

**SPATIAL GENETIC DEMIXING IN MICROBIAL COLONIES
WITH GROWTH DELAY**



A thesis submitted towards the partial fulfillment of
BS-MS dual degree programme
from JANUARY 2020 to DECEMBER 2020

by

VISHRUT PATEL

BSMS REG. No. 20141022

under the guidance of

DR CHAITANYA ATHALE

AFFILIATION OF THE SUPERVISOR

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH
PUNE

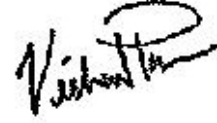
Certificate

This is to certify that this dissertation entitled "Spatial genetic demixing in microbial colonies with growth delay" submitted towards the partial fulfillment of the BS-MS degree at the Indian Institute of Science Education and Research, Pune represents original research carried out by Patel Vishrut Yogesh at Indian Institute of Science Education and Research Pune, under the supervision of Dr Chaitanya Athale from 02-Jan-2020 to 20-Dec-2020.



डॉ. चैतन्य अथले / Dr. Chaitanya Athale
आसोसो प्रोफेसर / Associate Professor
भारतीय विज्ञान शिक्षण संशोधन संस्थान
Indian Institute of Science Education & Research
Pune / Pune - 411 003, India

Supervisor:
DR CHAITANYA
ATHALE
ASSOCIATE
PROFESSOR,
DIVISION OF
BIOLOGY, IISER
PUNE, INDIA



PATEL VISHRUT
YOGESH
ROLL No.
20141022
BS-MS
IISER PUNE

DATE:
14/01/2021

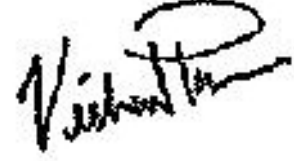
Declaration

I, hereby declare that the matter embodied in the report titled “Spatial genetic demixing in microbial colonies with growth delay” is the results of the investigations carried out by me at the Indian Institute of Science Education and Research, Pune from the period 17-01-2020 to 02-01-2021 under the supervision of Dr Chaitanya Athale and the same has not been submitted elsewhere for any other degree.



डॉ. चैतन्य अथले / Dr. Chaitanya Athale
आसोसो प्रोफेसर / Associate Professor
भारतीय विज्ञान शिक्षण संशोधन संस्थान
Indian Institute of Science Education & Research
पुणे / Pune - 411 003, India

Supervisor:
DR CHAITANYA
ATHALE
ASSOCIATE
PROFESSOR,
DIVISION OF
BIOLOGY, IISER
PUNE, INDIA



PATEL VISHRUT
YOGESH
ROLL NO.
20141022
BS-MS
IISER PUNE

DATE:
14/01/2021

Acknowledgements

I would like to express my deepest appreciation for my supervisor and mentor, Dr Chaitanya Athale, who has been a continual source of inspiration, insight and motivation throughout this project. Discussions with him were essential in keeping the spirit of scientific inquiry alive in me, especially during the time of a global pandemic. Without his guidance and persistent help this thesis would not have been possible.

I would like to thank Prof Deepak Dhar for sharing his knowledge and insight on some theoretical aspects of the study. I would like to thank the members of the CyCelS Lab for being a source of abundant help and constructive critique which was vital in culmination of this project.

I am extremely grateful for my friends: Reema, Danish, Suman, Shashank, Nandini and Anjali, whose support for me has been with me not just through this project but my time at IISER Pune in its entirety. I also to say thanks to everyone who played make-believe with me and some dice. This appreciation extends to Critical Role and other members of the TTRPG community at large.

Finally, I would like to acknowledge various members of IISER faculty, administration and support, including but not limited to the Dean, the Medical Committee, the Fifth Year Thesis Committee, the IT department, the hostel staff. They have been key in providing the infrastructure and environment within which pursuits of science of mine and others are possible without many hurdles.

Abstract

A microbial colony population expanding in two-dimensions is subject to greater genetic drift compared to well-mixed populations. Time-series fluorescence images of bacteria *Escherichia coli* and budding yeast *Saccharomyces cerevisiae* colonies show that during the course of colony growth, a well-mixed population of two fluorescently labeled strains, segregates into well-defined sector-like domains with fractal boundaries (Hallatschek *et al.*, 2007). The dynamics of these domain boundaries are responsible for the sectoring pattern at large. As the colony grows, these domain boundaries do not propagate linearly, as the cells with different genotypes on either side of the domain boundary undergo neutral competition at this interface. These domain boundaries in the sectoring pattern are akin to trajectories of random walkers undergoing anomalous diffusion. Through previous research, we note that different microbial colonies, such as that of *E. coli* and *S. cerevisiae*, differ in the amount of boundary wandering. To answer these questions, we have implemented the Eden Model, a grid based stochastic simulation in Python. This model of microbial colony growth allows us to test models that can give rise to the experimentally observed differences in boundary wandering. We propose a modification of the Eden Model that delays the cell division of newly reproduced cells the model. This is implemented as a two-state process: a newly born cell is in a nascent state for a fixed maturation time after which they reach an active state which can divide. We quantitatively show that by incorporating this growth delay into the Eden Model, the domain boundary wandering becomes less pronounced in our model and consequently, reduces the genetic drift in the growing populations. We discuss how our model integrates biological cell cycle dynamics into colony growth models to predict the genetic drift in colony populations.

Contents

Abbreviations and Symbols	3
1 Introduction	7
1.1 Spatial gene segregation and genetic drift	9
1.2 Previous work on genetic sectoring	10
2 Theory and Model	12
2.1 Sectoring dynamics	12
2.2 Eden Model (EM)	17
2.3 Eden Model with growth delay (EMwD)	19
2.3.1 Division as a first order process	19
2.3.2 Growth delay	21
3 Methods	23
3.1 Eden Model (EM) implementation	23
3.1.1 The grid structure	23
3.1.2 Growth process	24
3.1.3 Initializations	24
3.1.4 Hexagonal grid implementation	27
3.2 Eden Model with growth delay (EMwD) implementation	28
3.2.1 Growth process	28
3.3 Mean Square Deviation analysis of the domain boundary	29
3.3.1 MSD over multiple realizations	29
3.3.2 MSD over segmented windows	31
4 Results	32
4.1 Sectoring patterns in the Eden model of microbial colony growth	32
4.1.1 Number of sectors in linear and circular growth	32
4.1.2 Boundary wandering as a super-diffusive random walk	34
4.1.3 MSD over segmented windows	36
4.2 Effect of growth delay on demixing and sector boundaries	40

4.2.1	Effect of growth delay on boundary wandering	40
5	Discussion	47
	References	51
A	Source Code	53

Abbreviations and Symbols

Here is the list of abbreviations and symbols frequently used in this report and their descriptions.

Reference for abbreviations and symbols

- α Diffusive exponent (The exponent in the diffusion equation for domain boundary wandering $\sigma^2 = \text{MSD} = 2Dr^\alpha$.)
- δ Maturation time (This is the time period after which a nascent cell transitions to an active dividing cell in the Eden Model with growth delay (EMwD) formulation. This is the key parameter which we vary in our various simulations of the model.)
- $\mathcal{C}_{\text{init}}$ Circular initialization (Refers to the initial condition for the simulation space which is similar to the circular inoculation of a microbial colony. See subsection 3.1.3 for details.)
- $\mathcal{LS}_{\text{init}}$ Linear singular initialization (Refers to the initial condition for the simulation space which generates just a single domain boundary on a linear growth front of colony growth. See subsection 3.1.3 for details.)
- $\mathcal{L}_{\text{init}}$ Circular initialization (Refers to the initial condition for the simulation space which is similar to the linear inoculation of a microbial colony. See subsection 3.1.3 for details.)
- $\mathcal{SS}_{\text{init}}$ Semicircular singular initialization (Refers to the initial condition for the simulation space which generates just a single domain boundary on a circular growth front of colony growth. See subsection 3.1.3 for details.)
- τ Mean division time (This is the parameter which determines the rate of division of cells in EMwD. For the purpose of all the simulation results described in this text, assume $\tau = 1.0$.)
- D Diffusive coefficient (The coefficient in the diffusion equation for domain boundary wandering $\sigma^2 = \text{MSD} = 2Dr^\alpha$.)

EM Eden Model (Refers to the Type-C Eden Model with individuals of multiple distinct genetic identities.)

EMwD Eden Model with growth delay (Refers to modified Type-C Eden Model with a delayed growth state of cells as described in section 2.3.)

Chapter 1

Introduction

A bacterial colony growing on a solid substrate such as an agar plate with nutrient rich media has a very different population growth pattern compared to a well-mixed fluid media. The former scenario shows much more enhanced genetic drift compared to the latter even at high population sizes. This is due to the constraint that population growth is limited to the narrow perimeter of the colony which is described as the growth frontier. This leads to repeated genetic bottlenecks and segregation of the bacterial gene pools into sector-like regions of reduced genetic diversity. Experiments involving plating of well-mixed *E. coli* population consisting of CFP- and YFP-tagged strains in equal ratios and imaging under fluorescent microscopy shows genetic sectoring pattern on the scale of a few millimeters (Hallatschek *et al.*, 2007).

On the microscale, the cells are organising during the colony growth by cell growth, division and mechanical contact forces between the cells as they push against each other. These small scale interactions lead to the pattern at large. A previous study simulating bacterial colony growth *in silico* (Rudge *et al.*, 2013) indicate that mechanical instability arising due to uniaxial growth in rod-shaped cells leads to the fractal boundary between genetically segregated population regions, which shall be referred to as **domains**.

This genetic sectoring pattern also is observable in 2-dimensional models where most of the biological and physical processes responsible for the colony spread is simplified to reproduction and spatially exclusive site occupation rules. The genetic sectoring behaviour is salient enough to appear in such reductionist models, it follows that it would be a feature of any spatially expanding population, not just microbial populations.

