deterministic models, formulated through ordinary differential equations (ODEs), provide insights into how mitochondrial populations evolve over time under specific conditions. However, we will see how the random occurrence and proliferation of mitochondrial mutants is a more complex phenomenon, that is best captured by stochastic models.

The competition between normal mitochondria and mutants can be understood in terms of a recently introduced evolutionary mechanism, by the stochastic survival of the densest or also named SSD. This phenomenon highlights the importance of considering both deterministic and stochastic factors in understanding mitochondrial dynamics. By integrating these models, we aim to offer a comprehensive analysis of mitochondrial behavior, an essential first step for developing potential therapeutic interventions for mitochondrial diseases.

We simulate a population of mitochondrial DNA (mtDNA), consisting of wildtype (w) and mutant (m) types, controlled by the nucleus. We are particularly interested in heteroplasmy (h), the proportion of the population that is mutant. Our focus is on understanding how an existing mutation can expand clonally to high heteroplasmy levels. In our neutral model in SSD, both wildtypes and mutants have the same constant degradation rate (μ) and replication rate (λ), defined by :

$$\lambda(w,m) = \mu + c(N_{ss} - w - \delta m)$$

Here, c, N_{ss}, and δ are parameters. The replication rate is adjusted based on the difference between the current population size ($w + \delta m$) and a target population size (N_{ss}). Importantly, mutants contribute less to the population than wildtypes, indicated by the parameter $0 \leq \delta < 1$. The parameter c measures the strength of the nucleus's control over replication.

1. Patrick F. Chinnery and David C. Samuels, *Relaxed Replication of mtDNA : A Model with Implications for the Expression of Disease*

2. A. Hiona and C. Leeuwenburgh, *The role of mitochondrial dna mutations in aging and sarcopenia : implications for the mitochondrial vicious cycle theory of aging.*

3. A. D. de Grey, *A proposed refinement of the mitochondrial free radical theory of aging.*

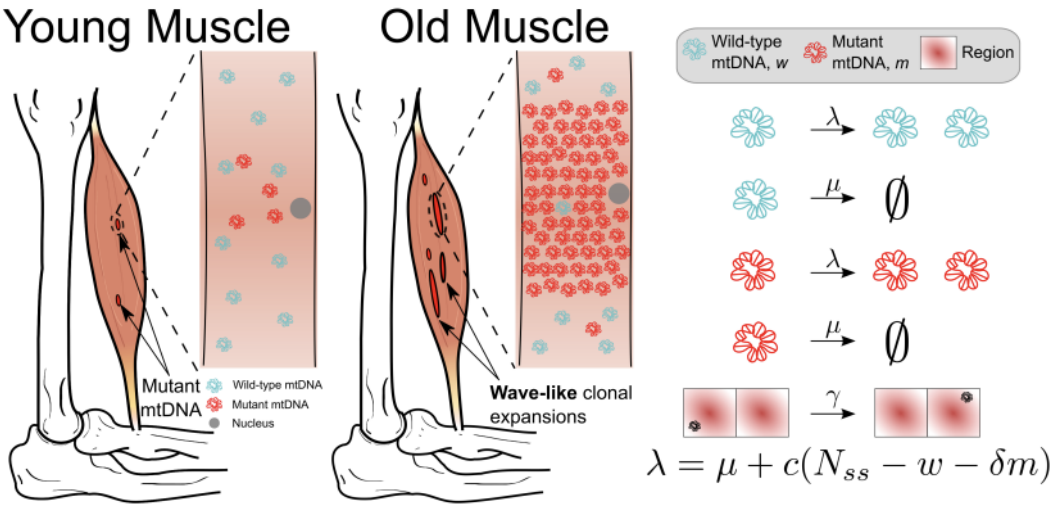


FIGURE 1 – Model applied to muscle fibres. *Source : Stochastic survival of the densest and mitochondrial DNA clonal expansion in ageing,2022, F.Insalata*

The first aim of the whole research project was to present a mechanistic model for the clonal expansion of mtDNA deletions, without delving into the exact molecular details that lead to the higher density of mutants (SSD). Then, I used an other model, termed Stochastic Survival of the Slowest (SSS), in which mutants exhibit slower birth and death rates, so that I can compare it to the SSD model and the Replicative Advantage (RA) model,where mutants have a higher replication rate, in practice and in a quantitative way..

Given the difficulties in directly observing mtDNA mutation dynamics due to destructive measurement processes and long timescales, distinguishing between different models becomes a significant challenge. Longitudinal studies, which involve repeated observations of the same subjects over time, are impractical, and cross-sectional studies, which compare different subjects at a single point in time, introduce variability due to genetic, environmental, and lifestyle differences. As a final step, I focused on point mutations, another class of mtDNA mutations, and utilized the site frequency spectrum (SFS) to distinguish between models such as Replicative Advantage (RA), SSD, and SSS without the need for longitudinal or cross-sectional data. By leveraging advanced high-throughput DNA sequencing technologies, I demonstrated that SFS, which can be obtained from a single time point, is a powerful tool for differentiating between these models, offering a significant contribution to the field and providing a method that could assist biologists in their future research.

2 Deterministic models

2.1 Carrying capacity

We introduce a notion that is really important in population and logistic growth. First of all, some definitions :

Term	Definition
Exponential growth	Population growth that isn't limited by resource availability, allowing the population growth rate to keep increasing over time.
Logistic growth	Population growth that is limited by resource availability, causing the population growth rate to slow down as the population size gets larger.
Carrying capacity	The maximum number of organisms or populations that an ecosystem can support.

TABLE 1 – Definitions of population growth terms

In nature, exponential growth can not last because space and resources are limited. Factors like food, predators, and disease set an ecosystem's carrying capacity, the largest population it can support. The logistic growth model includes carrying capacity, showing a more realistic population growth. Populations grow fast at first with plenty of resources. But as they get bigger and near carrying capacity, growth slows due to resource limits, forming an S-shaped curve. This model shows how reproduction increases population size, but resource limits keep it balanced at carrying capacity.

In our model, we will say mutants are the densest because we add δ which will always be between 0 and 1 for the SSD model. For the Replicative Advantage one, $\delta = 1$ and for SSS, it will depend if we consider mutants only slowest (i.e $\delta = 1$), or densest and slowest (i.e $\delta < 1$). To see this, consider the carrying capacities of w and m respectively. The wildtype carrying capacity w^* satisfies

$$cm(N_{ss} - w - \delta m) = 0$$

for $w > 0$ and $m = 0$.

$$w^* = N_{ss}.$$

Similarly, the mutant carrying capacity m^* satisfies

$$cm(N_{ss} - w - \delta m) = 0$$

for $m > 0$ and $w = 0$.

$$m^* = \frac{N_{ss}}{\delta}.$$

Hence, the carrying capacities, for the SSD model, are $w^* = N_{ss}$ and $m^* = \frac{N_{ss}}{\delta} > N_{ss} = w^*$, as $\delta < 1$.

2.2 A Simple Model Highlighting the Need for Stochastic Approaches

2.2.1 Replicative Advantage model (RA)

Here we just want to show a basic model that justifies the idea that to simulate population growth, the deterministic way is not adequate. Our 2 species have the same carrying capacity ($\delta = 1$) but the mutants, have a higher replication rate.

The system of ordinary differential equations [4] corresponding to these rates is :

$$\frac{dw}{dt} = c(N_{ss} - w - m)\,w$$
$$\frac{dm}{dt} = [c(N_{ss} - w - m) + k_{ra}]\,m.$$

where $k_{ra} > 0$ is the replicative advantage coefficient.

In a setting with two species, we will often ask questions about the proportion of mutants, denoted as h or also named heteroplasmy :

$$h = \frac{m}{w+m}.$$

If we want to apply stochastic methods, we have to consider chemical reactions. this point will be explain later, but here we just want to focus on the results.

$$W \xrightarrow{\mu} \emptyset \qquad M \xrightarrow{\mu} \emptyset$$
$$W \xrightarrow{\lambda_w(w,m)} W+1 \quad \text{and} \quad M \xrightarrow{\lambda_m(w,m)} M+1 \,.$$

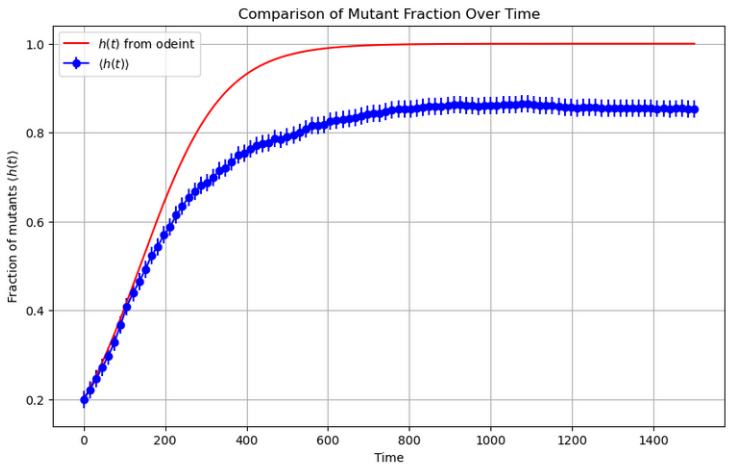


FIGURE 2 – Comparison deterministic and stochastic dynamics : $\mu = 0.05$, $c = 0.01$, $N_{ss} = 50$, $\epsilon = 0.01$, $w_0 = 40$, $m_0 = 10$, $T = 1500$, and 100 simulations.

We observe a discrepancy : the ensemble average $\langle h(t)\rangle$ does not follow the solution $h(t)$ of the ODE. This discrepancy arises due to the inherent stochasticity of the process, leading to fluctuations in the population dynamics. Indeed, when we examine the evolution of the mutants over time in each simulation, we sometimes get different results. There are simulations where the mutants reach zero, and they cannot replicate. We have a mutant extinction, which is due to the stochasticity of the model, distinguishing it from the deterministic one. And sometimes, there are also simulations where mutants reach fixation. This explains why we observe the graph

4. All the parameters in the equation are referred in the introduction

in Figure 2.

That is why we will focus on stochastic modelling to account for the behaviour of mutants.

This RA model is the basis of the work and the basic hypothesis on how mutants behave. However we can ask ourselves if mutants having a better replication rate is the good case to simulate the deletions? We will see that this model can be not the only one to explain the dynamic of mtDNA deletions and that is what we will study in this report.

3 Using stochastic models

To comprehend population growth, we initially explored models operating in discrete time, where time progresses in distinct steps. However, real-world scenarios are often more intricate, requiring a continuous perspective on time advancement.

This research project uses an algorithm known as the Gillespie or Stochastic Simulation Algorithm. Acquiring proficiency in this method was imperative for investigating muscle ageing and conducting simulations on diverse population dynamics.

3.1 Gillespie Algorithm

The Gillespie algorithm, is a method in probability theory for generating statistically accurate trajectories of stochastic systems where reaction rates are known. This algorithm is particularly effective for simulating chemical reactions in systems with low numbers of molecules. It is a variant of the dynamic Monte Carlo method and is widely used in computational systems biology.

In a reaction chamber with a finite number of molecules, the algorithm simulates the process by sampling random waiting times until a reaction occurs, rather than discretizing time, otherwise there would be times where no reaction occurs. The waiting time for a reaction is exponentially distributed, with the rate determined by the sum of the individual reaction rates.

3.1.1 Poisson Point Process

In this section, we consider a scenario where different chemicals can undergo reactions with one another, governed by various reaction rules. It is important to note that the term "chemicals" can refer to molecules, cells, or even organisms, while "reactions" encompass any type of interaction between them.

We focus on cases where the number of chemicals is so small that stochastic effects become significant. Under these conditions, chemical reactions can be modeled as a Poisson Process. This implies that each reaction occurs randomly at a rate that depends only on the current state of the system, without any influence from previous states or events. This kind of process is also known as a memoryless process.

If the rate of an event is denoted by λ, then the mean time to the next event is given by $\tau = \frac{1}{\lambda}$. Picture breaking down time into super tiny bits, like Δt. So, the chance of something happening in just one of these bits is $\lambda * \Delta t$.

3.1.2 How the algorithm works

The Gillespie algorithm simulates the exact time course of a set of chemical reactions. It takes into account the random nature of reaction events and provides a way to generate statistically correct trajectories of the system over time. The algorithm consists of several key steps to determine which reaction occurs next and the time interval until that reaction.

Steps of the Gillespie Algorithm

1. **Initialization** : Set the initial number of molecules for each species and the initial time.
2. **Reaction Propensities** : Calculate the propensity functions for each possible reaction. The propensity function a_j for reaction j is given by :

$$a_j = c_j \cdot h_j$$

where c_j is the rate constant and h_j is the number of possible combinations of reactant molecules for reaction j.

3. **Total Propensity** : Compute the sum of all propensity functions :

$$a_0 = \sum_j a_j$$

4. **Time to Next Reaction** : Generate the time τ until the next reaction occurs using an exponential distribution :

$$\tau = \frac{1}{a_0} \ln\left(\frac{1}{r_1}\right)$$

where r_1 is a uniform random number between 0 and 1.

5. **Reaction Selection** : Select which reaction will occur next. This is done by generating another uniform random number r_2 and finding the smallest integer j such that :

$$\sum_{i=1}^{j} a_i > r_2 \cdot a_0$$

6. **Update** : Update the number of molecules of each species according to the stoichiometry of the selected reaction. Also, update the time by adding τ to the current time.
7. **Iterate** : Repeat steps 2-6 until the desired end time is reached.

Example to understand its dynamic

Consider a simple system with two species A and B and two reactions :
— $A \rightarrow B$ with rate constant k_1
— $B \rightarrow A$ with rate constant k_2
The steps of the Gillespie algorithm for this system would be :

1. **Initialization** : Start with N_A molecules of A and N_B molecules of B, and set the initial time $t = 0$.
2. **Reaction Propensities** : Calculate $a_1 = k_1 \cdot N_A$ and $a_2 = k_2 \cdot N_B$.
3. **Total Propensity** : Compute $a_0 = a_1 + a_2$.
4. **Time to Next Reaction** : Generate τ using the exponential distribution.
5. **Reaction Selection** : Use r_2 to determine whether reaction 1 or reaction 2 occurs.
6. **Update** : Adjust N_A and N_B according to the chosen reaction and update the time t.
7. **Iterate** : Continue the process until the simulation end time is reached.

This method ensures that the stochastic nature of chemical reactions is accurately captured, providing insights into the dynamics of the system that might be missed by deterministic approaches.

3.2 Obtaining effective stochastic differential equations SDE's

This section summarizes the methodology employed by Professor Nick Jones and Dr. Ferdinando Insalata. Specifically, they used a method called stochastic dimensionality reduction. This method was used to derive a single stochastic differential equation (SDE) from an Ito system. [5] Given the fundamental importance of this method to the overall work, I have outlined the key concepts to elucidate its functionality. Understanding these principles was crucial for me to effectively apply the algorithm in deriving the final SDE and determining the drift term. Here, the general form of the SDEs we deal with has two terms [6] : drift + diffusion. We are interested in the drift because we can understand from there if there is a bias in the dynamics that makes the number or individuals grow/stay constant/decrease depending on its sign and value.

Stochastic reduction simplifies complex systems for easier analysis and simulation. This process involves transitioning from a Chemical Master Equation (CME) to a Fokker-Planck equation via the Kramers-Moyal expansion, then converting the Fokker-Planck equation to a system of Itô Stochastic Differential Equations (SDEs), and finally reducing this system to a single effective SDE using the central manifold approach.

From the rate functions, we can derive a set of differential equations that describe how the probability of being in different states specifically, having different numbers of molecules changes over time. This set of equations is known as the CME.

$$\frac{dP(\mathbf{n},t)}{dt} = \sum_{\mathbf{n}'\neq\mathbf{n}} \left[W(\mathbf{n}|\mathbf{n}')P(\mathbf{n}',t) - W(\mathbf{n}'|\mathbf{n})P(\mathbf{n},t)\right] \tag{1}$$

where, $W(n|n')$ is the transition rate from state n' to n.

Using the Kramers-Moyal expansion, we approximate discrete transitions by continuous derivatives, leading to a Fokker-Planck equation :

$$\frac{\partial P(\mathbf{x},t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}\left[A_i(\mathbf{x})P(\mathbf{x},t)\right] + \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}\left[B_{ij}(\mathbf{x})P(\mathbf{x},t)\right] \tag{2}$$

where $A_i(\mathbf{x})$ and $B_{ij}(\mathbf{x})$ are drift and diffusion terms. This Fokker-Planck equation corresponds to the Itô SDE :

$$d\mathbf{X}_t = \mathbf{A}(\mathbf{X}_t)dt + \mathbf{B}(\mathbf{X}_t)d\mathbf{W}_t \tag{3}$$

where $\mathbf{W}_t$ is a Wiener process. In our study, we have

$$\mathbf{A} := \mathbf{ST}$$

where $\mathbf{A}$ is a vector of length N, $\mathbf{S}$ is the $N \times R$ stoichiometry matrix, and $\mathbf{T}$ is the vector of transition rates of length R.

5. Todd L Parsons, and Tim Rogers, *Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift : a constructive approach with examples from theoretical biology*

6. Like in Eq.3

When the system has distinct timescales, we can further reduce the system by averaging out the fast variables. Considering a system of Itô SDEs :

$$\begin{cases} dX_t = f(X_t, Y_t)dt + \sigma_X(X_t, Y_t)dW_t^X \\ dY_t = \frac{1}{\epsilon}g(X_t, Y_t)dt + \frac{1}{\sqrt{\epsilon}}\sigma_Y(X_t, Y_t)dW_t^Y \end{cases} \tag{4}$$

with ϵ being small, the fast variable Y_t reaches a quasi-stationary distribution conditional on X_t. This leads to a reduced SDE for X_t :

$$dX_t = \overline{f}(X_t)dt + \overline{\sigma_X}(X_t)dW_t^X \tag{5}$$

where $\overline{f}(X_t)$ and $\overline{\sigma_X}(X_t)$ are the averaged drift and diffusion terms over the stationary distribution of Y_t.

3.3 Error bars

You will see throughout this study, lots of graphics that present error bars. In our stochastic simulations, we calculate the errors on the average value shown in plots as the standard error of the mean $\sigma/\sqrt{(N)}$, with σ the standard deviation and N the number of simulations.

4 Stochastic survival of the densest : SSD

Our approach is based on a straightforward stochastic model that describes the dynamics of a population of mtDNA, consisting of wildtype and mutant alleles, regulated centrally by the nucleus. The primary variable of interest is heteroplasmy (h), defined as the fraction of the population that is mutant. This model primarily examines how a pre-existing mutation can proliferate and achieve high heteroplasmy through clonal expansion.

We will see models that are either neutral, meaning both wildtype and mutant alleles share the same constant degradation rate, μ, and an identical replication rate, λ, given by :

$$\lambda(w,m) = \mu + c(N_{\text{ss}} - w - \delta m),$$

where c, N_{ss}, and δ are parameters. However, we will also explore models that incorporate a higher degradation rate for mutants, where μ for mutants is given by :

$$\mu + \epsilon$$

The replication rate is adjusted according to the deviation of the current population size, represented by $w + \delta m$, from a target population size N_{ss}. Importantly, mutants contribute less to the overall population size than wildtypes, with the parameter $0 \leq \delta < 1$ indicating the relative contribution. We demonstrate that the condition $0 \leq \delta < 1$ implies that mutants are the more densely packed species.

4.1 One unit

4.1.1 Neutral model

$$\begin{aligned} W &\xrightarrow{\mu} \emptyset \\ M &\xrightarrow{\mu} \emptyset \\ W &\xrightarrow{\lambda(w,m)} W+1 \\ M &\xrightarrow{\lambda(w,m)} M+1 \end{aligned}$$

where :
— $W \xrightarrow{\mu} \emptyset$: Wildtype individuals degrade at rate μ.
— $M \xrightarrow{\mu} \emptyset$: Mutant individuals degrade at rate μ.
— $W \xrightarrow{\lambda(w,m)} W+1$: Wildtype individuals replicate at rate $\lambda(w,m)$.
— $M \xrightarrow{\lambda(w,m)} M+1$: Mutant individuals replicate at rate $\lambda(w,m)$.
We consider a network with $N = 2$ species and $R = 4$ reactions. The stoichiometry matrix S is given by :

$$S = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}$$

The global rates for wildtypes or mutants are determined by multiplying the per molecule rates μ and λ by the population sizes w or m, respectively.

Moreover, with that system we tend to say that the number of mutants will increase with time but the heteroplasmy will fluctuate around its initial value h_0. Indeed, mutants have a higher carrying capacity $\delta m < N_{ss}$.It might be puzzling that while the average number of mutants

$\langle m \rangle$ increases, the average fraction of mutants $\langle h \rangle$ remains constant. Here's a simpler way to understand this :
Every individual in the system, whether mutant or wildtype, has the same death and replication rates. This means each individual has an equal chance of generating descendants that will eventually take over the entire population. Each individual, therefore, has the same probability of becoming the common ancestor of a future population.
This probability for each individual to become the common ancestor is $\frac{1}{w_0+m_0}$, where w_0 and m_0 are the initial numbers of wildtypes and mutants, respectively. Since there are initially m_0 mutants, the probability that a mutant will take over the whole population is $\frac{m_0}{w_0+m_0} = h_0$, where h_0 is the initial fraction of mutants.
At mutant fixation, the steady-state mutant population is $\frac{N_{ss}}{\delta}$, and the mutant fraction is 1. Thus, for $t \to \infty$:

$$\langle m \rangle \to h_0 \cdot \frac{N_{ss}}{\delta}$$

and

$$\langle h \rangle \to h_0 \cdot 1 = h_0.$$

In other words, $\langle m \rangle$ increases to its final value because the population size at fixation ($\frac{N_{ss}}{\delta}$) is larger than the initial mutant count m_0 for $0 < \delta < 1$. Conversely, the final value of $\langle h \rangle$ remains the same as the initial fraction h_0, explaining why $\langle h \rangle$ is constant over time.

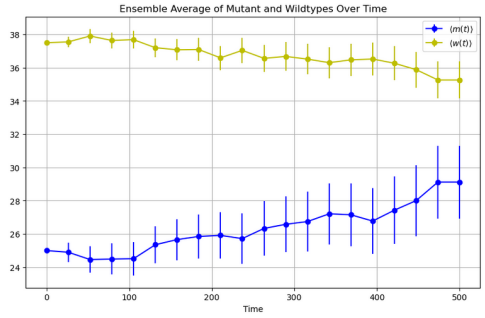


FIGURE 3 – Evolution of mutants and wildtypes : $\mu = 0.05$, $c = 0.01$, $N_{ss} = 50$, $w_0 = 37.5$, $m_0 = 25$, $T = 5000$, and 200 simulations.

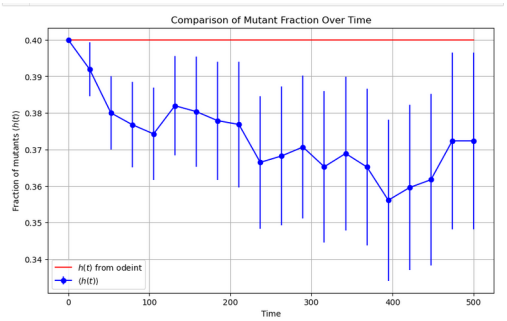


FIGURE 4 – Evolution of mean mutant fraction in ODE and in stochastic model : $\mu = 0.05$, $c = 0.01$, $N_{ss} = 50$, $w_0 = 37.5$, $m_0 = 25$, $T = 5000$, and 200 simulations

In summary, despite the increasing mutant population, the mean mutant fraction stays constant due to the identical birth and death rates of mutants and wildtypes. The large error bars at later time points indicate an evolution in $\langle h \rangle$, but the overall trend confirms that $\langle h \rangle$ remains steady around 0.4.

4.1.2 Selective disadvantage model

It is intriguing to examine the impact of introducing preferential degradation for mutants by assigning them a higher replication rate of $\mu + \epsilon$, where $\epsilon > 0$.

Here's how it works :

In our original model, mutants and wildtypes had the same degradation rate μ. To model preferential elimination, we now assign mutants a degradation rate of $\mu + \epsilon$, with $\epsilon > 0$. This means mutants die at a faster rate compared to wildtypes.

This adjustment can influence the dynamics of the system, potentially reducing the number of mutants over time due to their higher elimination rate. But we will see that this first thought is not totally true.