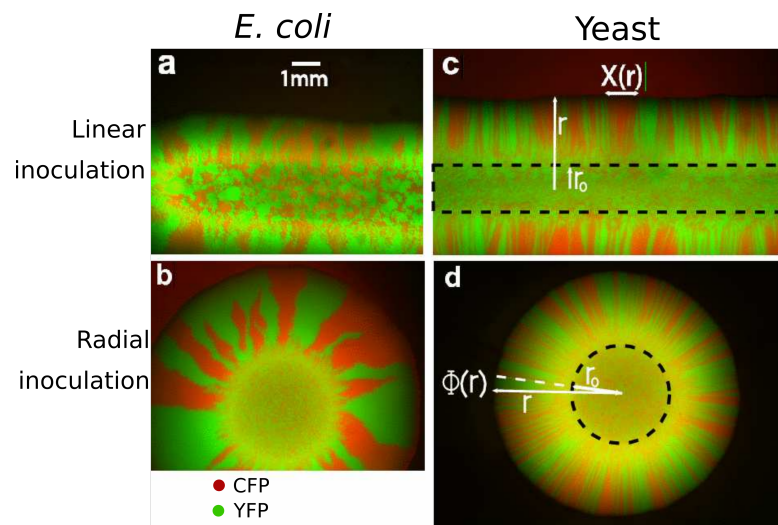


Figure 1.1: **Images reproduced from Hallatschek *et al.* (2007)**). Micrographs from fluorescent microscopy experiments in which well-mixed microbial populations with 50:50 YFP and CFP labelled strains were inoculated on agar plates. The images show that the two genetically distinct subpopulations segregate into sector-like domains with a sharp boundary separating them. Subfigures (a) and (b) depict *E. coli* colonies and subfigures (c) and (d) depict *S. cerevisiae* (yeast) colonies. (a) and (c) are linear inoculations while (b) and (d) are circular inoculations. *S. cerevisiae* colonies show less domain boundary wandering and greater number of sectors compared to the *E. coli* colony.

1.1 Spatial gene segregation and genetic drift

The genetically segregated domains which are visible under fluorescent imaging of a microbial colony inoculated with a mixture of CFP- and YFP-labelled cells, are a manifestation of genetic drift which acts at the frontier of the expanding colony. Genetic drift in a population usually refers to the variation of different genotypes in a small population, because of the randomness (in sampling) in the reproductive process, as opposed to any phenotypic differences among the individuals.¹

Although the populations in this case are not small (the initial droplet in figure 1.1 A consists of $\approx 10^6$ cells), the effective population of reproducing individuals is a small fraction of the population. Only a small number of cells at the very edge of the colony, the growth frontier, are able to pass on their genes to the next layer of outwardly spreading cells. This “genetic bottleneck” is strong enough to segregate the population into domains of genetic near-homogeneity. This is only possible because the individuals are immobile barring the mechanical outward force due to the individual cell growth and are neutral competing for space as a resource. In a well-mixed, population as a counterexample, genetic drift is usually weak, and the dynamics of genotypic frequencies is almost deterministic during the population growth. This soft constraint of 2-dimensional spread of a population is often encountered in nature, in fact the surface colonizing bacteria and archae are known to arrive fairly early in the evolutionary history of life on earth. In that regard, the genetic drift at the colony growth frontier played a major role in the evolution of primitive life on Earth.

This enhanced genetic drift can lead to disappearance or fixation of certain genotypes, beneficial and deleterious alike, in the reproducing subpopulation of a colony. In terms of population genetics, there is a decrease in the population diversity at the frontier and, consequently, a decrease in diversity in the colony population at large. From a practical standpoint, in a microbial population these effects can be important factors in affecting the evolution of antibiotic drug resistance, which is a huge challenge for us in medicine and healthcare. If we can understand what factors affect the sectoring pattern, we can perhaps have an insight into what contributes and controls the rate of loss of genetic diversity, we can perhaps be able to design drugs to target not just the microbial growth, but any factors which can reduce the chance of an antibiotic-resistant mutant proliferation and survival.

¹For our discussions throughout, we assume that there are no phenotypic differences across any individuals, which also holds for the experiments.

1.2 Previous work on genetic sectoring

There have been many attempts to model the self-organized pattern of a microbial colony, within and at large, from grid-based approaches of diffusion-limited aggregation (DLA) model (Matsuyama and Matsushita, 1993) to *in silico* three-dimensional biophysical modelling with cell-cell mechanical interactions (Rudge *et al.*, 2013). Largely, the focus had been on the spread of the colony, the branching patterns at the frontier and other mechanical or physical aspects of the such as biofilm formations.

The spatial structuring of genes in a microbial colony was initially observed by fluorescent imaging of colonies with two neutrally labelled (with GFP and YFP) subpopulations of bacteria (Hallatschek *et al.*, 2007) and since, then there have been many studies to elucidate this phenomenon. Theoretical studies such as (Saito and Müller-Krumbhaar, 1995), (Hallatschek and Nelson, 2008) and (Korolev *et al.*, 2011) have found multiple consistent descriptions of this phenomenon. Computer simulation attempts to reproduce the qualitative genetic sectoring pattern have used stepping-stone model, off-lattice colony growth models (Korolev *et al.*, 2011), metapopulation models with demes on a lattice (Gralka *et al.*, 2016) and variations on Eden Model (Hallatschek and Nelson, 2008) (Kuhr *et al.*, 2011). Of all these models, Eden Model is the simplest and more intuitively aligns with our analogue for non-motile microbial colony on hard agar plates, compared to others which have cell migration in space. All the models, show genetic sectoring, but by the nature of Eden Model, the monoallelic domain boundaries are exact making it more amenable to the boundary wandering analysis that we wished to do.

Eden Model serves as the starting point of the study as the coarsest model of the system we wish to understand. The Eden Model, first described by Murray Eden in 1961 to study the biological growth pattern (Eden, 1961), is a stochastic growth process on a two-dimensional (square) lattice. While the original study only focused on the overall spread of the colony, I present a many-allele variant of the Eden Model (EM) which is used to understand the genetic patterning within the colony. Despite being lattice-based and a very simplified description of reproduction and spread, it matches well with the essential aspects of the fractal genetic patterning that happens due to a narrow reproducing frontier in a two-dimensional range expansion. It has previously been considered to study spread of deleterious mutation (Kuhr *et al.*, 2011) but here we consider neutral competition only.

While, the sectoring pattern has been well-described and reproduced by many avenues, little has been known as to how biological factors at the cellular level, such as cell shape, growth rate, division rates, etc. factor into.

Consider figure 1.1; contrasting the genetic sectoring patterns of *E. coli* and *S. Cerevisiae*, we can see that the domain boundary wandering in the yeast colonies is observably less than that in the bacterial colonies. This shows that the spatiogenetic sectoring in microbial colonies although potentially universal, show appreciable diversity. Previous research focusing on the mechanical interactions within the colony (Farrell *et al.*, 2017), suggest that the difference in the cell shape for the rod-shaped *E. coli* and the quasi-spherical *S. Cerevisiae* can be one of the factors that differentiate the spatiogenetic patterns of their respective colonies. Another difference between the two microbes is the cell division rate and I honed my focus on this aspect to see if it results in the desired effect on the sectoring pattern. In this study, I hope to model a transition from the excessive domain wandering pattern of *E. coli* to the almost spoke-like sectors in yeast colonies, by modulating the division process at the level of an individual.

I shall start by describing the genetic sectoring in colony growth in terms of the wandering and colliding domain boundaries surrounding the monoallelic domains in the colony, a theoretical outlook presented by (Hallatschek and Nelson, 2008). This simplifies the problem by mapping the two-dimensional sectoring pattern to a one-dimensional fractal boundary which is akin to random walk, whose spread relates to the roughness of the front of the colony. I, then, will describe EM with multiple genetic identities and how the simulations for this growth process compares to our experimental observations.

After which, I shall propose a modified version of EM, in which implement the condition that the cell which just got produced cannot immediately reproduce but must mature over a time period before being able to divide. I showed that increasing this maturation period does indeed reduce the domain boundary wandering qualitatively. To quantify this effect, I focused on two main simulation initial conditions Linear singular initialization ($\mathcal{LS}_{\text{init}}$) and Semicircular singular initialization ($\mathcal{SS}_{\text{init}}$), both of which only have two monoallelic domains and a single boundary separating them. This allows us to quantify the wandering of a single domain boundary at a time as a diffusive random trajectory. It was found that the additional maturation time in the cell growth step pronounceably decreased the domain boundary wandering and genetic drift in the system. I shall finally discuss how the maturation time affects the colony frontier roughness and implications of my findings.

Chapter 2

Theory and Model

2.1 Sectoring dynamics

The key idea at play in how we interpret the sectoring dynamics as part of this study is to look at the erratic path of the domain boundaries or the sharp contour separating the domains.

Let us define the domain size as the distance, in the case of a linear inoculation, and the separation angle, in the case of circular inoculation, between the boundaries encompassing the domain. In a deterministic model of growth, a genetic sector or domain should remain of a constant size, except for the circular growth case where the sector angle grows as the radius of the colony. But if we let stochastic variations to play a role, the sectors will inevitably show size variations as there is neutral competition taking between distinct genetic type variants at the boundary separating the two domains. Throughout this study, the assumption that there is no difference in the reproductive success of the individuals in the population regardless of their genotype (their colour). Hence, the variation in the domain size is completely due to neutral genetic drift.

The phenomenological outlook we take to analyse is this variation dynamics is the one developed by (Hallatschek and Nelson, 2008). I will summarize this approach and list the key results relevant for interpreting the quantitative results that I will present later.

An illustrative explanation of their model can be seen in the figure 2.1. Each domain is described by the pair of random walkers (the tips of the encompassing domain boundaries where they meet the growth front). Their trajectories in space and time are sufficient to describe the dynamics of the particular domain. Since these walkers are constrained to the growth front, this effectively maps the problem of sectoring dynamics in two dimensions to

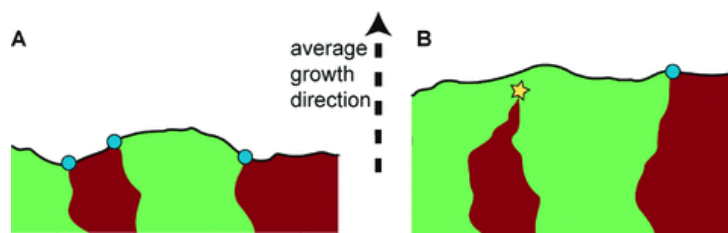


Figure 2.1: *Illustration reproduced from Hallatschek and Nelson (2008)* A visualization of how the sectoring pattern in a microbial colony coarsens over time. A. As the colony growth front (black contour) is moving with an average upward velocity, there are four domains in contact with the front. The three domain boundaries can be seen as the trail of the tips of the domain boundaries (blue circles). B. The second image depicts the subsequent snapshot after some time. By chance, the two domain boundary tips (encompassing the red domain on the left), consequently annihilate each other (at the yellow star). The enclosed domain (red domain on the left) is pinched off from the growth frontier and, henceforth trapped in the bulk of the colony. The two green domains merge to form a larger single domain. We model these dynamics by the tips of the boundaries performing a random walk along the the one-dimensional contour of the growth front with the self-annihilating (or coalescing, in case of three distinct domain colours) behaviour.