By reducing the dimension of the system [7], we derive an effective stochastic differential equation (SDE) for the mutant population. We have this equation.

$$dm = \left[\frac{2(1-\delta)\mu}{N_{ss}} - \varepsilon\right] m\left(1-\frac{\delta m}{N_{ss}}\right)dt + \frac{1}{N_{ss}}\sqrt{2m\mu(N_{ss}-\delta m)(N_{ss}+m-\delta m)}dW$$

The drift in m is positive for

$$\delta m < N_{ss} \quad \text{and} \quad \varepsilon < \frac{2(1-\delta)\mu}{N_{ss}}$$

It is important to see that ϵ is contained between 0 and $\frac{2(1-\delta)\mu}{N_{ss}}$, because if not, we will not observe a stochastic reversal selection.

In the stochastic scenario, there exists a critical value $\varepsilon_M = \frac{2(1-\delta)\mu}{N_{ss}}$ such that, for $\varepsilon < \varepsilon_M$, the mean mutant population $\langle m \rangle$ still increases, indicating a higher likelihood of mutant fixation compared to wildtype.

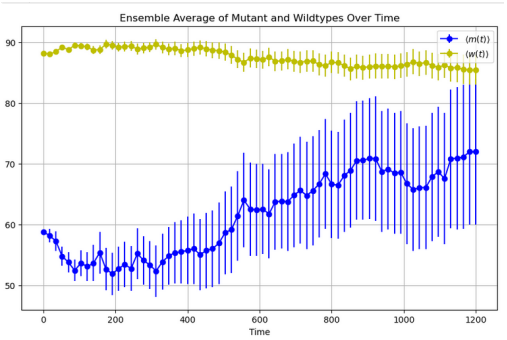


FIGURE 5 – Evolution of wildtypes and mutants : $\mu = 0.05$, $c = 0.01$, $N_{ss} = 100$,$\delta = 0.2$, $w_0 = 88$, $m_0 = 58$, $T = 1200$, $\epsilon = 0.000338$ and 100 simulations

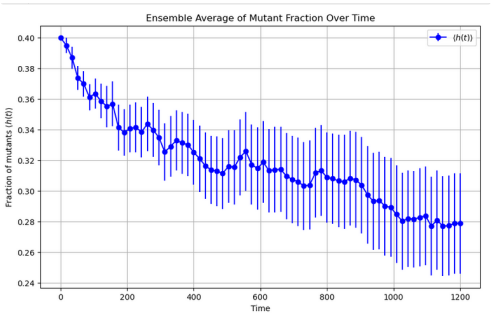


FIGURE 6 – Evolution of mean mutant fraction when mutants have a higher degradation rate : $\mu = 0.05$, $c = 0.01$, $N_{ss} = 100$,$\delta = 0.2$, $w_0 = 88$, $m_0 = 58$, $T = 1200$, $\epsilon = 0.000338$ and 100 simulations

When analyzing the latest plot, we observe a decrease in $\langle h \rangle$, contrasting with the constant behavior seen in the plot from the previous week. This divergence may stem from a fundamental alteration in the model : the introduction of differing death rates for wildtypes and mutants. Consequently, each individual no longer possesses an equal probability of initiating a lineage capable of dominating the entire system. Consequently, as t tends to infinity, the calculated value of $\langle h \rangle$ differs from h_0 observed in the previous model. Thus, $h_\infty < h_0$.

This specific evolution is called stochastic reversal selection (see Figure.5) : even if mutants are less fit due to the higher degradation rate, their number grows because of stochastic effects. To reach this we needed to have : higher mutant carrying capacity, noise, and higher mutant death rate.

However, if we want to take it to the next level and be able to make a link between these models and muscle fibres, we need to look at another model.

7. For detailed derivations and the impact of varying the degradation rate for mutants, refer to Appendix 1.

4.2 Two units

We formulated both deterministic and stochastic models consisting of one unit, with the latter representing the simplest system that exhibits stochastic survival of the densest. An immediate progression involves introducing spatial structure, where the system comprises compartments, each endowed with resources shared among its inhabitants, interconnected through some form of coupling.

While the death rate remains constant and uniform across both units (μ), the replication rate of each unit (λ_i) varies, dependent solely on the population within that unit : $\lambda_i = \mu + c(N_{ss} - w_i - \delta m_i)$, for $i = 1,2$. This unit-specific replication rate encapsulates compartmentalization, with the population in each unit vying for the resources (N_{ss}) within that unit. Additionally, individuals migrate between the systems at a constant rate γ. The stochastic formulation of the model is obtained by replicating Eq. (4.5) for w_2 and m_2 and adding lines accounting for the migration of individuals between the two units. The full chemical reaction network is :

$$
\begin{aligned}
W_1 &\xrightarrow{\mu} \emptyset \\
M_1 &\xrightarrow{\mu} \emptyset \\
W_1 &\xrightarrow{\lambda_1(w_1,m_1)} W_1 + 1 \\
M_1 &\xrightarrow{\lambda_1(w_1,m_1)} M_1 + 1 \\
W_2 &\xrightarrow{\mu} \emptyset \\
M_2 &\xrightarrow{\mu} \emptyset \\
W_2 &\xrightarrow{\lambda_2(w_2,m_2)} W_2 + 1 \\
M_2 &\xrightarrow{\lambda_2(w_2,m_2)} M_2 + 1 \\
W_1 &\xrightarrow{\gamma} W_2 \\
M_1 &\xrightarrow{\gamma} M_2 \\
W_2 &\xrightarrow{\gamma} W_1 \\
M_2 &\xrightarrow{\gamma} M_1 .
\end{aligned}
$$

The last four lines mean that when a wildtype (or mutant) migrates, it is then counted as a wildtype (mutant) in the other unit.

4.2.1 Neutral model

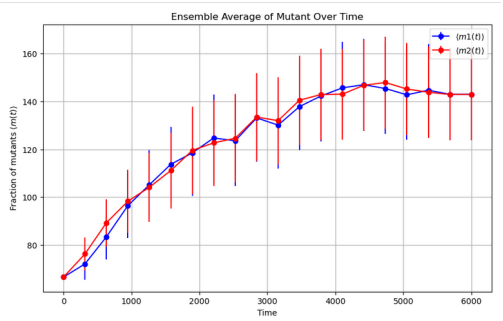


FIGURE 7 – Evolution of mutants : $\mu = 0.05$, $c = 0.01$, $N_{ss} = 100$,$\delta = 0.5$, $\gamma = 0.05$, $w_0 = 66$, $m_0 = 55$, $T = 1200$, and 50 simulations

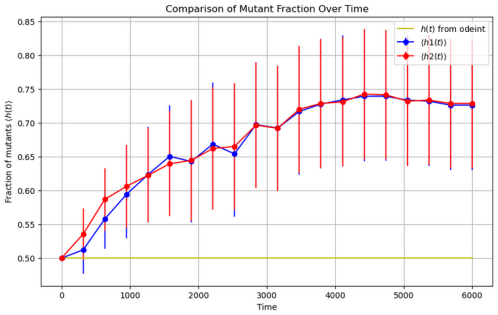


FIGURE 8 – Evolution of mean mutant fraction : $\mu = 0.05$, $c = 0.01$, $N_{ss} = 100$,$\delta = 0.5$, $\gamma = 0.05$, $w_0 = 66$, $m_0 = 55$, $T = 1200$, and 50 simulations

We observe a notable increase in the mutant fraction, a phenomenon attributable to stochastic noise-induced selection. Interestingly, setting $\delta = 1$ abolishes this specific evolutionary trend, underscoring the role of δ in the dynamics of mutant evolution.

When considering the dynamics within each unit, it's apparent that diffusion occurs at a much slower rate compared to birth and death processes. Consequently, units typically reach fixation before any significant diffusion event takes place. This relative independence of evolution within each unit persists until an individual migrates between units.

Upon migration of a mutant to a wild-type unit, its fixation probability, or the likelihood of assuming control of that unit, is relatively high, albeit inversely proportional to the carrying capacity N_{ss}. Conversely, when a wildtype migrates, its fixation probability decreases due to the higher presence of mutants, if there are more mutants in the other units.

Given the higher migration and fixation rates for mutants, the overall probability of mutants eventually dominating both units exceeds that of wildtypes. [8]. Consequently, the average fraction of mutants increases from its initial value. Notably, experiments exploring increased γ reveal a divergence in the evolution of h. Fast diffusion tends to homogenize the behavior of units, negating the observed increase in the average fraction of mutants.

8. I am talking here on the overall rate, given by rate times number of individuals

4.2.2 Selective disadvantage model

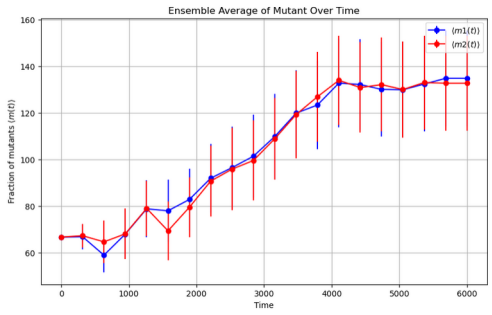


FIGURE 9 – Evolution of mutants : $\mu = 0.05$, $c = 0.01$, $N_{ss} = 100$,$\delta = 0.5$, $\gamma = 0.05$, $w_0 = 66$, $m_0 = 55$, $T = 1200$, $\epsilon = 0.000338$ and 50 simulations

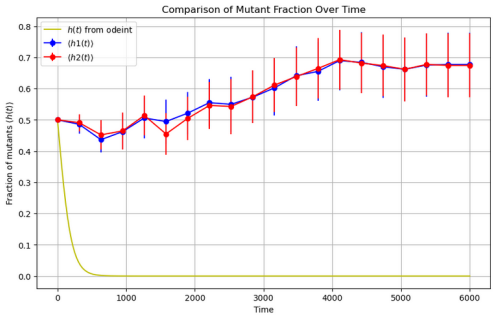


FIGURE 10 – Evolution of mean mutant fraction when mutants have a higher degradation rate : $\mu = 0.05$, $c = 0.01$, $N_{ss} = 100$,$\delta = 0.5$, $\gamma = 0.05$, $w_0 = 66$, $m_0 = 55$, $\epsilon = 0.000338$ $T = 1200$, and 50 simulations

We observe that the evolution of $\langle h \rangle$ shows an increase, whereas in the model with one unit, there was not an increase of heteroplasmy. This difference arises because, in the model with two units, species migration is a crucial factor :
Mutants exhibit a higher likelihood of establishing themselves upon migrating to a wild-type unit, a probability that remains notable despite their elevated degradation rate.

Additionally, stochastic fluctuations play a pivotal role :
Random fluctuations may favor mutants in specific units, enabling their average fraction to increase across the entire system. The intuitive explanation provided is apt, with migrations and noise serving as key factors (alongside density) for this effect to manifest.

4.3 SSD conclusion

Professor Jones' research group has developed a spatial stochastic model that accurately predicts the clonal expansion of mitochondrial deletions in skeletal muscle fibers. Remarkably, the new model demonstrates that deletions can expand even if they are preferentially eliminated. This phenomenon hinges on the increased density of deletions and the spatial and stochastic nature of the model.

While not claiming SSD as the definitive mechanism underlying muscle ageing, this study suggests that SSD could offer valuable insights for understanding the phenomenon and devising potential treatments. Notably, the wave speed of mutants in SSD models is inversely related to copy numbers, implying that increasing copy numbers through treatments or medications could potentially mitigate the wave speed, thus limiting the expansion of mutants.

5 Stochastic survival of the slowest : SSS

The initial phase was focused on the SSD model. It took me about a month to study and understand its objectives and dynamics. This understanding laid the groundwork for delving deeper into the project and initiating a new phase based on the SSD model's principles. Therefore, it was crucial to present all the parameters and outcomes as they form the foundation of our work.

From now on, my focus shifts to a new model where the population of wildtypes is referred to as *the slowest in everything*. This term reflects changes in their birth and degradation rates. My task was to implement this new model and explore its outcomes, so that I could find out whether a site frequency spectrum (SFS) [9] could distinguish between SSS, SSD and RA.

5.1 Mutants densest + Wildtypes Slowest

In this model, we have those characteristics for mutants and wildtypes.

1. **1 unit** : 2 species only
2. **Mutants densest** : higher carrying capacity
3. **Wildtypes slowest** : in death and birth

5.1.1 One unit

We still have the four reactions like in SSD but here, the rates of birth and death of wildtypes have changed :

$$\begin{array}{c} W \xrightarrow{\mu-\rho} \emptyset \\ W \xrightarrow{\lambda_w(w,m)-\rho} W+1 \end{array} \quad \text{and} \quad \begin{array}{c} M \xrightarrow{\mu} \emptyset \\ M \xrightarrow{\lambda_m(w,m)} M+1 \end{array} .$$

Here ρ is a parameter that implicates that wildtypes are slower. We force ρ to be positive and with the stochastic reduction dimension, we can get a range of values for ρ. Considering that we have added a new parameter, we needed to see if the stochastic differential equation would change [10]. We are here only presenting the drift term of the equation.

1. **Deterministic model**

$$\begin{aligned} \frac{dw}{dt} &= c(N_{ss} - w - \delta * m)w \\ \frac{dm}{dt} &= c(N_{ss} - w - \delta * m)m. \end{aligned}$$

2. **Stochastic model**
   We found a new final stochastic equation where the drift term is

$$\frac{2m(-N_{ss} + m\delta)(\rho + (-1+\delta)\mu)}{N_{ss}^2}$$

Again, we want the drift term to be positive. Also, we consider that the number of mutants can't get higher than N_{ss}/δ. So the first parenthesis $(-N_{ss} + m\delta)$ is negative. Therefore, the second

9. This will be explained in the next section
10. To do so, we used the same method of stochastic reduction dimension

one needs to be negative and we find a ρ max which is equal to : $-\mu \cdot (\delta - 1)$. It is the maximum possible decrease in the rates for mutants and degradation that allows the drift to be positive.

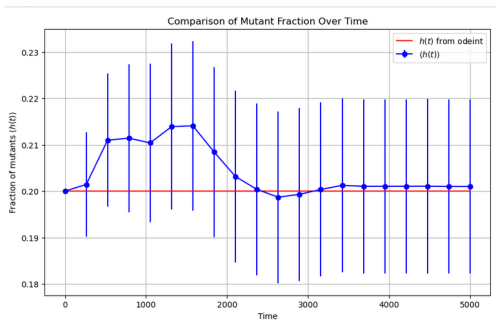
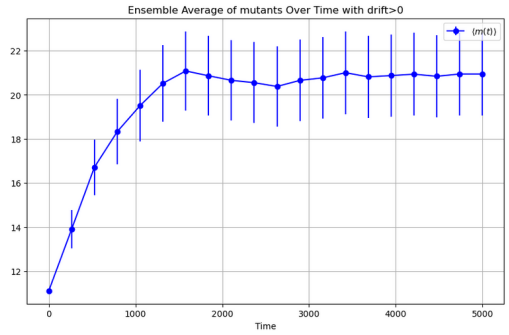
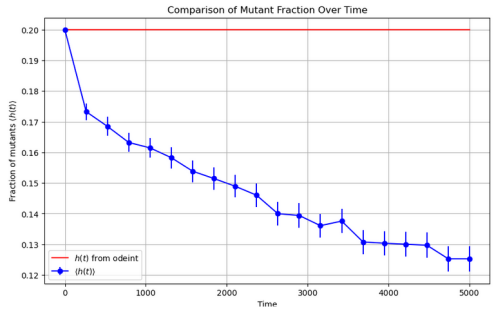
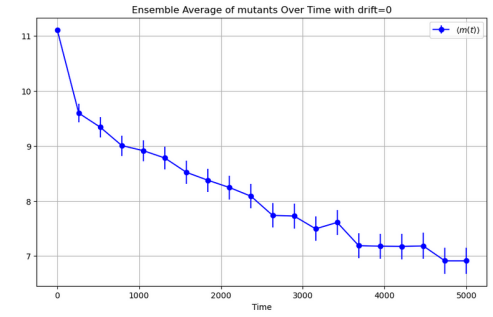
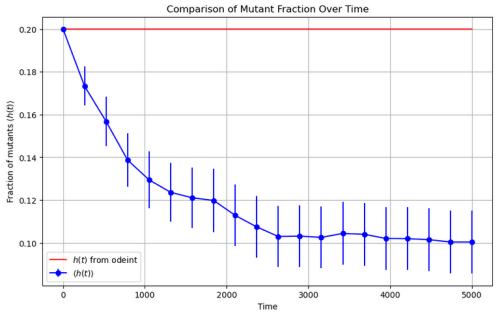
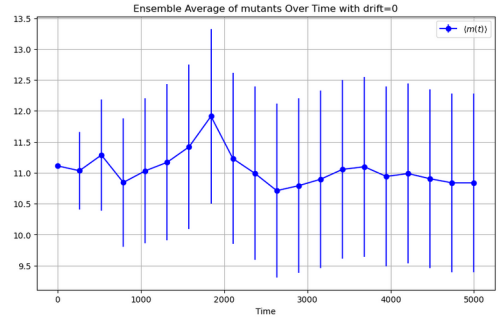
Sign of the drift term	**Heteroplasmy evolution**	**Mutants evolution**
>0		
<0		
=0		

TABLE 2 – Comparison of different models based on the sign of the drift term : $\mu = 0.05$, $c = 0.01$, $N_{ss} = 100$, $w_0 = 88$, $m_0 = 22$, $T = 5000$, and 50 simulations.

Here, we can see that when the drift is positive (the most relevant case) we still have an increase in mutants but their fraction remains constant. It is pretty similar as the SSD model with one unit in a neutral model.

5.1.2 Two units

We are here simulating the model with the first model's characteristics but this time with 2 units.

The method used to obtain a condition on ρ can no longer be applied in this model, so we chose to pursue the simulations with the $\rho_{\max}$ found in the previous model.
Also, for the first results in Table 3, we are not changing the migration rate, but only the death and birth one for wildtypes.

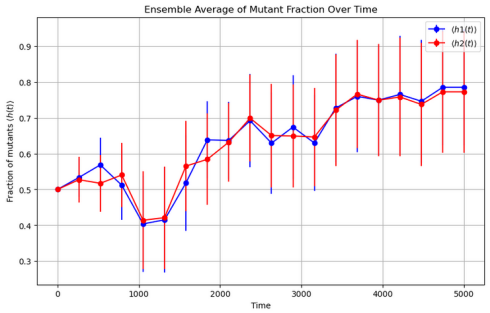
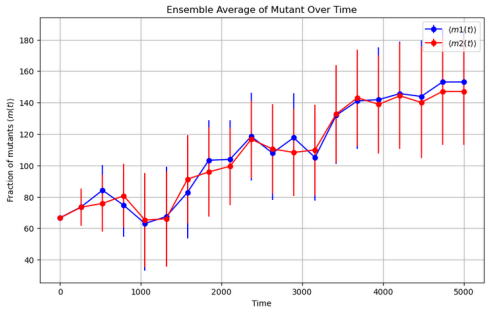
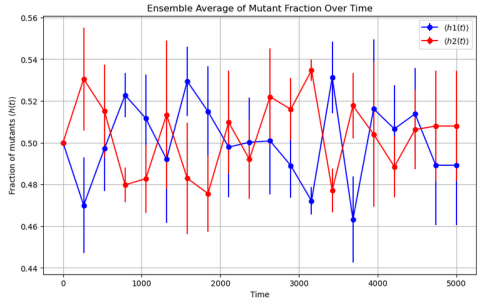
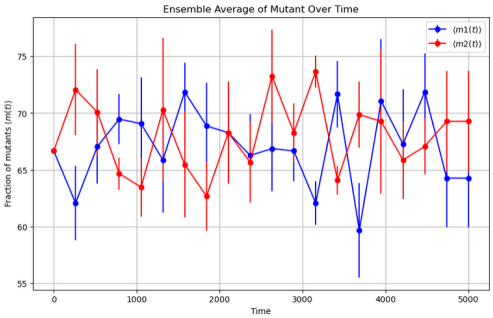
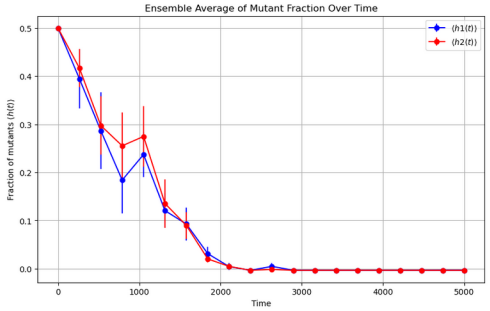
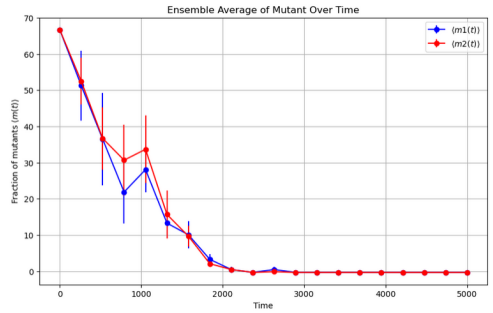
Sign of the drift term	Heteroplasmy evolution	Mutants evolution
>0		
<0		
=0		

TABLE 3 – Comparison of different models based on the sign of the drift term : $\mu = 0.05$, $c = 0.01$, $N_{ss} = 100$, $\gamma = 0.05$, $w_0 = 88$, $m_0 = 22$, $T = 5000$, and 50 simulations.

We can see that, having slower wildtypes with a drift positive, aims to having an increase in the population of mutants and also in heteroplasmy. We can see a noise induced selection. However, if we focus on a drift negative or null, our mutants are either going extinct or either evolving in a random way without having a real concrete evolution. This is due to the construction of a stochastic equation. And this encourages us to focus on cases where the drift is positive.

5.2 Mutants only slowest + Wildtypes neutral

Here we are analysing a case where $\delta = 1$. When the wildtypes are no longer the slowest, we have this deterministic system that is quite similar as the one seen above but without a higher

degradation rate for mutants.

$$\frac{dw}{dt} = c(N_{ss} - w - m)w$$
$$\frac{dm}{dt} = c(N_{ss} - w - m)m$$

With this model, h_{ODE} (heteroplasmy value derived from the solution of the deterministic equations) will stay constant at h0. And when we use the reduction algorithm, we have this drift term :

$$\frac{2m(-N_{ss} + m)\rho}{N_{ss}^2}$$

Here, if ρ is negative, we can't make the mutants "slower" because when we will want to change for example the death rate which is μ, normally we would do $\mu - \rho$, but with a $\rho < 0$ it would increase the death rate and not decrease it, and we would be in the model *Mutants with a higher degradation rate*.
So we can only focus on the dynamic where $\rho > 0$ and $\rho = 0$. This is important as it explains why we can't have a drift negative for mutants densest+slowest. Indeed, with a negative drift mutants cannot decrease.

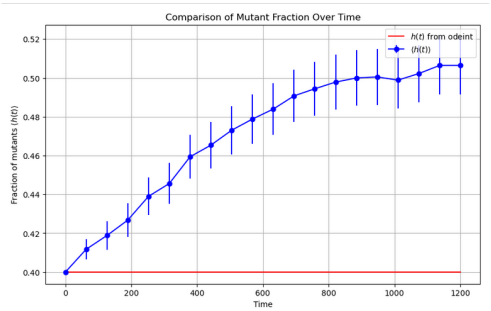


FIGURE 11 – Heteroplasmy evolution

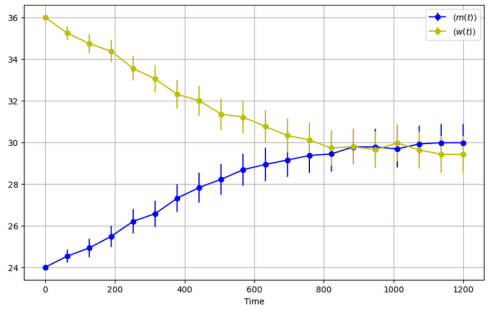


FIGURE 12 – Mutants evolution.