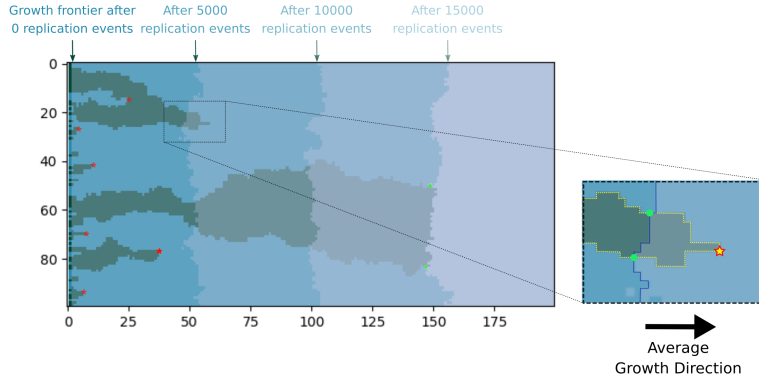


Figure 2.2: Analogous illustration of figure 2.1 overlaid and mapped onto an instance of an EM simulation output. The image shows the progression of the linearly inoculated colony with two genetic variants initialized at random positions at $x=0$ over the course of 1.5×10^4 divisions. Red stars indicate some of the collisions of neighbouring domain boundaries. In the zoomed inset image, a more clear delineation of the growth frontier after 5000 divisions (blue line) and the domain boundaries (yellow lines). The subsequent boundary collision is marked by the yellow star. A unit distance in x-y directions is the dimension of a single cell or grid site.

a simpler one-dimensional description tenable to analytical techniques.

Let us first consider the case of linear inoculation. The analysis starts by assuming that the separation between a pair of domain boundaries (i.e. the domain size) is a continuous random variable $X(r)$ where r marks the location of the growth front and X fluctuates as the separation between two independent random walks with diffusive coefficient D_X . Implying, if the front has the average displacement of $\Delta r = r - r_0$, then the associated increment $\Delta X = X(r) - X(r_0)$ has mean 0 and a variance linearly proportional to Δr . That gives us,

$$\langle \Delta X \rangle = 0; \quad \langle \Delta X^2 \rangle = 4D_X \Delta r \equiv \sigma_X^2(r) \quad (2.1)$$

A domain is closed off when the domain boundaries collide or the domain size reaches 0. On an average, the domain size at r should be comparable to $\sigma_X(r)$, thus as long as the habitat span L remains larger than $\sigma_X(r)$, it is likely that the domain has survived and remains in contact with the growth front. Thus, the effective number of surviving domains $N(\Delta r)$ at

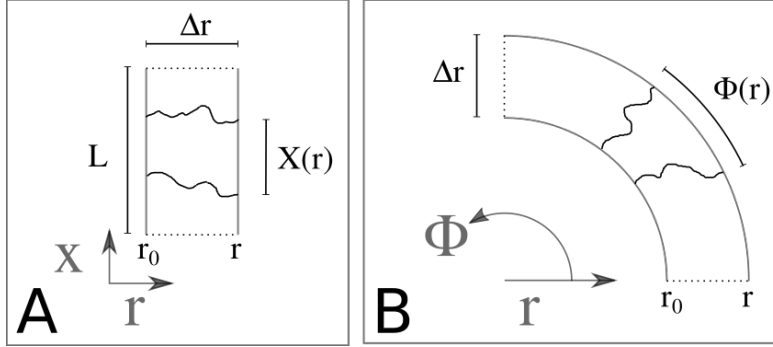


Figure 2.3: Summary of all the measurable discussed in section 2.1. A. Illustration of colony growth front with linear inoculation, with an example domain with separation $X(r)$ when the front is at r . L is the habitat span in the direction perpendicular to the growth direction. B. Illustration of colony growth front with linear inoculation, with an example domain with angular separation $\Phi(r)$ when the front is at radius r .

effective time r decreases as Δr increases, eventually leading to the fixation of a single domain at the growth front, given enough growth time. The crude but informative relation is,

$$N(r|r_0) \propto \frac{L}{\sqrt{D_X \Delta r}} \quad (2.2)$$

Let us now come to the case of a circular inoculation. As the growth front is curved we have to consider the domain sizes in terms of the subtended from the centre of the inoculation. This is parametrization can be done as,

$$\Delta \Phi \equiv \frac{\Delta X}{r} \quad (2.3)$$

This change of variables means that the angular diffusive coefficient D_Φ is no longer a constant and varies with r as follows.

$$D_\Phi(r) \equiv \lim_{\Delta r \rightarrow 0} \frac{\langle \Delta \Phi^2 \rangle}{4 \Delta r} = \frac{D_X}{r^2} \quad (2.4)$$

Thus, from a colony growth from r_0 to r gives rise to a mean square angular displacement of

$$\sigma_\Phi^2 \equiv \langle \Delta \Phi^2 \rangle = 4 \int_{r_0}^r dr' D_\Phi(r') = 4 D_X (r_0^{-1} - r^{-1}) \quad (2.5)$$

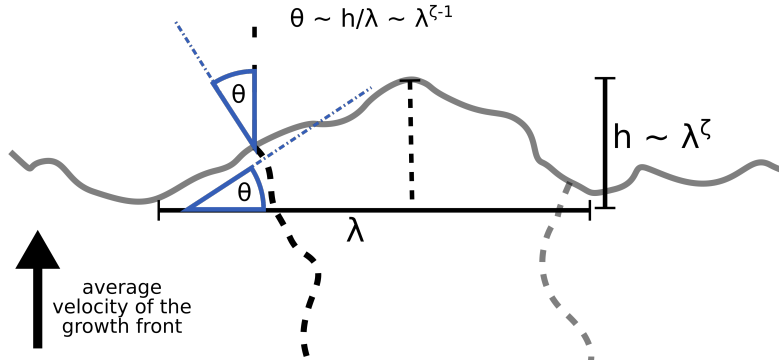


Figure 2.4: An illustration of how roughness of the growth front can lead to a tilt in the instantaneous velocity of domain boundary tip walker.

This consequence of this, in contrast with the linear inoculation case, is that the mean square angular displacement converges to a finite value of $\frac{4D_X}{r_0}$ implying the sectoring stabilizes as the colony grows. In fact, we expect a finite number of domains surviving at long times $N(\infty|r_0)$,

$$N(\infty|r_0) \propto \sqrt{\frac{r_0}{D_X}} \quad (2.6)$$

Thus far, we assumed that the growth front is smooth and progresses uniformly for across the habitat span. Now, if the growth front contour is rough, the domain boundary wandering no longer can be assumed to be diffusive. Consider that across a span of length λ if the surface roughness (the height difference from the mean position of the surface) $h \sim \lambda^\zeta$, where ζ denotes the roughness exponent of the surface. In that case, the average velocity of the domain boundary can no longer be assumed to be the same as the average velocity of the growth front as depicted in the Fig. .

In fact there will be a slight tilt θ to the mean velocity of the walker with mean 0 and root mean square of the order $\lambda^{\zeta-1}$. This lends the walker a varying drift velocity in the direction perpendicular to the mean velocity of the growth front which has a mean 0 across time. Effectively, changing the diffusive dynamics of ΔX to an analogous anomalous diffusion regime,

$$\sigma_X^2(r) = 4D_X \Delta r^{2\zeta} = 4D_X \Delta r^\alpha \quad (2.7)$$

Here, α shall be referred to as the anomalous diffusion coefficient.

Propagating this correction into equations 2.2 through 2.6, we get the following,

$$N(r|r_0) \propto \frac{L}{\sqrt{D_X} \Delta r^\zeta} = \frac{L}{\sqrt{D_X} \Delta r^{\frac{\alpha}{2}}} \quad (2.8)$$

$$N(\infty|r_0) \propto \frac{r_0^{1-\zeta}}{\sqrt{D_X}} = \frac{r_0^{1-\frac{\alpha}{2}}}{\sqrt{D_X}} \quad (2.9)$$

The quantities N (number of sectors surviving at the frontier), r/r_0 (distance of colony propagation in units of mean cell length scale) and L (colony span in the transverse direction to the expansion) can be readily measured both in experiments and simulations. By image analysis of experimentally grown colonies we can isolate domain boundaries visible in the fluorescence micrographs and perform MSD analysis to calculate the D and α values. The same can be said about the simulation outcomes. Thus, all the quantities in 2.8 serve as the points to reconcile the spatiogenic patterns found in experiments to those in the simulations.

Notably, both the results from the analysis of domain boundaries found in the micrographs of experiments done in (Hallatschek *et al.*, 2007) and the analysis of the domain boundaries of the EM simulations, which is presented later in the thesis, show the above-mentioned anomalous diffusive trajectories. Later, I shall describe the results for domain boundary dynamics in terms of Diffusive coefficient (D)¹ and Diffusive exponent (α) for the EM and the variations thereof with varying parameters. Equations 2.8 relate those quantities to the number of surviving sectors for given amount of front progression and expected fixation times.

2.2 Eden Model (EM)

In this section, I explain how the EM growth process works. The process is defined on a 2-dimensional lattice in which an individual (cell) occupies a lattice site and can reproduce another individual with the same genetic identity to one of its unoccupied neighbouring site on the lattice (Eden, 1961). For our discussion, the lattice is square and the neighbourhood considered is the Von Neumann neighbourhood i.e. the two vertically adjacent and the two horizontally adjacent lattice sites.

Any individual with an unoccupied neighbouring site can replicate and hence, is called an active site. During each growth step, one of the active site is chosen with uniform probability over all active sites and is chosen to

¹Hereinafter, the subscript X is dropped from D_X for ease of notation as the context for the usage of diffusive coefficient is obvious.

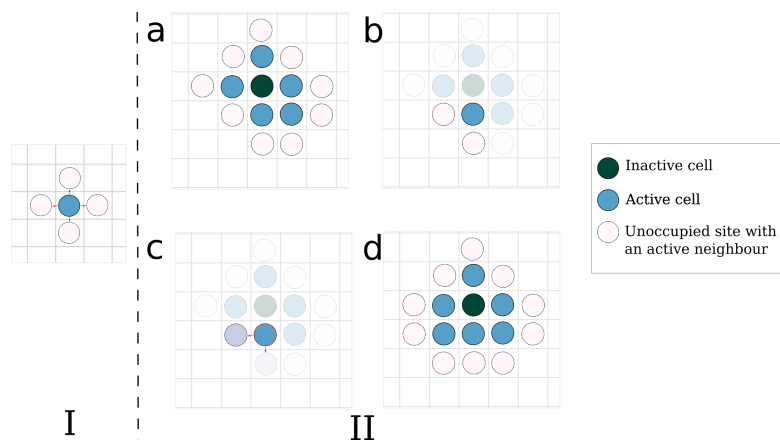


Figure 2.5: **Illustrative EM schematic.** I. A cell occupying a site can reproduce another cell to one of its four neighbouring sites. II. An illustration of sequence of substeps in a step in the growth process. (a) *Identify active sites.* An occupied site is an active site if it has one or more unoccupied neighbour, and considered inactive otherwise. (b) *Select an active site to replicate* This selection is done with uniform probability over all active sites. (c) *Select an unoccupied neighbour of the chosen active site.* Selection is done with uniform probability over the unoccupied neighbours. (d) *Reproduce a new cell to the selected unoccupied site* This completes the growth process step

reproduce to one of its unoccupied neighbouring site with equal probability over each unoccupied neighbouring site. This process was previously referred to as Type-C Eden model by Jullien and Botet (Jullien and Botet, 1985). The individuals do not die and thus, an occupied site remains so indefinitely. Thus, of the standing population only the active site individuals participate in the growth of the population which is mostly confined to the periphery of the cluster formed, hence we loosely refer to the set of active site as the frontier. Figure 2.5 describes on growth step of the process.