FIGURE 13 – Evolution when the drift is positive (>0 sign of the drift term),$\mu = 0.07$, $c = 0.0025$, $N_{ss} = 60$, $\gamma = 0.05$, $w_0 = 36$, $m_0 = 24$, $T = 1200$, and 1000 simulations, 1 UNIT

If we look only at the model with 2 UNITS and a drift positive, we have those results :

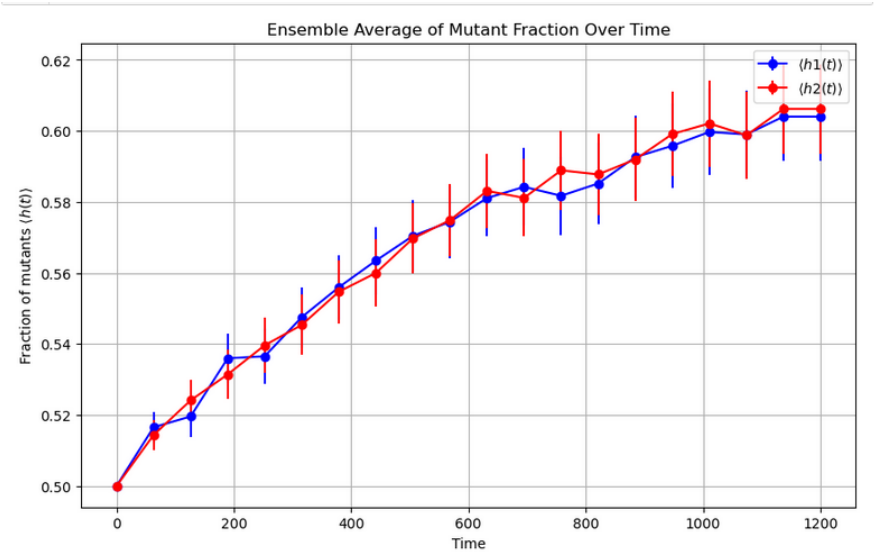


FIGURE 14 – Heteroplasmy evolution, 2 UNITS,$\mu = 0.07$, $c = 0.0025$, $N_{ss} = 60$, $\gamma = 0.05$, $w_0 = 36$, $m_0 = 24$, $T = 1200$, $\rho = 0.025$ and 1000 simulations

Although mutants replicate at a slower rate, we observe that they can still expand and dominate the population. This phenomenon is known as a noise-induced stochastic reversal of selection. In contrast, within a deterministic model, both species exhibit identical behavior, as the underlying equations are the same. According to natural selection, one would expect that the less advantaged species is unlikely to take over. However, stochastic simulations reveal the opposite outcome, highlighting the influence of randomness in evolutionary dynamics.

5.3 Mutants densest-slowest + Wildtypes neutral

In this model, we have those characteristics for mutants and wildtypes.

1. **1 unit** : 2 species only
2. **Mutants densest** : higher carrying capacity
3. **Mutants slowest**
4. **Wildtypes neutral**

Here we add ρ for the mutants to be slower.

5.3.1 One unit

$$\begin{array}{c} W \xrightarrow{\mu} \emptyset \\ W \xrightarrow{\lambda_w(w,m)} W+1 \end{array} \quad \text{and} \quad \begin{array}{c} M \xrightarrow{\mu-\rho} \emptyset \\ M \xrightarrow{\lambda_m(w,m)-\rho} M+1 \end{array}.$$

Like before, we want to find the new equations that simulate this new model.

1. **Deterministic model**

$$\begin{aligned} \frac{dw}{dt} &= c(N_{ss} - w - \delta * m)w \\ \frac{dm}{dt} &= c(N_{ss} - w - \delta * m)m. \end{aligned}$$

2. **Stochastic model**
We used the same method to find the drift term in the SDE's equations and we have this :

$$\frac{2m\,(\mathrm{Nss}-m\delta)\left(\rho\delta+\mu-\delta\mu\right)}{\mathrm{Nss}\rho^2}$$

Here, we don't have a $\rho_{\max}$ but we have a $\rho_{\min}$ which is

$$\frac{\mu}{\delta}\cdot(\delta-1)$$

This value is negative so, we have to be careful when we will want to see the dynamics for different signs of the drift term. Indeed, we are in a similar case of mutants only slowest, but now the ρ_{min} is not 0 but a negative value. Therefore, like we saw previously, it is only interesting and relevant to study the case where the drift is positive.

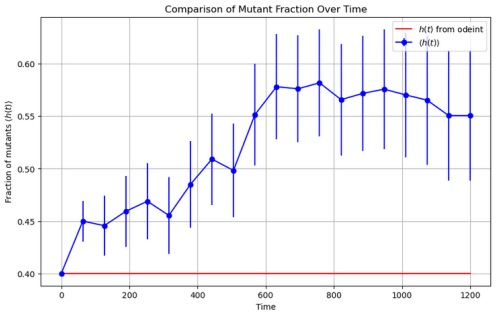


FIGURE 15 – Heteroplasmy evolution.

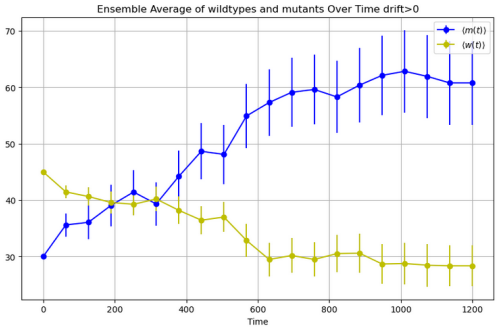


FIGURE 16 – Mutants evolution.

FIGURE 17 – Evolution when the drift is positive (>0 sign of the drift term).

5.3.2 Two units

Again, we couldn't use the method explain in Appendix 1, so we used the $\rho_{\min}$ found previously. We have again an increase for mutants and its fraction, so we are still in a *stochastic noise induced selection*. This is quite remarkable as the mutants have slower rates but they still manage to expand and take over the population.

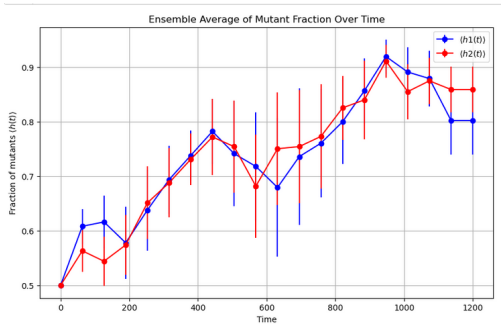


FIGURE 18 – Heteroplasmy evolution.

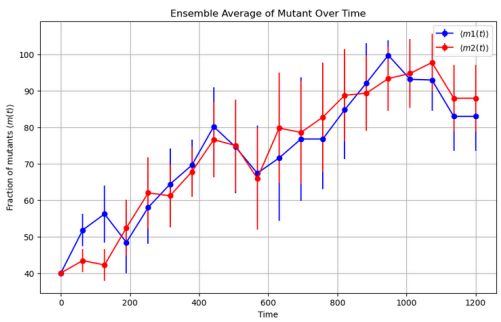


FIGURE 19 – Mutants evolution.

FIGURE 20 – Evolution with 2 units (>0 sign of the drift term).

5.4 SSS conclusion

To sum up, we have now seen different models with different characteristics to simulate the evolution of mutants and wildtypes in a mtDNA cell. And here are the ones that are most the relevant.

1. Replicative advantage RA : mutants higher replication rate
2. SSD : mutants densest+wildtypes neutral
3. SSS : mutants densest+slowest + wildtypes neutral
4. SSS : mutants only slowest + wildtypes neutral

With those 4 models, we can see the effect of the density (δ) and the disadvantage given to the mutants in terms of slower reproduction and degradation(ρ).
The results with 2 UNITS, show that in each cases, mutants and <h> increase if mutants are denser or slower. Even if they are only slower, they still manage to increase, which is surprising and shows a specific selection due to stochasticity.
Also, <h> increases and we still have this *noise induced selection* like we saw in SSD-2 UNITS.

Although we observe an increase in heteroplasmy in all models with two units, there is a difference in the pattern of heteroplasmy growth over time in the different models (see Fig2,8,10,14 and 18).

Another question can be raised with all these different models. Indeed, how could we distinguish these models and let the biologists know how the deletions occur in muscle fibres?

6 Distinguish the evolutionary mechanism at play

Our aim is to give biologists some data that would give them enough information to understand how the mutants expand and how we can simulate the deletions in a muscle fibre.

We have right now the Replicative Advantage model that is used a lot but we discovered that SSD or SSS can also be good models to simulate mtDNA deletions.

The new problematic is :
Given a high level of mutant fraction is observed at a point of time with no information about them before, is there a way to distinguish them ?

Point mutations involve additions, deletions, or alterations to mtDNA at the level of a single base pair. These mtDNA point mutations can be used to derive a summary statistic known as the site frequency spectrum (SFS), which we will define in the next section. We suggest that by comparing the site frequency spectrum, which can be obtained from a single time point, we can determine whether RA, SSD or SSS more accurately accounts for the accumulation of mtDNA deletions.

6.1 Site frequency spectra (SFS)

6.1.1 Use and definition

First of all, a quick definition to understand what is a site frequency spectrum.

The Site Frequency Spectrum (SFS) is a summary statistic in population genetics that describes the distribution of allele frequencies at polymorphic sites within a population. Mathematically, if x_i represents the frequency of the i-th allele, the SFS is the vector $\mathbf{S} = (S_1, S_2, \ldots, S_n)$, where S_i counts the number of sites with allele frequency i/n.

In our study, the site frequency spectrum (SFS) refers to the distribution of heteroplsamy frequency of mtDNA point mutations in skeletal muscle cells.

6.1.2 Point mutations

Besides the previously described chemical reaction network, we also need to consider the point mutations that occur in each mtDNA molecule. During each replication event, there is a probability that the newly replicated molecule will acquire some point mutations. We can model the number of point mutations gained as a $\text{Bin}(16569, \frac{1}{440{,}000})$ distribution, given that the human mitochondrial genome contains 16569 base pairs, and there is approximately one error per 440,000 nucleotides. Each point mutation is given a distinct identity, even if multiple mutations result from the same replication event.

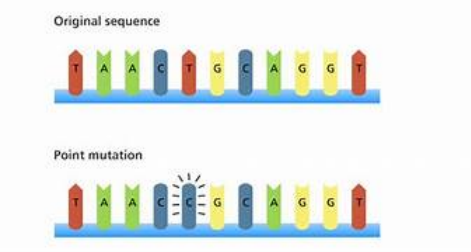


FIGURE 21 – Point mutations visualization

6.1.3 Statistical tests

To compare those 3 models, we are going to use statistical tests.

Rank-biserial correlation and Mann-Whitney U statistic

The rank-biserial correlation coefficient, introduced by Cureton, is a statistical measure used to compare two groups of data. It is particularly useful in situations where the data do not meet the assumptions necessary for parametric tests.

The rank-biserial correlation quantifies the degree of difference between two groups by comparing the ranks of their respective data points. This coefficient ranges from -1 to 1. A value of 1 indicates that all data points in one group are higher than all data points in the other group, while a value of -1 indicates the opposite. A value of 0 suggests no difference between the groups. The rank-biserial correlation is a valuable tool in non-parametric statistics, providing insight into the relative "largeness" of one group compared to another.

Definition : *The rank-biserial correlation coefficient* (r_{rb}) is defined as :

$$r_{rb} = \frac{2U}{n_1 n_2} - 1$$

where U is the Mann-Whitney U statistic, and n_1 and n_2 are the sample sizes of the two groups. You can also find it written that way :

$$r_{rb} = \frac{2 \times (Y_1 - Y_0)}{n}$$

Where :
— n = number of data pairs in the sample,
— Y_0 = mean of Y scores for data pairs with $x = 0$,
— Y_1 = mean of Y scores for data pairs with $x = 1$.
Example : Let's say you had the following data :
— Dichotomous variable : 1, 1, 1, 0, 1
— Ordinal variable : 3, 1, 5, 4, 2
The calculations would be :

$$Y_0 = 4 \quad \text{(only one ordinal variable is paired with 0)}$$

$$Y_1 = \frac{3+1+5+2}{4} = \frac{11}{4} = 2.75$$

$$n = 5$$

Giving a rank-biserial correlation coefficient of :

$$r_{rb} = \frac{2 \times (2.75 - 4)}{5} = \frac{-2.5}{5} = -0.5$$

Definition : *The Mann-Whitney test,* is a non-parametric statistical test used to assess whether there is a significant difference between the distributions of two independent samples. Mathematically, it ranks all observations from both samples together and compares the sums of ranks between the two groups.

Consider two datasets $x_1 \in \mathbb{R}^{n_1}$ and $x_2 \in \mathbb{R}^{n_2}$. The Mann-Whitney U test is a non parametric method for testing the alternative hypothesis H_1 against the null hypothesis H_0 :

$$H_0 : \text{The populations } x_1 \text{ and } x_2 \text{ have identical distributions,}$$

$$H_1 : \text{The distributions of the populations } x_1 \text{ and } x_2 \text{ differ.}$$

Like we saw before with the rank-biserial correlation coefficient :

$$U = \frac{n_1 n_2}{2}(1 - |r_{rb}|),$$

which approximates a normal distribution with mean m_U and variance σ_U^2 given by :

$$m_U = \frac{n_1 n_2}{2},$$

$$\sigma_U^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12} - \frac{n_1 n_2 \sum_{r=1}^{R} (t_r^3 - t_r)}{12(n_1 + n_2)(n_1 + n_2 - 1)},$$

where R is the number of unique ranks and t_r represents the number of ties at the r-th rank. The two-tailed p-value is determined by :

$$p = 2\Phi\left(-\frac{U - m_U}{\sigma_U}\right),$$

with $\Phi(\cdot)$ denoting the cumulative density function of a standard normal distribution.

6.2 SSD vs RA

An important aspect of this section is the rationale behind focusing solely on simulations where the mutants have reached fixation, as well as the reasons for matching either carrying capacities, fixation times, or wave speeds.

From a biological perspective, when analyzing a muscle fibre, we observe a certain number of mutant copies. However, the timing of their occurrence and the mechanisms of their emergence remain unknown. We have decided to select simulations where mutants have taken over the entire population. This approach aligns with the observation that in pathology, we only detect mutants once they have become prevalent. In addition, the exact time for mutants to reach fixation can vary in biological systems. By focusing on simulations where mutants have reached fixation, we avoid uncertainties related to the timing of mutant appearance and ensure a consistent reference point. This approach allows us to compare different models based on two key factors :

1. **Matching Carrying Capacities :** To compare the models effectively, we align the carrying capacities to ensure that the population sizes are equivalent across models. This way, any observed differences can be attributed solely to the underlying mechanisms at work, rather than discrepancies in population size.
2. **Matching Average Time to Fixation :** It is impractical to assume precise knowledge of mutant carrying capacities for all biological systems. Experimental data on mutant carrying capacity are often specific to individual systems and obtaining such data for every system is infeasible. Therefore, we can't only match carrying capacities. Assuming access to sparse mutant copy number data, we can estimate the rate of increase in mutant fraction. Unlike matching carrying capacities, this task is more challenging due to the distinct profile shapes of SSD and RA, like we saw in the previous sections. Consequently, we match the rate of increase using mutant fixation times. However, since there is no analytical formula for fixation time, it must be matched numerically.

6.2.1 First step - SFS stability

To achieve accurate simulations, the system is allowed to evolve for a sufficient period in the absence of deletions. This means we wait for a certain amount of time before introducing mtDNA deletions, allowing point mutations to stabilize. Only then are deletion mutants introduced. It is important to note that point mutations occur in both deletion mutants and wildtypes, and they do not alter the dynamics between deletion mutants and wildtypes.

To determine the appropriate time for introducing deletions, we analyze when the rank-biserial correlation value reaches and remains at zero. I did not develop the code to find this specific time ; it was provided to me. The time at which deletions are introduced varies depending on the number of units in the model :

$$T_{\text{intro\_2units}} = 2500 \text{ days}$$

$$T_{\text{intro\_chain\_units}} = 10000 \text{ days}$$

For instance, when simulating the SSD model, we must wait for a certain period before deletions occur. This allows us to observe the evolution of heteroplasmy accurately.

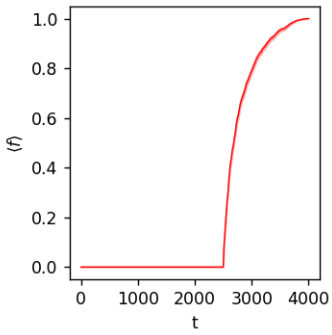


FIGURE 22 – SFS reaches stability then addition of deletions with the SSD model

6.2.2 2 Units model

Currently, we are analyzing a system with two units. In the next section, we will extend our study to a chain of units.

Matching carrying capacity

As mentioned earlier, we need to find a way to match the carrying capacity for both models. Recall that the SSD carrying capacity for mutants is given by $m^*_{SSD} = \frac{N_{ss}}{\delta}$, and for RA it is $m^*_{RA} = N_{ss} + \frac{k_{ra}}{c}$.

Given the biologists' data indicating that patients with diseases associated with mtDNA deletions exhibit a five-fold increase in mtDNA copy number compared to healthy individuals, we use $N'_{ss} = 5N_{ss}$.

Additionally, we must choose an appropriate value for k_{ra}. We can set $k_{ra} = cN_{ss}\left(\frac{1}{\delta} - 1\right)$. Assuming $\delta = 0.2$, we find $k_{ra} = 0.2$. With these values, we obtain the following results :

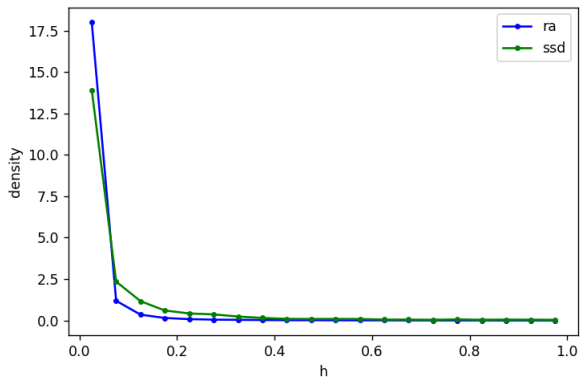


FIGURE 23 – SFS comparison with 2 units when heteroplasmy reaches 1

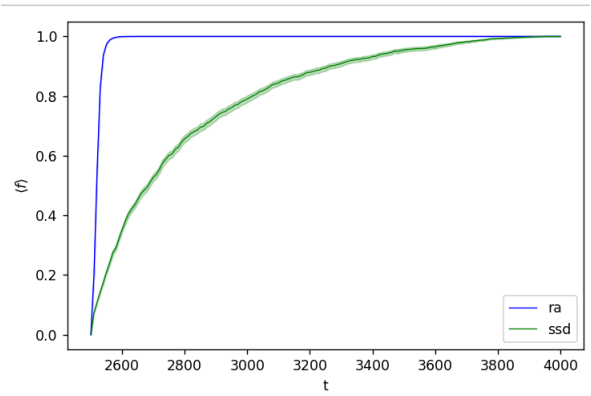


FIGURE 24 – Simulation results after waiting for the SFS to reach stability

The rank-biserial correlation coefficient is $r_b = -0.35026799946524667$, the U-statistic is $U = 282444216.0$, with $n_1 = 110627$ and $n_2 = 7859$, and the p-value is $p = 0$.
The rank-biserial correlation coefficient of -0.35 suggests a moderate negative association between the fixation times of mutants in the two compared models. This r_b value indeed indicates a significant distinction between the two models. To further discern the Site Frequency Spectrum (SFS) between the two models, we can examine the Mann-Whitney test results. Here, n_1 and n_2 represent the total number of point mutations in the SFS of the RA and SSD models, respectively. The distributions are evidently different as the p-value is equal to 0, and the SFS of RA is more concentrated at low h.

This can be explained by the fact that, with Replicative Advantage, individuals have a higher replication rate, which also leads to a higher rate of acquiring point mutations. Since heteroplasmy is defined as $\frac{1}{\text{total population at that time}}$, the value of h is minimized under these conditions.

With these results, we can assert that using the SFS is a reliable and valid method to distinguish between the two models. However, in practice, we do not have complete data nor knowledge, as we do in our simulations . By this, we mean that biologists, when analyzing muscle fibre, do not know how these mutants appeared. They can only observe that there are mutants in the cell (i.e., mutants have reached fixation). Therefore, we must consider the SFS of the models because we can only obtain a snapshot in time, which circumvents the inaccessibility of the mutants' dynamic history.

Matching average time to fixation

We can see that the average time of fixation of the proportion of mutants is not quite the same for the two models in Figure 26. Therefore, we can match their mutant average time to fixation. I was given the values for $N_{ss,RA}$ and k_{ra}, which the research team had previously found, to determine a fixation time that was approximately the same. These values are $N_{ss,RA} = 96.8$ and $k_{ra} = 0.008$. [11]

The idea was to match the values for RA and SSD using the following equation :

$$N_{RA_{ss}} + \frac{k_{ra}}{c} = \frac{N_{SSD_{ss}}}{\delta} \tag{6}$$

where $N_{RA_{ss}}$ and $N_{SSD_{ss}}$ are the target copy number parameters of the RA and SSD models, respectively. For the values of $N_{ss,SSD}$, we used 20 and $\delta = 0.2$.

Model	Parameters values	Average time of fixation (days)
SSD	$N_{ss} = 20, \delta = 0.2,$	3281
RA	$N_{ss} = 96.8, k_{ra} = 0.008$	3344

TABLE 4 – Average time to fixation match

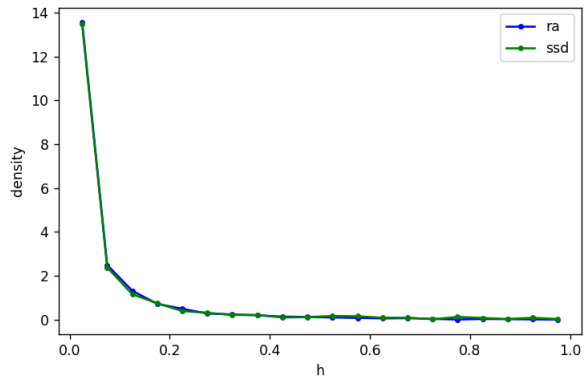


FIGURE 25 – SFS with 2 units when heteroplasmy reaches 1

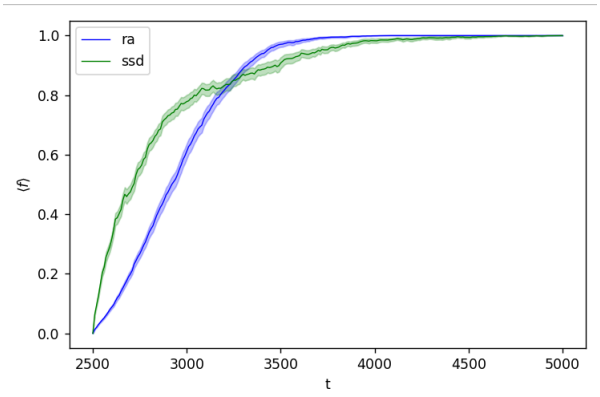


FIGURE 26 – Simulation results after waiting for the SFS to reach stability

If we just look at the graphs and especially Figure 27, the SFS are not significantly different, and the absolute value of the rank-biserial coefficient is 0.025. The Mann-Whitney test yields a p-value lower than 0.05, strongly suggesting that both distributions are indeed different and that the SFS can still distinguish both models :

The rank-biserial correlation coefficient is $r_b = -0.02570784127431842$ and the p-value is $p = 1.1692045723310603 \times 10^{-20}$.

6.2.3 Muscle fibres = chain of units

In this section, we simulate muscle fibers as a chain of units. Due to time constraints, I was unable to simulate the models with an extensive chain of units, as these simulations are

11. I was given the values that were found before my arrival to allow me more time to work on comparing both models to SSS

time-consuming. Instead, I chose to focus my research on a model involving two cells, with the consent of my tutor, using the new SSS model. Consequently, this section presents results obtained by the group rather than my own work. Nevertheless, I reviewed these results to understand their implications. This part is essential for our work, as simulating a chain of units can more accurately represent the functioning of muscle fibers. Therefore, these results are important for biologists and served as a foundational step for me to simulate muscle fibers using the SSS model. Because it was not my own, all the results and explanation are in the Appendix 3.

6.3 SSS results

In this section, we aim to analyze the Site Frequency Spectrum (SFS) for the stochastic survival of the slowest model.

Our focus is on a model where mutants are only the slowest and not the densest, i.e., where $\delta = 1$. Using the same methodology applied in comparing SSD and RA models, we seek to match either the carrying capacity, fixation time, or wave speed when employing a chain of units.

6.3.1 Comparison with SSD

2 units - Matching time fixation

Initially, we focus on a model of SSS where $\delta = 1$, meaning the mutants are only slower but not denser. Thus, the mutant carrying capacity is N_{ss} for SSS and $\frac{N_{ss}}{\delta}$ for SSD. In our case, the value does not change because we are adjusting both the birth and death rates simultaneously. Therefore, our focus shifts to matching fixation time.