In our implementation, we focus on the genetic identity of individual cells. The daughter cell has the same genetic identity as the parent. We assume that there are no mutations in the process. The total number of alleles during the initialization of an instance is treated as a variable parameter.

We consider two geometry of the growth spread: linear growth and circular growth. In linear growth, a single line of active cells is initialized on the leftmost edge of the square grid and thus the process propagates the spread in left to right direction. While, for the circular growth conditions, we initialize

the active cells on an approximate circle of sites on the square grid around an inactive core. In this case, the growth process leads to an approximate circular cell cluster growing radially outward. In both case, the initial active cells are assigned their genetic character with uniform probability over the total number of alleles chosen at the parameter.

This version of the model shows a constant rate of population growth i.e. 1 per growth step. Although this does not correspond to the biological reality in which we expect the population growth rate to be proportional to the number of reproducing individuals at any given time, the actual growth cluster and the arrangement of the individuals within them do not depend on the time elapsed but the total division events that happen during the course of the simulation. Thus, for the sake of simplicity of implementation there is no consideration given to the time rate of division, but just the cluster spread and absolute number of population growth.

I will describe the results from the EM simulations which are consistent and validated by literature in chapter 4. Having said that, the range of behaviour of the EM is only defined by the EM Type-C only depends on the genetic identities of the cells and their location at the start of the simulation and does not have any other parameter barring the size of the grid/habitat.

2.3 Eden Model with growth delay (EMwD)

Figure 1.1 shows a qualitatively different levels of boundary wandering and one can also clearly see the difference in the numbers of the domains in contact with the growth front between the bacterial (*E. coli*) and the yeast (*S. cerevisiae*) colonies. I wanted to integrate one of the biological differences between the two organisms into the grid-based simulations I discussed so far and to see if that could lead to a difference in the genetic pattern if that factor is varied. With that in mind, I decided to put in a growth delay factor into the system. This modified growth model is called EMwD.

2.3.1 Division as a first order process

As I mentioned previously in 2.2 the simulation time does not correspond to any meaningful time for the population save for the subsequence of each division/birth events. To amend this, I made an assumption about the rate at which the cells divide. The division of all active cells is a first order process with the time order parameter τ . In other words, each cell with an unoccupied neighbouring site divides at constant rate and the average time for an active cell to undergo division is Mean division time (τ). Thus, an

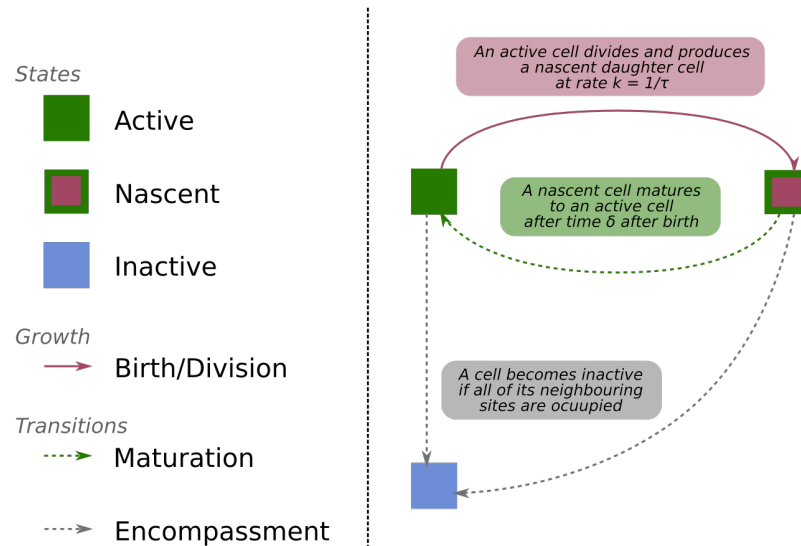


Figure 2.6: **Illustrative EM with growth delay schematic.** The main cell states and the various growth and transition process are labelled on the left of the figure. The squares denote the cell states: Active (green), nascent (green with pink inset) and inactive (blue). The right side depicts how the various processes relate the cell states to each other. The pink arrow denotes a birth/division event in which an active cell produces a nascent daughter cell at an unoccupied neighbouring site as described in 2.3.1. The green dashed arrow indicates the transition from a nascent to active cell after the maturation time δ has elapsed as discussed in 2.3.2. The grey dashed arrows indicate the encompassment of an active/nascent cell by other cells which makes it unable to reproduce and hence deemed inactive.

active cell at any given time (t_{current}) will divide after time (T_{div}) which is a random variable which follows the distribution below:

$$\mathbb{P}(t \leq T_{div} < t + dt) = \frac{1}{\tau} \exp\left(\frac{-t}{\tau}\right) dt, \text{ if } t > 0, \text{ else, } 0 \quad (2.10)$$

This assumption means that the population growth rate at any time is proportional to the number active (read reproducing cell) and, in the case of a well-mixed population, the population would grow exponentially. These two implications validate our assumption with enough scrutiny that our simplistic model merits.

If this growth rate process was included in the EM, the statistical behaviour of the spatial organization of the individuals, remains unchanged, so we can carry the observations from the original formulation of the EM to the one assuming division as a first order process, regarding the behaviour of the sectors and the domain boundary wandering.

This assumption gives our model a meaningful continuous time axis.

2.3.2 Growth delay

A bacterium recently formed from a division is less likely to divide compared to one which has had the time to harvest nutrients, grow in terms of cell volume and mass and replicate its genetic information. So, I propose to introduce a maturation time or a growth delay for a cell to model that idea.

Consider that an active cell (mother) divides to produce a daughter cell which will occupy one of the mother's unoccupied neighbouring sites. Now, I shall label this daughter cell as **nascent**. Only after Maturation time (δ) has passed after its birth will it become active and can potentially divide to add to the population. In other words, each cell division in our model is asymmetric division that produces an active cell at the original site of the mother cell and a nascent cell at one of its unoccupied neighbouring sites.

The condition for a cell to become inactive remains unchanged: an active/inactive cell becomes inactive if it no longer has an unoccupied neighbouring site i.e. it is completely encompassed by other cells. This modified model is summarized in figure 2.6.

Thus, our modified model takes two parameters to describe its growth dynamics: τ the mean division time for an active cell and δ the time a nascent cell takes to mature to an active cell. Both of these are in arbitrary time units and since, there is no other time scale defining factor in the system, the system behaviour only depends on the ratio $\frac{\tau}{\delta}$ up to a rescaling in the time axis. From this observation and the reduction of the parameter space, for all the simulations performed $\tau = 1$ was fixed. In other words, for our

purposes a unit time is equivalent to the mean division time for an active site.

Moreover, note that EM growth process is a special case of this modified model in which $\delta = 0$. If $\delta = 0$, then effectively the nascent cell state is removed from the system and each daughter cell becomes active immediately, as was the case in the EM. Thus, the growth process reverts to the EM, albeit now the process is defined on a continuous time axis instead of the discrete growth steps. It can be argued that if we let both the EM and the EMwD, but $\delta = 0$, grow until N divisions take place, any possible final configuration of the system is equally likely in both processes.

This opens up a spectrum of behaviour controlled δ of which the EM is an extreme point. Our expectation is that increasing δ for a fixed τ , the behaviour of the model will go from Eden-like to deviating from it.

Chapter 3

Methods

In this chapter I shall describe the practical computational details of the implementation of the models and the analysis of the simulations. All the programs are in `python` utilizing mainly the scientific computation packages `numpy` and `scipy`.

3.1 Eden Model (EM) implementation

For the case of the primitive EM, we consider only discrete growth steps, in each of which one cell divides and the population grows by one.

3.1.1 The grid structure

The `EdenModel` class/object functions on the main data attribute of an $N \times M$ `numpy` array with integer entries which represents the lattice in which the simulation takes place. From hereinafter, I shall refer to this array as the grid. The `numpy` array with periodic boundary conditions was implemented by subclassing the existing `numpy.ndarray` based on the `python` code provided by Alex McFarlane on their github blog post (McFarlane, 2016).

Each entry in the grid represents a lattice site indexed by the usual array indexing `'ij'` in `numpy`. The array takes integer values. An entry with a value 0 indicates an empty site. Whereas a positive integer entry represents a cell occupying the site with the genetic identity equivalent to the integer value. Thus, the genetic identities on a lattice are 1, 2, and so on. A -1 entry denotes a blocked site, which is used only in different initializations.

The adjacency on our lattice is defined a `nbhd` method which takes a site index as an argument and returns the indices of the sites adjacent to

it. In general, the adjacent neighbourhood of each site is the Von Neumann neighbourhood, i.e. the two vertically and the two horizontally adjacent sites. Though it should be noted that redefining this method can allow simulation in a different lattice configuration, such as hexagonal site grid as discussed in subsection 3.1.4.

3.1.2 Growth process

During a realization of a EM simulation, the dynamic variables are the grid itself and `active`, a list of site indices of cells which are active. The EM growth step described in is implemented as follows. At the start of a growth step, an index is chosen from with `active` with uniform probability to be divided and a daughter cell of the same genetic identity is added to one of its unoccupied neighbouring site (`open_nbhd(index)` returns indices from `nbhd(index)` with grid entry 0). Then, it is checked that the daughter cell is active and if any of its neighbours have become inactive, based on which `active` is updated. Figure 3.1 describes the algorithm flow in detail.

3.1.3 Initializations

EM simulations were performed mostly four different initializations of the grid, each defined by the grid size $N \times M$ and additional parameters.