We set $N_{ss,SSD} = 20$, $\delta_{SSD} = 0.2$, and $c = 2.5 \times 10^{-3}$. For the SSS model, given that $\delta = 1$, we have $N_{ss,SSS} = 5 \times 20 = 100$ [12]. The time when the SFS reaches stability is 2500 days with 2 units. We then experimented with different values of ρ in the SSS model to approximate the same fixation time as SSD. More over, it is important to take into account that ρ can not be superior as the degradation parameter μ. Other wise, we would have a degradation rate negative which will break the Gillespie Algorithm. Here $\mu = 0.07$.

Model	Parameters values	Average time to fixation (days)
SSD	$N_{ss} = 20$, $\delta = 0.2$, $\rho = 0$	3281
SSS	$N_{ss} = 100$, $\delta = 1$, $\rho = 0.05$	4502
SSS	$N_{ss} = 100$, $\delta = 1$, $\rho = 0.06$	4389
SSS	$N_{ss} = 100$, $\delta = 1$, $\rho = 0.04$	4461
SSS	$N_{ss} = 100$, $\delta = 1$, $\rho = 0.001$	4254
SSS	$N_{ss} = 100$, $\delta = 1$, $\rho = 0.0001$	4239
SSS	$N_{ss} = 100$, $\delta = 1$, $\rho = 0.00001$	4062

TABLE 5 – Average to time fixation with different values of the rate difference ρ

6.3.2 Comparison with RA

2 units - Matching time fixation

Remember what we saw in the section 6.3.2. We were using the equation 6 to match the time fixation when we were comparing SSD with RA. Here, we will use the same logic but to compare

12. Based on experimental data showing that there are five times more mtDNA mutant cells than wildtypes

SSS with RA.

$$N_{RA_{ss}} + k_{ra}/c = N_{SSS_{ss}} = N_{SSD_{ss}}/\delta_{SSD} \tag{7}$$

To do so we used $N_{ss,RA} = 96.8$, $k_{ra} = 0.008$, $\delta_{SSD} = 0.2$, $\delta_{SSS} = 1$, $c = 2.5e-3$ and $N_{ss,SSS} = 100$ we tried different values of ρ to obtain approximately the same time fixation. For the RA model, the time to fixation is 3344 days. Our first attempt was to try with a very low ρ to see if it could get close enough to the RA fixation time.

Model	Parameters values	Average time to fixation (days)
RA	$N_{ss} = 96.8$, $k_{ra} = 0.008$	3344
SSS	$N_{ss} = 100$, $\delta = 1$, $\rho = 0.05$	4502
SSS	$N_{ss} = 100$, $\delta = 1$, $\rho = 0.06$	4389
SSS	$N_{ss} = 100$, $\delta = 1$, $\rho = 0.04$	4461
SSS	$N_{ss} = 100$, $\delta = 1$, $\rho = 0.001$	4254
SSS	$N_{ss} = 100$, $\delta = 1$, $\rho = 0.0001$	4239
SSS	$N_{ss} = 100$, $\delta = 1$, $\rho = 0.00001$	4062

TABLE 6 – Average time to fixation with different values of the rate difference ρ

6.3.3 Results

The following plots present the results for the optimal value of ρ in terms of achieving the average time to fixation in both the SSD and RA models. As shown, it was challenging to obtain a perfect fit that could precisely match the time to fixation.

Due to time constraints, I was unable to explore a wider range of parameter values that might have led to a better fit. However, with the best value of ρ that I could identify, I was still able to achieve some meaningful results.

Comparison with SSD

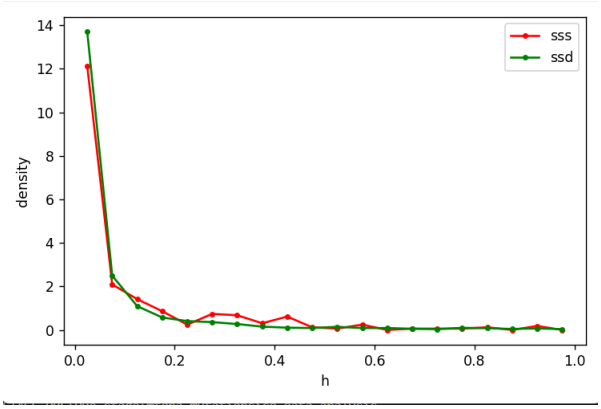


FIGURE 27 – SFS comparison when heteroplasmy reaches 1 with $\rho = 1e-5$

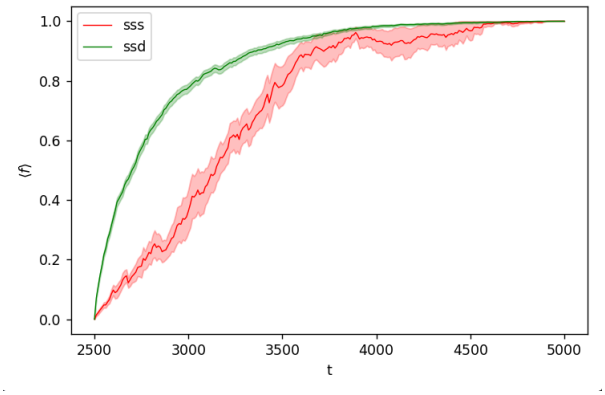


FIGURE 28 – Simulation results with $\rho = 1e-5$

The rank-biserial correlation coefficient is $r_b = -0.06029670096010727$, the U-statistic is $U = 718156.5$, with $n_1 = 4703$ and $n_2 = 325$, and the p-value is $p = 0.06864980775028104$. These results indicate that the distributions of fixation times are statistically different. The Mann-Whitney test, with a very low p-value and a low r_b coefficient, demonstrates that the models

are not identical. We gained preliminary insights into how the SSS and SSD models could be distinguished using the Site Frequency Spectrum (SFS).

Comparison with RA

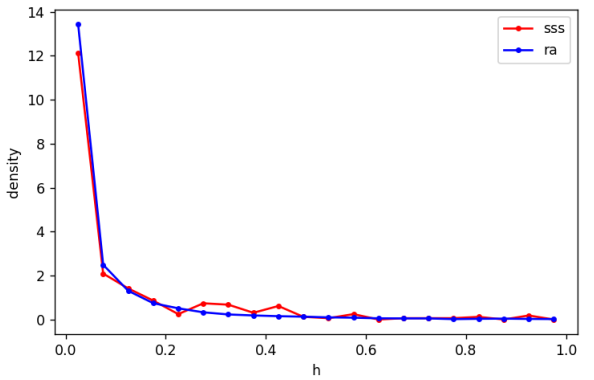


FIGURE 29 – SFS comparison when heteroplasmy reaches 1 with $\rho = 1e-5$

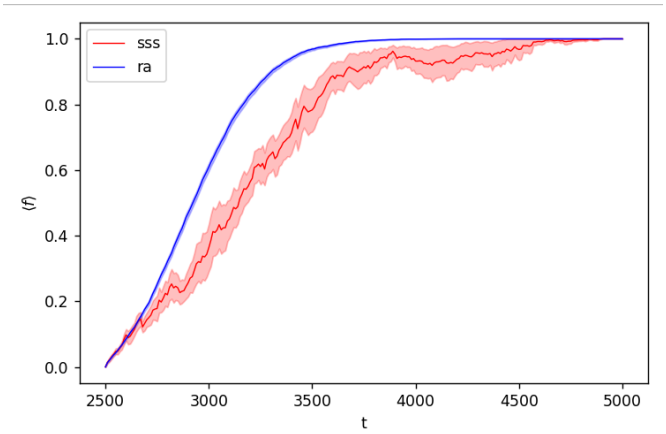


FIGURE 30 – Simulation results with $\rho = 1e-5$

The SFS are quite the same again, and a Mann-Whitney test was made and we found these results : The rank-biserial correlation coefficient is $r_b = -0.057067829898976824$, the U-statistic is $U = 2172444.9999999995$, with $n_1 = 14178$ and $n_2 = 325$, and the p-value is $p = 0.007810264732685118$.The p-value indicates a statistically significant difference between the two distributions. The rank-biserial correlation coefficient (r_b) suggests a very weak negative association. The results indicate a statistically significant difference between the two distributions. Therefore, it can be a good way for the biologists to distinguish both models and understand how the mutants behave : if it's either SSS or RA.

Little analysis

Due to time constraints and the fact that running numerous simulations requires extended computational time, I was unable to determine a more optimal value for the rate difference that would result in a closer time to fixation for SSS compared to SSD and RA. Additionally, as shown in Figures 30 and 28, the error bars for the SSS model are considerably larger. This suggests that in the SSS model, it is more challenging for mutants to reach fixation, indicating that a significantly higher number of simulations would be necessary to observe mutants consistently achieving fixation.

Moreover, I had to use a very low value of ρ to accelerate the fixation of SSS mutants, bringing their average time to fixation closer to that of the SSD and RA models. This outcome was unexpected. Initially, it was anticipated that increasing ρ would slow down the mutants but enhance the dynamics of mutant population growth. This is an area that requires further investigation in future work, particularly to thoroughly assess the impact of the rate difference, ρ, on the model (especially considering that the drift term of the SDE must remain positive).

7 Conclusion

This study has allowed us to highlight new stochastic models that could explain mtDNA deletions. Although Replicative Advantage (RA) is the most well-known mechanism, our findings indicate that it alone cannot account for these deletions due to the inconsistency in wave speed data. Therefore, it was crucial to explore alternative mechanisms. Here, we have introduced two new models : Stochastic Survival of the Densest (SSD) and Stochastic Survival of the Slowest (SSS), which can potentially explain these mutations. Our work aimed to identify which of these models best explains mtDNA mutations in aging cells. This is particularly important as mtDNA mutations are linked to age-related diseases such as sarcopenia.

We discovered that even when mutants are slower, they can still expand and take over the entire population (SSS). These models can be distinguished using the Site Frequency Spectrum (SFS), which can be derived from the frequency of mtDNA point mutations at a given time. This method also allows for the quantitative distinction of the causes of mutation accumulation.

When comparing SSD to RA, the results show significant p-values, indicating that SFS can distinguish between these models. Building on the work done with a chain of units [13], future research could explore these models under more realistic experimental constraints by analyzing the SFS at the wavefront of observed deletion mutant takeovers within a chain of units. This approach could also help us understand why there is such a difference in the SFS of SSD and RA within these chains. If the SFS remains jagged under more realistic parameters, we might conclude that SSD is not a suitable model for explaining how deletion mutants spread across muscle cells. This necessitated testing another model, SSS, to see if similar results are observed when comparing SSS with SSD and RA.

Although I did not have the time to simulate a full chain of units due to time constraints, I managed to obtain some results using two units. Even in this simplified scenario, significant p-values were observed, suggesting that SFS can be an effective tool for distinguishing these models. However, this case is not realistic, as only two units were used instead of a chain, which would more accurately represent a muscle fiber. Future research should focus on implementing a chain of units to determine whether the SFS remains jagged and, consequently, whether SSS should be rejected as a viable model.

This SFS analysis can also be applied to other models. Future research can narrow down the potential mechanisms underlying mtDNA mutation accumulation that were previously hypothesized but not fully supported by quantitative evidence.

13. Check the Appendix 3

8 Computations

Transition to High-Performance Computing for SFS Simulations

To effectively utilize the Site Frequency Spectrum (SFS), a substantial number of simulations is required. Consequently, I was no longer able to rely on the Python scripts I had used in the previous sections. At this stage of the internship, it became essential to leverage the computing power of the Imperial server, which provided access to fast and efficient hardware for running new simulations.
Additionally, preliminary work had been carried out by another member of the research group. I needed to take over their code, understand its functionality, and then adapt it to suit my own research needs. This task proved to be long and challenging, as it involved learning to correctly use a high-performance computing (HPC) server and modifying the existing code to fit my specific requirements.
Despite these challenges, once I mastered the use of the HPC server and successfully adapted the code, I was able to achieve rapid results. For instance, using my own Gillespie algorithm, simulations for the 2-unit SSS model took approximately 30 minutes to run. However, with the HPC server, the same simulations were completed in just 3 minutes. This significant improvement in computational efficiency allowed me to conduct more extensive analyses and obtain results much faster.