1. Circular initialization ($\mathcal{C}_{\text{init}}$) [r_0 (initial radius), n_g (number of genotypes)]: This is similar to the circular inoculation of a microbial colony. A circle of radius r_0 centred at the centre of the grid is initialized with cells of genetic identities chosen uniformly from 1 to n_g . All the sites in the interior of this circle are blocked.
2. Circular initialization ($\mathcal{L}_{\text{init}}$) [n_g (number of genotypes)]: This is similar to the linear inoculation of a microbial colony. The first column of the grid ($j = 0$) is blocked and the second column of the grid is initialized with cells of genetic identities chosen uniformly from 1 to n_g .
3. $\mathcal{LS}_{\text{init}}$: This is used to study an isolated domain boundary wandering in linear growth front. The first column of the grid ($j = 0$) is blocked. The second column is initialized with cells of genetic type 1 and 2 in the top and the bottom half respectively. Additionally, the top and the bottom rows of the grids are also blocked.
4. $\mathcal{SS}_{\text{init}}$ [r_0 (initial radius)]: This is used to study an isolated domain boundary wandering in circular growth front. A semicircle centered at

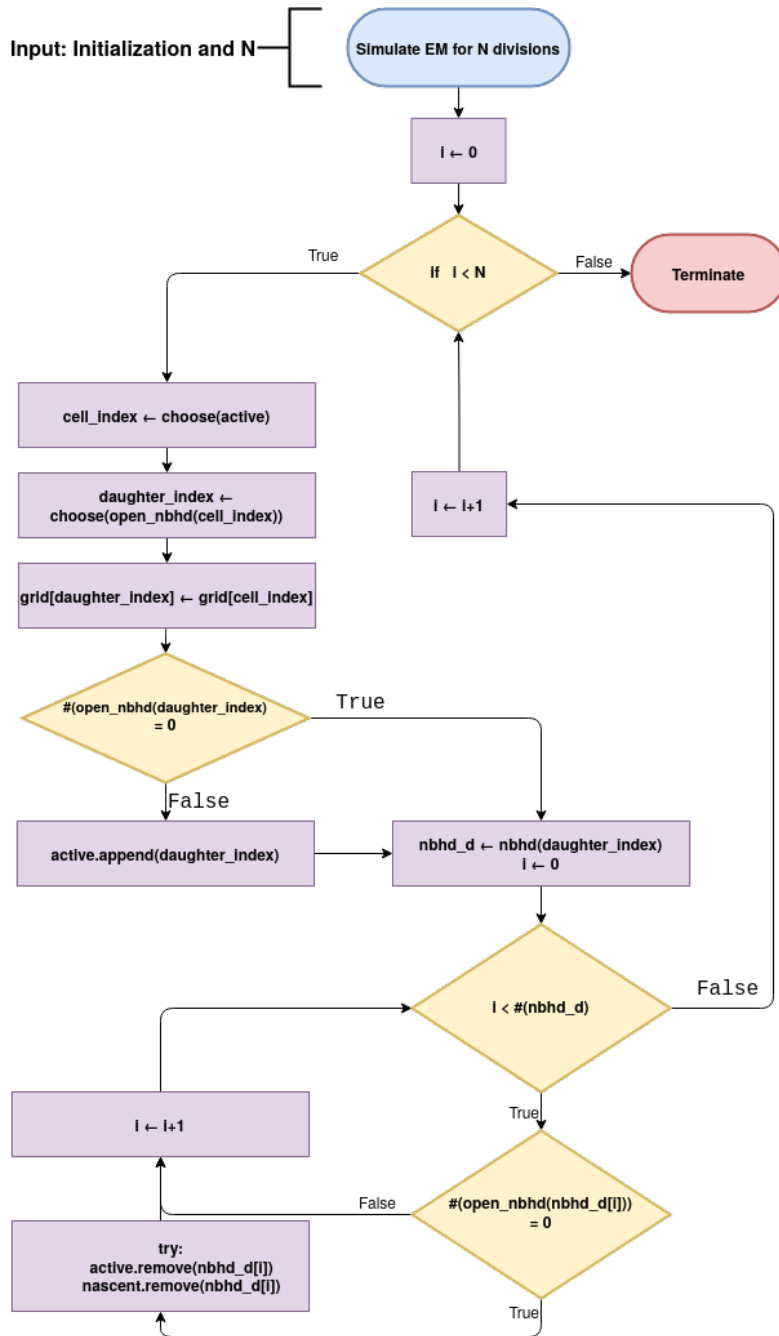


Figure 3.1: **EM implementation flow diagram for N divisions.** The rectangular blocks represent a process or a variable assignment. The diamonds represent a condition decision. The rounded rectangles are start and terminate states in the algorithm.

the midpoint of the left-most edge of the grid is initialized with the top quarter circle lined with cells of genetic identity 1 and the bottom quarter circle with those of genetic identity 2. The interior of the semicircle is blocked.

Figure 3.2 shows all the initialization and the subsequent relization after 7500 growth steps.

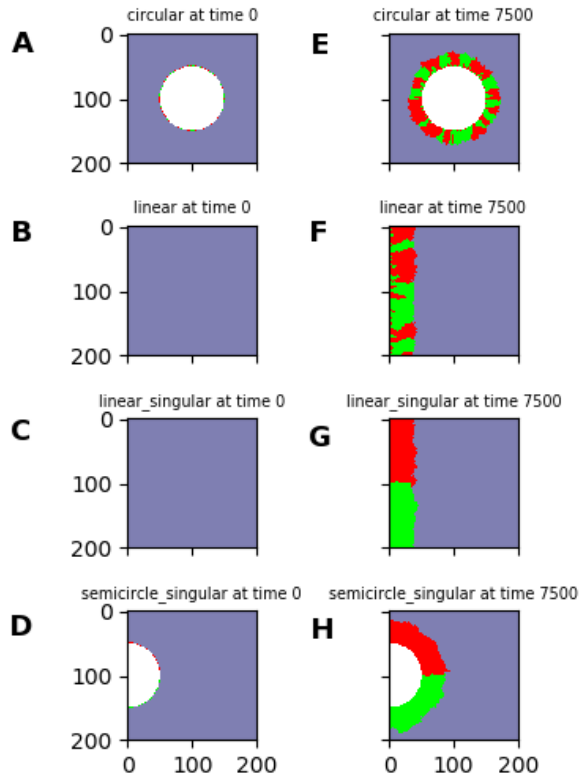


Figure 3.2: **Initializations implemented.** Subfigures A, B, C and D show the four chief initializations: Circular initialization ($\mathcal{C}_{\text{init}}$), Circular initialization ($\mathcal{L}_{\text{init}}$), Linear singular initialization ($\mathcal{LS}_{\text{init}}$) and Semicircular singular initialization ($\mathcal{SS}_{\text{init}}$), respectively, on 200×200 grid before the growth simulation has been started. Subfigures E, F, G and H show an outcome of the simulation for $\mathcal{C}_{\text{init}}$, $\mathcal{L}_{\text{init}}$, $\mathcal{LS}_{\text{init}}$ and $\mathcal{SS}_{\text{init}}$ after 7500 growth steps/divisions in the EM growth process. Red and green denote the two genetically distinct cells present in the colony. Blue denote empty site in the grid and white denote blocked sites.

3.1.4 Hexagonal grid implementation

A 2-dimensional triangular lattice is one in which each lattice point is a unit away from six neighbouring sites. We refer to it as the hexagonal grid because each lattice site can be thought of a hexagon with a neighbour sharing one of the hexagon's edge each. We can define the EM on this lattice as well, just by changing the cell neighbourhood in the growth step. A hexagonal grid version of EM or the one defined on a triangular lattice was also implemented, by redefining the `nbhd((i, j))` to return $[(i + 1, j), (i + 1, j + 1), (i, j + 1), (i - 1, j), (i - 1, j - 1), (i, j - 1)]$. This allowed the program to store the hexagonal grid in the $M \times N$ `numpy.ndarray` as usual. Except in this case, the indexing of the array does not match up to the hexagonal grid site coordinates, but are a skew linear transformation of it.

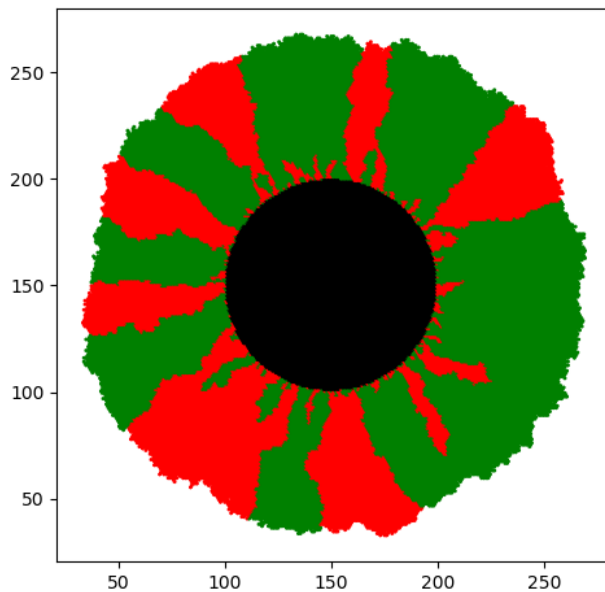


Figure 3.3: **An instance of circular colony growth simulation on a hexagonal grid.** Simulation output of circular colony growth with 2 genetic variants. The initial colony radius was $r_0 = 50$. The simulation ran for 4×10^4 growth steps. Red and green denote the two genetically distinct cells present in the colony. Black denote blocked sites.

3.2 Eden Model with growth delay (EMwD) implementation

The object class for EMwD was implemented by subclassing upon the primitive EM class `EdenModel` with the grid attribute remaining intact. This object takes the mean division time τ and the maturation time δ as parameter arguments.

The list of indices of active cells `active` is the same as in the EM, see section 3.1. To track the nascent cells, a queue consisting of tuples of nascent cell index and the global time at which it will transition to an active cell is defined as `nascent`. The global time is tracked by the variable `time`.

3.2.1 Growth process

The growth process is based on Doob-Gillespie algorithm in which the division is a first order process and the maturation events (i.e. nascent to active transitions) take place at a fixed time which are determined by the birth time of the nascent cells in question. The growth process progresses by determining which of the events: a division or a maturation event takes place first. Consider the following conditions at the start of a growth step and what a growth step looks like for each of them:

1. `num(nascent) > 0` and `num(active) = 0`

Then, a nascent to active transition will happen first. The first element in the queue `nascent` is popped and assigned to variables `cell_index` and `t_m` respectively. `cell_index` is added to `active` and the global `time` is set to `t_m`. In other words, the first nascent cell to mature is made active and the time progress to the time at which that happens.

2. `num(nascent) = 0` and `num(active) > 0`

Then, a division event will happen first. A random increment time `dt` is drawn from an exponential probability distribution with scale parameter $\frac{\tau}{N_{active}}$ where τ is the mean division time of an active cell and N_{active} is the number of active cells in the system. The global `time` is incremented to `time + dt`. Then, an active cell is chosen with uniform probability to be divided and a new daughter cell is added to the grid. The daughter cell index and the time at which the daughter cell will mature, `time + δ` is appended to `nascent`. Then, the algorithm check for any encompassment event and updates the `active` and `nascent` accordingly.

3. `num(nascent) > 0` and `num(active) > 0`

A random increment time `dt` is drawn from an exponential probability distribution with scale parameter $\frac{\tau}{N_{active}}$. If `time + dt < t_next`, the time at which the next maturation event happens, the division event happens at `time + dt` as described in case (2). Else, a nascent to active transition happens at `t_next` similar to case (1).

The algorithm flow of the EMwD is shown in figure 3.4. The initial conditions of the simulation space defined for EM remains the same for EMwD.

3.3 Mean Square Deviation analysis of the domain boundary

Domain boundary points are extracted from the grid after the simulation by marking them at the vertices shared by four sites. In other words, a domain boundary point is said to exist at (i, j) if any of the following cell pairs identified by their indices $[(i, j), (i + 1, j)]$, $[(i, j), (i, j + 1)]$ or $[(i, j), (i + 1, j + 1)]$, differ in their genetic identity. The initial conditions \mathcal{LS}_{init} and \mathcal{SS}_{init} are used to extract isolated domain boundaries. The coordinate axis is set with the leftmost boundary point as the origin and the obvious x- and y-axis are considered.

3.3.1 MSD over multiple realizations

To calculate the MSD of the boundary from its starting position in the y-axis across various realizations (N), the boundary points from the N realizations are collated into a single dataset $[(x_i, y_i)]_{i=1}^K$. Then, MSD is calculated by the following equation.

$$\text{MSD}(x) = \frac{\sum_{\forall i, x_i=x} y_i^2}{n_x} \quad (3.1)$$

where, n_x is the number of points such that $x_i = x$. This method is only used for to analyse domain boundary wandering in case of the \mathcal{LS}_{init} initial conditions since, average direction of growth for a domain boundary cannot be considered constant in case of \mathcal{SS}_{init} initial condition. Moreover, the fit value of the diffusion coefficient from the $\text{MSD}(x)$ vs x plots are not accurate. For this, the next method for estimating MSD profile of the domain boundaries is preferred.

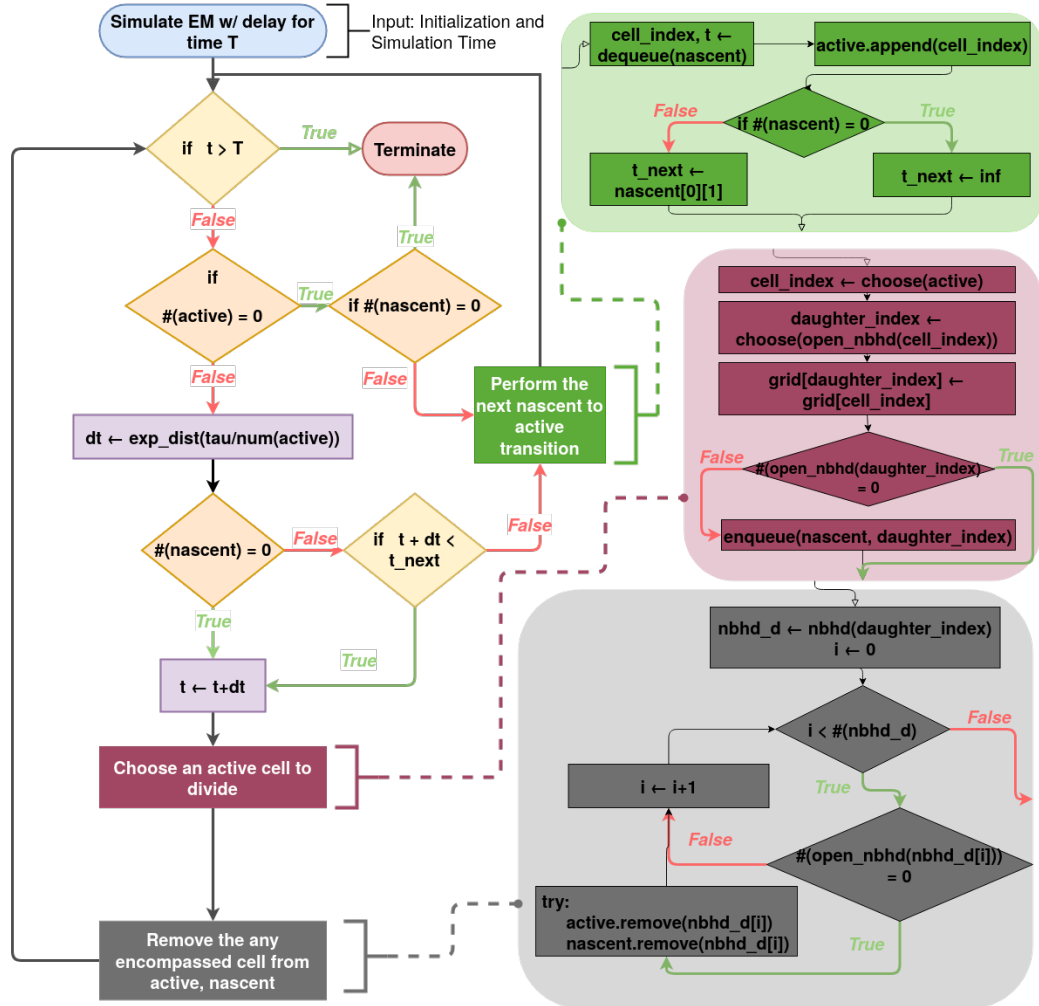


Figure 3.4: **EMwD implementation flow diagram.** The main processes of maturation event, division event and updating the active and nascent containers are shown in green, pink and grey rectangles, respectively. The substeps in each of the processes are detailed in the flow blocks connected by the dashed lines. The rectangular blocks represent a process or a variable assignment. The diamonds represent a condition decision. The rounded rectangles are start and terminate states in the algorithm. *Symbols in the flow diagram*- T: total time, dt: time-step, t: instantaneous time

3.3.2 MSD over segmented windows

Consider a domain boundary data sorted as $[(0, 0) = (x_0, y_0), \dots, (x_i, y_i), \dots, (x_N, y_N)]$ such that $x_i \leq x_{i+1}$. From this trajectory, we take a contiguous segment $Seg(p, q) = [(x_i, y_i)]_{i=p}^q$ and fit a line through it using least square regression $\Lambda_{p,q} \equiv y = m_{p,q}x + c_{p,q}$. We then consider transformed coordinates for the $Seg(p, q)$ as the x-axis along the line $\Lambda(p, q)$ and the origin as the projection of (x_p, y_p) onto $\Lambda(p, q)$. Under this transformation our segment looks like $Seg_{\Lambda}(p, q) = [(0, Y_1), \dots, (X_j, Y_j), \dots, (L_{p,q}, Y_{q-p})]_{j=1}^{q-p}$. Here $L_{p,q}$ is the length of the segment in the transformed coordinates.

$$MSD(p, q) = \frac{\sum_{j=2}^{q-p} Y_j^2 \cdot (X_j - X_{j-1})}{L_{p,q}} \quad (3.2)$$

For our analysis, we calculate the MSD for segments of length $L = l, 2l, \dots, m_{\text{cutoff}}l$ such that $m_{\text{cutoff}}l$ is approximately 60 to 70% of the total trajectory length.

Chapter 4

Results

In this chapter, I shall summarize my findings from the simulation of Eden model with cells with distinct genetic identities and my modification of the Eden Model with incorporates a delay in cell division after the birth of a new cell in the system.

4.1 Sectoring patterns in the Eden model of microbial colony growth

We started this investigation of the genetic patterning in the two-dimensional expansion of a microbial colony with the EM with the specification that the cells now carry a genetic identity or label which they pass on to their offspring. The growth rule is agnostic to this genetic label and thus, phenotypically all the cell agents in the system are identical. Hence, it is surprising at first glance, that this version of the Eden model produces genetic sectoring patterns quite similar to those observed in experiments with two strains of neutrally labelled microbes. The simulation outputs from figure 4.1(b) is fairly comparable to those of figure 1.1.

In retrospect, it is obvious that the genetic sectoring would be inevitable as all the ingredients for it are there; neutral competition, genetic bottleneck due to small reproducing population at the growth frontier and the stochasticity of the growth rule are sufficient factors to capture this behaviour in simulations.

4.1.1 Number of sectors in linear and circular growth

In figure 4.2(a), it can be seen how quickly the domain boundary collisions lead to reduced genetic diversity in the population. In the case of linear

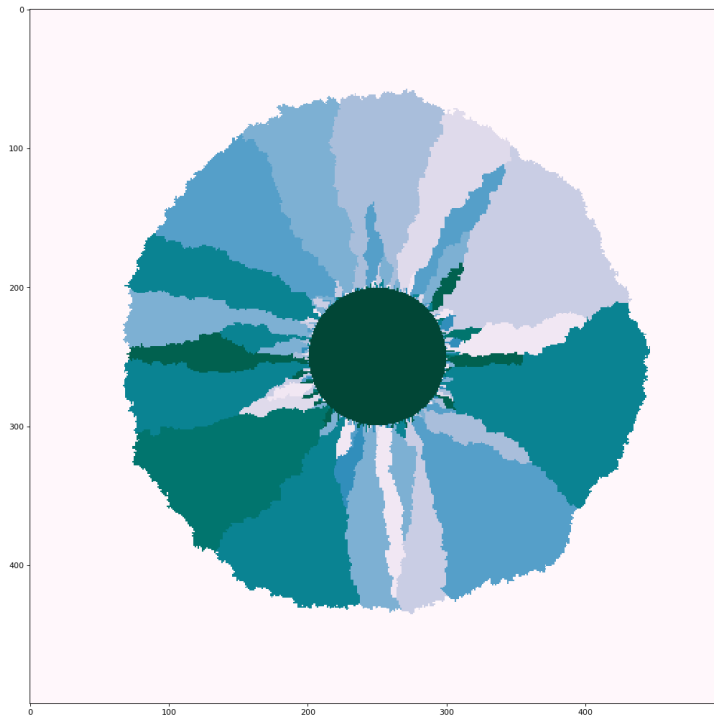
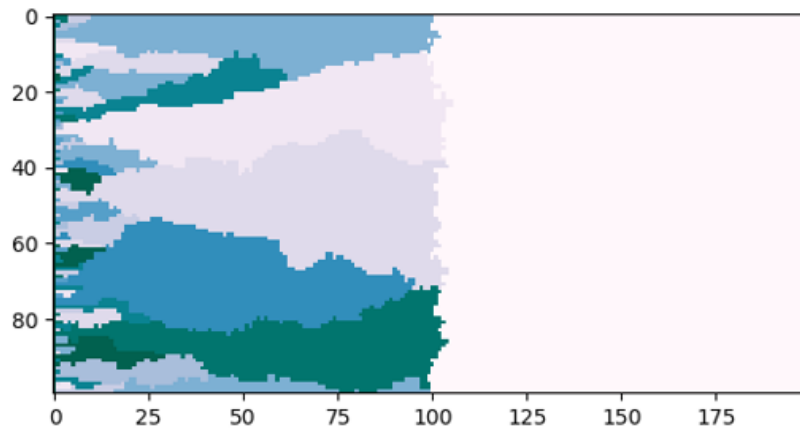