Codes

Finally, you can find all the codes in this Github page.

9 Appendix

9.1 SDE section 4.1.2

To simplify the analysis of our system, we can derive an approximate stochastic differential equation (SDE) for just one species. This process, known as *stochastic dimensional reduction*[14], leverages the fact that the system tends to fluctuate around the condition $\lambda(w,m) = \mu$, or equivalently around configurations where $w + \delta m = N_{ss}$.

If we look at the four reactions cited previously without a higher degradation rate, we obtain this system of SDEs :

$$dw = cw(N_{ss} - w - \delta m)\,dt + w\sqrt{c(N_{ss} - w - \delta m) + 2\mu}\,dW_1, \tag{8}$$

$$dm = cm(N_{ss} - w - \delta m)\,dt + m\sqrt{c(N_{ss} - w - \delta m) + 2\mu}\,dW_2, \tag{9}$$

where dW_1 and dW_2 are independent Wiener increments (Gaussian noise with zero mean and variance dt).
Therefore, with the reduction we can obtain this final SDE.

$$dm = \frac{2(1-\delta)\mu}{N_{ss}} m\left(1 - \frac{\delta m}{N_{ss}}\right) dt + \frac{1}{N_{ss}}\sqrt{2m\mu(N_{ss} - \delta m)(N_{ss} + m - \delta m)}\,dW.$$

.

If we look a bit closer at this equation, we can see a drift term (the first one), which has to be positive. And the second one is noise term with dW being a Wiener increment.
We will adapt this method, to the model with a higher degradation rate and it gives us the SDE for the SSD higher degradation rate model in section 4.1.2. Also, we used it to find the drift for the SSS model.

9.2 Wave Speed Results

This section summarizes the findings of F. Insalata and N. Jones[15]. Their analysis reveals that the RA model no longer best represents the observed wave speed of mutants. Instead, the SSD model more accurately aligns with experimental observations, highlighting its relevance for explaining mtDNA deletions.
Professor Jones's group found that in the SSD model, the wave speed of mutants closely matches observed data, showing proportionality to the inverse of the copy number N_{ss}. The figure below illustrates the spatial profile of mutant fractions over time, showing wave-like expansion.

14. The theory is explained in the section *Using stochastic models*.
15. Source : Stochastic Survival of the Densest and Mitochondrial DNA Clonal Expansion in Ageing, 2022

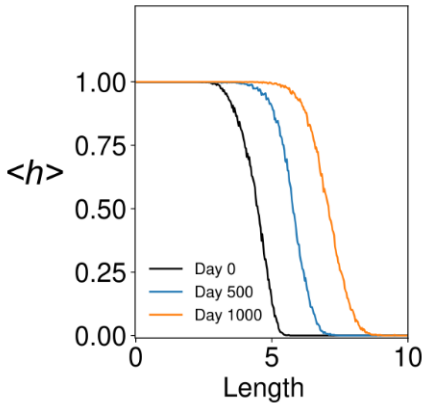


FIGURE 31 – Noise-induced traveling wave of the denser species in a chain of units, with hopping (diffusion) between nearest neighbors

The x-axis represents the position along the chain, and the y-axis shows the average mutant fraction $\langle h \rangle$. The plot demonstrates how mutant distribution evolves over time with three distinct time intervals.
Though I did not contribute directly to this part of the study, the results underscore the importance of the SSD model in predicting wave speeds. Experimental data reveals a sigmoid-shaped spatial profile, indicative of wave-like expansion. While the RA model predicts a wave speed of approximately 40 μm/day, the SSD model's prediction is around 0.2 μm/day, which is more consistent with observations.
The simulated wave speed, given by :

$$v \approx 2\sqrt{kD}, \tag{10}$$

where D is the diffusion coefficient and $k = \sqrt{(1-\delta)^2 \mu\gamma} / N_{ss}^{2/3}$, aligns with the Fisher-Kolmogorov wave-speed formula. This formula reflects an effective selective advantage for mutants in the SSD model.

Probability distributions for wave speeds predicted by SSD and RA models were compared. The SSD model better reproduces observed wave speeds, whereas the RA model predicts speeds two orders of magnitude faster. Thus, the SSD model proves to be superior in mimicking the observed wave dynamics.

9.3 Muscle fibre-Chain of units (SSD vs RA)

Experimental data provided us with information on the number of cells to use in our simulation. Specifically, we had data on a particular section along the fiber that exhibits a saturated mutant fraction.

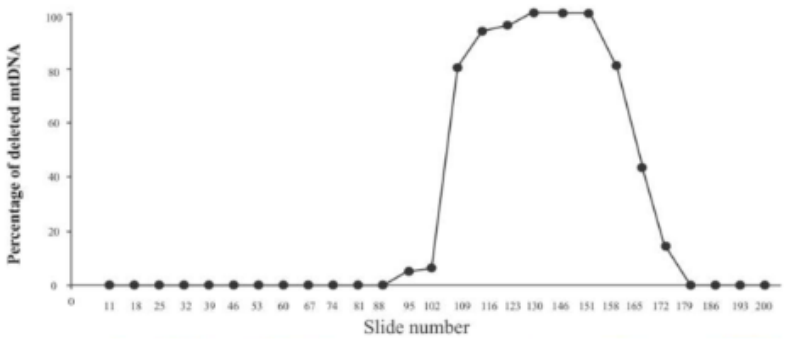


FIGURE 32 – Number of deletions in a muscle fiber according to the American Journal of Human Genetics

Based on these data, we chose to model our muscle fiber with a chain of one hundred units. As before, the first step involves determining the time at which we can introduce deletions. For this simulation, we selected an introduction time of 10,000 days.

Matching Wave Speed

As observed in Appendix 2, there is a notable difference in the traveling wave behavior between the RA and SSD models. To understand these differences, we analyze the expressions for the wave speed in the SSD model and the parameter k_{ra}.

In this context, the diffusion coefficient D can be defined as $D = \gamma L^2$, where $L = 1$. To match the wave speed of the RA and SSD models, the value of k_{ra} must be calculated using the following expression :

$$k_{ra} = \frac{\sqrt{(1-\delta)^2 \mu \gamma}}{N_{ss}^{2/3}} \tag{11}$$

Additionally, to ensure that the carrying capacity is equivalent in both models, we determine the value of $N_{ss,RA}$ using equation 6 from our study. The results of these calculations are presented below.

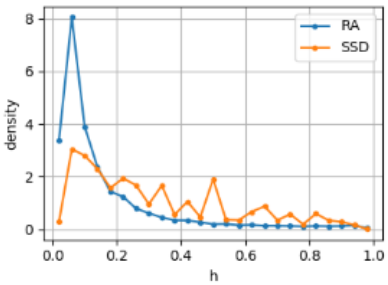


FIGURE 33 – SFS with 100 units - *Results obtained by another member of the research group*

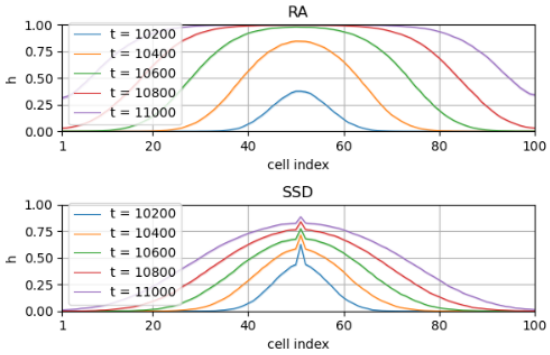


FIGURE 34 – Wave propagation with RA and SSD models - *Results obtained by another member of the research group*

The phenomenon of wave speed is also evident in the Site Frequency Spectrum (SFS) analysis. We can observe a traveling wave in the simulations. Notably, there are differences in the wave profiles, particularly in the center of the wave. In the SSD model, the wave has a steeper bump because deletions occur more slowly compared to the RA model.
Furthermore, it is apparent that the wave speed in the RA model is significantly faster. This discrepancy in wave speed impacts the SFS, where the SSD model exhibits more fluctuations, while the RA model shows a smoother distribution. These observations are crucial for understanding the dynamics of muscle fiber simulations and the implications for biological studies.

Matching Carrying Capacity

In this section, we apply the same method as explained in the model with two units, where the mutant carrying capacity is five times that of the wildtype carrying capacity.

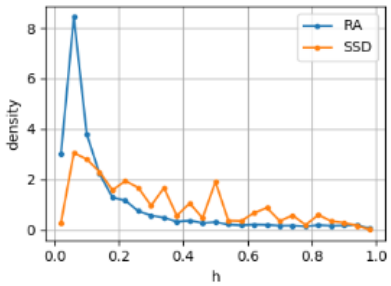


FIGURE 35 – SFS with 100 units - *Results obtained by another member of the research group*

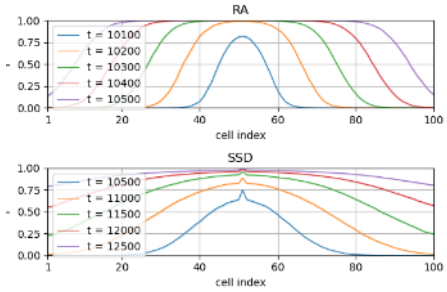


FIGURE 36 – Wave propagation with RA and SSD models - *Results obtained by another member of the research group*

As depicted in the figures, the time required for mutant deletions is longer in the SSD model. This observation aligns with experimental data, which indicate that the wave speed in the RA model is considerably faster than in the SSD model [16]. This difference is also reflected in the Site Frequency Spectrum (SFS) plots. Specifically, there is a drop in frequency at low h values, similar to the system with two units. However, with 100 units, there are more point mutations, resulting in a more pronounced drop.

Additionally, the SFS for the RA model is smoother, while the SFS for the SSD model appears jagged. This phenomenon has not yet been fully explained and requires further research.
Despite this, a Mann-Whitney test was conducted, yielding a p-value of $p = 1.37 \times 10^{-104414}$, indicating a significant difference between the two models. Thus, the SFS remains a robust statistical tool for distinguishing between the models.

However, future research could imply using more realistic parameters and if we do so and still see a jagged SFS for the SSD model, then maybe, this model can no longer be reliable and may be rejected as a possible model that could explain mtDNA deletions.

This finding is valuable for biologists, as it aids in determining which model better represents the data and helps in selecting appropriate treatments for muscle aging in specific patients.

16. For a more detailed explanation, refer to Appendix 2

10 Bibliography

[1] A. Hiona and C. Leeuwenburgh. *The role of mitochondrial dna mutations in aging and sarcopenia : implications for the mitochondrial vicious cycle theory of aging.* Experimental Gerontology, 43(1) :24–33,2008.

[2] A. D. de Grey. *A proposed refinement of the mitochondrial free radical theory of aging.* BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology, 19(2) : 161–166, -02 1997.

[3] Patrick F. Chinnery and David C. Samuels. *Relaxed Replication of mtDNA : A Model with Implications for the Expression of Disease* Am. J. Hum. Genet. 64 :1158–1165, 1999

[4] Khan Academy. *Population Growth and Carrying Capacity*. 2024.

[5] F. Insalata, H. Hoitzing, J. Aryaman, and N. Jones. *Stochastic Survival of the Densest and Mitochondrial DNA Clonal Expansion in Ageing*. Proceedings of the National Academy of Sciences, Volume 119, Number 32, Article e2203783119, 2022.

[6] J. H. Wakeley. *Coalescent Theory : An Introduction*. Roberts and Company Publishers, Greenwood Village, CO, 2009

[7] T. L. Parsons and T. Rogers. *Dimension Reduction for Stochastic Dynamical Systems Forced onto a Manifold by Large Drift : A Constructive Approach with Examples from Theoretical Biology*. Journal of Physics A : Mathematical and Theoretical, Volume 50, Number 41, Pages 435001, 2017.

[8] G. Constable, Y. Krumbeck, and T. Rogers. *An Invitation to Stochastic Mathematical Biology*. Notices of the American Mathematical Society, Volume 68, Number 11, Pages 1842–1854, 2021.

[9] Desmond J. Higham. *Modeling and Simulating Chemical Reactions*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2008.

[10] E. Bua, J. Johnson, A. Herbst, B. Delong, D. McKenzie, S. Salamat, and J. M. Aiken. *Mitochondrial dna-deletion mutations accumulate intracellularly to detrimental levels in aged human skeletal muscle fibers*, American Journal of Human Genetics, 2006