Figure 4.1: Subfigure (a): **Simulation output of linear colony growth with 10 genetic variants.** The simulation was initialized with Circular initialization ($\mathcal{L}_{\text{init}}$) with number of genetic variants = 10. The growth process was done for 10^4 steps/reproduction events. Grid size is 100×200 . Different shades of teal indicate cells of different genetic identity. Subfigure (b): **Simulation output of circular colony growth with 10 genetic variants.** The simulation was initialized with Circular initialization ($\mathcal{C}_{\text{init}}$) with initial radius $r_0 = 50$ and number of genetic variants = 10. The growth process was done for 10^5 steps/reproduction events. Grid size is 500×500 . The darkened initial core is ignored as it is inactive throughout the process and its constitution plays no role in the simulation. Different shades of teal indicate cells of different genetic identity.

growth, the size of the frontier (the height of the grid) remains fairly constant through the process. Thus, the unoccupied sites neighbouring the active cells remain bounded. Given enough time, two randomly propagating domain boundary will eventually collide, and the sector bound by them will close off. Hence, we expect if the linear growth simulation is propagated for long enough, there will be a fixation event i.e. only one sector will remain at the frontier.

Comparing this to circular growth (see figure 4.2(b)), we observe initially large number of sectors being closed off. But as the frontier (circumference of the radially growing colony) grows the spatial separation of domain boundaries increases on average leading to the events of sectors closing off becoming rarer as time progresses. In contrast with the linear growth front, the domain boundaries on a circular front seem to drift apart at a rate proportional to the velocity of the growth front. We can expect a non-zero number of sectors surviving on an average, in case of circular growth, and the sectoring pattern seems to stabilize.

Both of these observations agree with the experimental results and the theoretical expectations from the results described in 2.1.

When figures 4.2(b) and 4.2(a) are compared with 4.1(b) and 4.1(a), the obvious expectation of greater diversity at initialization leads to greater diversity at the end of growth is realized. But in the case, three or more genetic variants in the system, not all boundary collisions lead to boundary annihilation. Say, for example, a domain of type 1 neighbours a domain of type 2 on top and a domain of type 3 on the bottom, then if the two domain boundaries of the domain of type 1 were to collide, they will coalesce to form a single domain boundary between domains of type 2 and 3. Thus, the average rate at which domains grow (or contract) really is a function of the number of distinct genetic identities in the system, in addition to the number of domains and the span of the habitat. The mean field approximation for this calculation would rely on the expected fraction of boundary collisions which result in coalescence (as opposed to annihilation).

4.1.2 Boundarywantering as a super-diffusive random walk

With focus on the behaviour of singular domain boundary, 25 simulations of $\mathcal{LS}_{\text{init}}$ initialized EM grids was performed. A representative realization of such simulation is provided in figure 4.3(a). The collated 25 domain boundaries were collected and the MSD vs r (distance from the initial front) was performed as described in section 3.3.1. The $\ln(\text{MSD}(r))$ vs $\ln(r)$ plot is

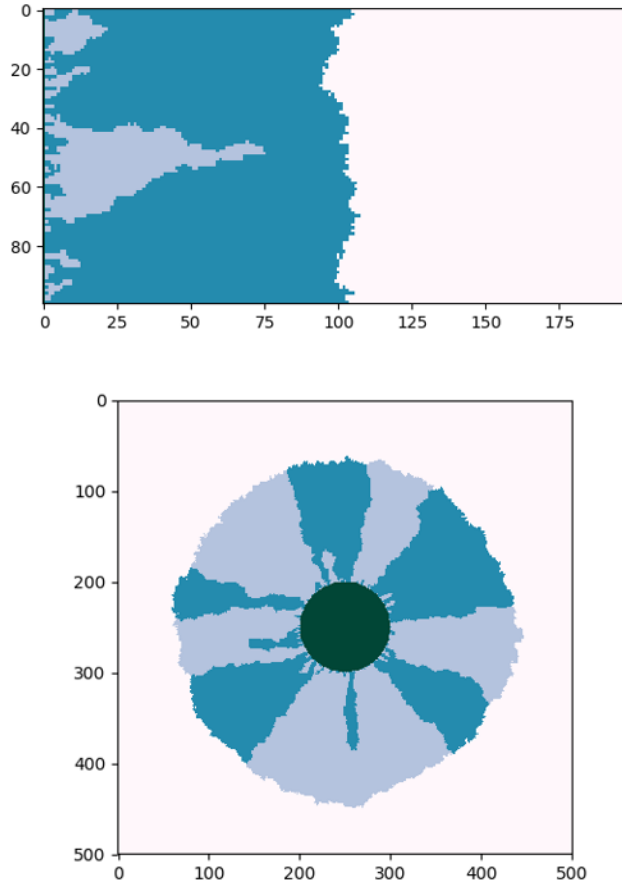


Figure 4.2: Subfigure (a): **Simulation output of linear colony growth with 2 genetic variants.** This simulation was on a rectangular grid with periodic boundary conditions on the top and bottom edge of the grid. The first column is row of inactive cells, the second column was initialized randomly with cells with one of 2 genetic identities. Different shades of teal indicate different genetic identity. The growth process was done for 10^4 steps/reproduction events. Grid size is 100×200 . Subfigure (b): **Simulation output of circular colony growth with 2 genetic variants.** The simulation was initialized with the circular C initialization with initial radius $r_0 = 50$ and number of genetic variants = 2. The growth process was done for 10^5 steps/reproduction events. Grid size is 500×500 . The darkened initial core is ignored as it is inactive throughout the process and its constitution plays no role in the simulation. Different shades of teal indicate cells of different genetic identity.

given in the figure

The close linear fit in figure 4.3(b) implies that domain boundary wandering can be approximated by the equation:

$$\text{MSD} = 2Dr^\alpha = 2Dr^{2\zeta} \quad (4.1)$$

The slope of the linear fit $\alpha = 2\zeta = 1.33$ indicates that if we were to consider r or in the linear case x as the time-coordinate then the domain boundary wandering is analogous to a **super-diffusive random walk**. The diffusive coefficient in that case would be $2D = 0.51$. The diffusive exponent has previously been calculated by simulations of the Eden model (Saito and Müller-Krumbhaar, 1995) which validates our implementation as consistent with it. This anomalous diffusive character can be attributed to the roughness of the growth frontier itself as described in section 2.1.

4.1.3 MSD over segmented windows

Furthermore, MSD analysis of segmented windows as described in section 3.3.2 was performed on singular domain boundaries obtained from both $\mathcal{LS}_{\text{init}}$ initialization and $\mathcal{SS}_{\text{init}}$ initialization for 25 runs on both. For the linear growth case, the segment windows ranged from 36 to 720 with increments of 36 while each total boundary lengths were ≈ 1000 . Similarly, for circular growth case, the segment windows ranged from 20 to 400 with increments of 20 while each total boundary lengths were ≈ 600 .

From the $\log_{10}(\text{MSD})$ vs $\log_{10}(\text{window size})$ plots for both $\mathcal{LS}_{\text{init}}$ and $\mathcal{SS}_{\text{init}}$ initializations, the line

$$y = \log_{10}(\text{MSD}) = \alpha \times \log_{10}(\text{window size}) + \log_{10}(2D) = mx + c \quad (4.2)$$

is fit using linear regression. The segmented window MSD plots are shown in figures 4.4(a), 4.4(b) and 4.5. Both fits give use similar values for the anomalous diffusion equation parameters. The $\mathcal{LS}_{\text{init}}$ simulations predict $\log_{10}(2D) = -1.8416 \pm 0.1663$ and $\alpha = 1.2903 \pm 0.0546$. The $\mathcal{SS}_{\text{init}}$ simulations predict $\log_{10}(2D) = -1.8913 \pm 0.1365$ and $\alpha = 1.3006 \pm 0.0546$. Both the results can be considered practically identical to each other. This is because only the local geometry of the growth frontier where the domain boundary intersects with it, affects the wandering of the domain boundary.

Remarkably, a previous study has found the diffusive exponent for the case of an *E. coli* colony to be $\zeta_{E. coli} = 0.65$ (Hallatschek *et al.*, 2007) which is very close to our results from the simulations of the Eden growth process $\zeta = 0.65$. This surprising similarity in the diffusive exponents have been noted by (Hallatschek *et al.*, 2007), although no speculations have been made

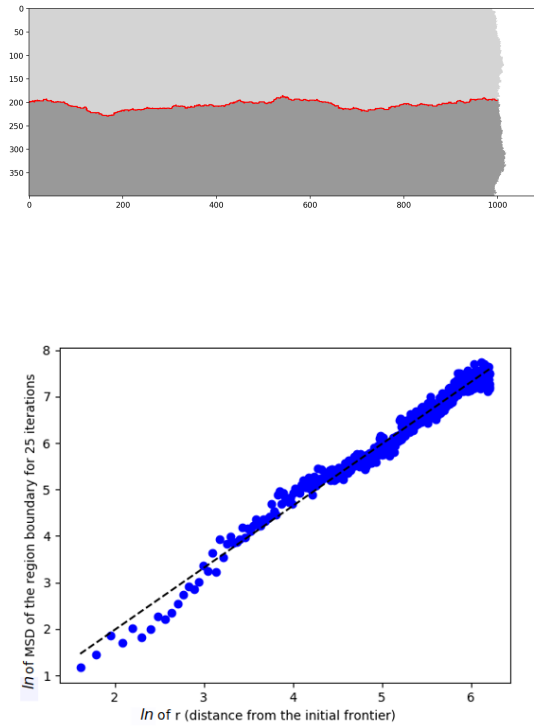


Figure 4.3: Subfigure (a): **Representative simulation output for singular domain boundary with a linear growth front** This simulation was done a rectangular grid with $\mathcal{LS}_{\text{init}}$. The top and bottom half of the first column of the grid was initialized with cells of genetic type 1 and 2 respectively. The boundary between the two domains is highlighted in red. The growth process was done for 4×10^5 steps/reproduction events. Grid size is 400×1200 . Subfigure (b): **\log_e of MSD of domain boundary vs. \log_e of distance from initial frontier** The y-value is the natural logarithm of the squares of vertical deviation of the domain boundary in (a) averaged over 25 instances of simulation and the x-axis represents the natural logarithm of the distance propagated by the frontier from the initial frontier. The blue dots indicate the simulated MSD values and the dashed black line is the linear fit (slope = 1.33, y-intercept = -0.680)

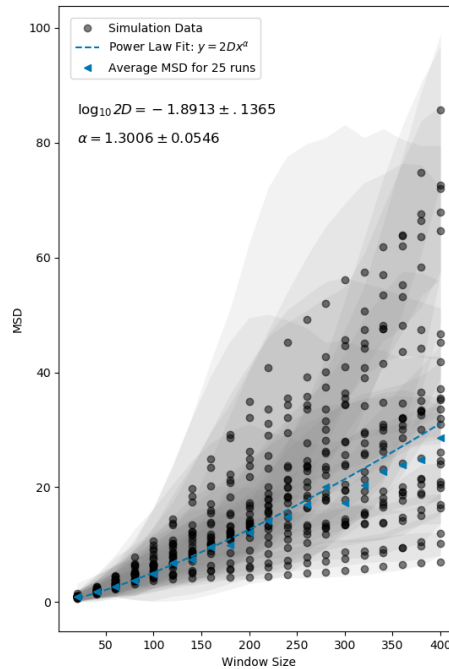
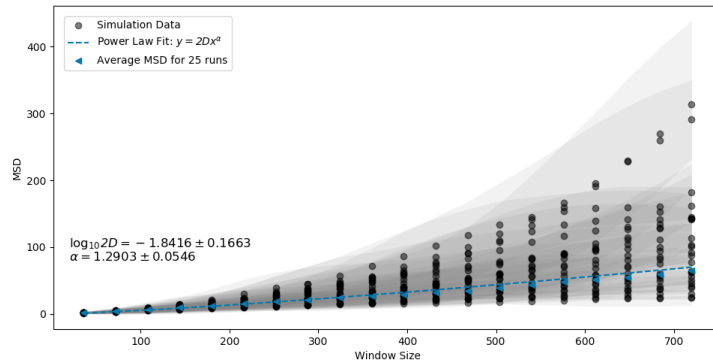


Figure 4.4: Subfigure (a): **Segmented window MSD vs window size for the linear growth case (linear scale)** The plot shows segmented window MSD calculation performed on 25 runs of singular domain boundary propagation with $\mathcal{L}\mathcal{S}_{\text{init}}$ initialization. The segment window sizes ranged from 36 to 720 with increments of 36 while each total boundary lengths were ≈ 1000 . Subfigure (b): **Segmented window MSD vs window size for the circular growth case (linear scale)** The plot shows segmented window MSD calculation performed on 25 runs of singular domain boundary propagation with $\mathcal{S}\mathcal{S}_{\text{init}}$ initialization. The segment window sizes ranged from 20 to 400 with increments of 20 while each total boundary lengths were ≈ 600 . In both the subfigures, black dots indicates the average MSD for the given window size for a single domain boundary realization and the semi-transparent grey filled curves represent standard deviation bounds for each black dot. The blue triangle represent the MSD averaged for all 25 boundaries. The dashed blue line indicates the power law fitted curve for $y = 2Dx^\alpha$.

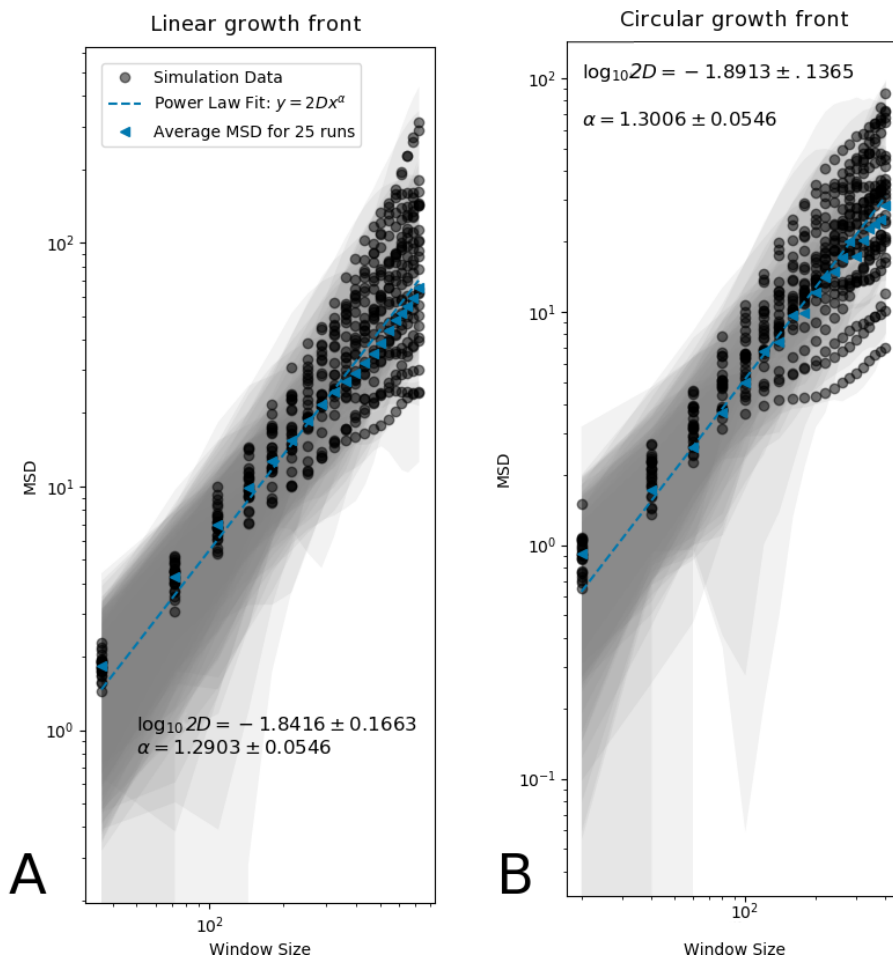


Figure 4.5: **Segmented window MSD vs window size for the linear and circular growth case (loglog scale)** Subfigure (a) is the plot in figure 4.4(a) in a log-log scale. Subfigure (b) is the plot in figure 4.4(b) in a log-log scale. Refer to figures 4.4(a) and 4.4(b) captions for details for the plots

as to why. This connection is further made stark by noting that *E. coli* is but one microorganism which forms such two-dimensional colonies on substrates, while it has been noted that yeast colonies do not show such pronounced boundary wandering.

4.2 Effect of growth delay on demixing and sector boundaries

In this section, the results regarding the Eden Model with growth delay are presented. The model formulation and the implementation are discussed in sections 2.3 and 3.2, respectively. Before proceeding, it should be noted that the parameters in the EM with growth delay model, namely τ and δ . As noted in 2.3, the growth process only depends on the ratio $\frac{\tau}{\delta}$ upto a time rescaling, $\tau = 1$ is fixed for all the simulations.

The obvious consequence of increasing δ is that the overall colony growth will be slowed down¹, as the time from birth to the first division event for each cell is increased by δ . Although, we cannot say that the linear relation at the individual level scales to population level. As seen in the figure 4.6, for the same simulation time T , the final colony size achieved decreases for increasing δ .

There are other details which stand out in the figure 4.6. Qualitatively, the domain boundary wandering becomes less prominent for increasing δ . Moreover, the significant presence of the cells in the nascent state at the growth frontier for higher values of δ . To make an analogy to reaction systems, the nascent to active transition becomes the rate limiting process for larger δ values for this multi state growth process. Because of this, the colony geometry is also affected. In EM starting with a single cell or a circular initial colony the final colony shape approaches a circle. While in EMwD, although the growth front becomes smoother, the overall curvature starts deviating from circularity for increasing δ . It is an artefact of the underlying square lattice. I speculate on this further in the next chapter.

4.2.1 Effect of growth delay on boundary wandering

The qualitative observations suggest that δ is smoothening the growth frontier and thus, dampening the boundary wandering phenomenon which causes the sectoring. Segmented window MSD analysis was performed in the same

¹This is the reason I referred to this model as Eden Model with growth delay.

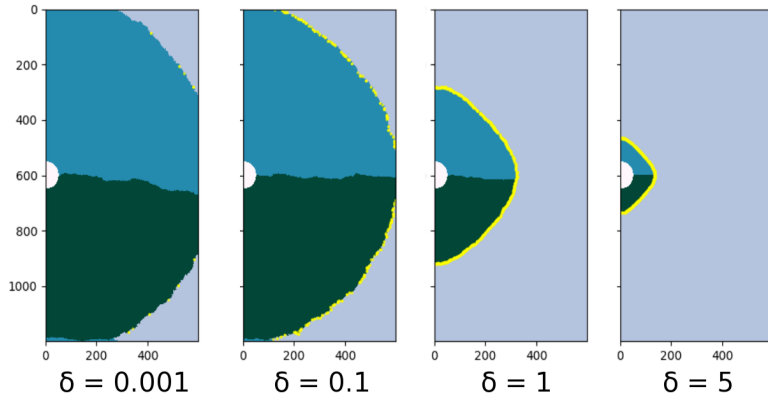


Figure 4.6: **Simulation output with $\mathcal{SS}_{\text{init}}$ for various δ values after growth time $T = 500$ and mean division time $\tau = 1$ fixed for each.** The four simulation outputs for $\delta = 0.01, 0.1, 1, 5$ are shown. The two shades of teal represent cells of different genetic identities. The yellow dots represent nascent cells.

fashion as in section 4.1.3 for a range of δ for both the LS and SS initializations.

For both $\mathcal{LS}_{\text{init}}$ and $\mathcal{SS}_{\text{init}}$, we can see from figures 4.7(a) and 4.7(b) that MSD profiles for $\delta \leq 0.1$ remains largely stable and similar to those of respective growth case in EM. For $\delta > 0.1$ we see the MSD values starting dipping, indicating quantitatively lesser boundary wandering.

The diffusive equation parameters D , the diffusion coefficient, and α , the diffusion exponent, derived from fitted curves to the segmented window MSD profile are plotted as a function of δ in figure 4.8.

The plots in figure 4.8 are surprising, as the initial expectation would be to presume that both the parameter D and α would have a monotonic dependence of δ , if any. In figure 4.8(a), at $\delta = 10^{-3}$, the predicted $\alpha = 1.13$ is significantly lower than the one obtained for the EM = 1.30. Then, it further increases to 1.31 at $\delta = 0.01$. This is in stark contrast with our intuition based in theory described in section 2.1 that increasing δ leads to smoothening of growth frontier which in turn, should reduce α . In the circular case, initially for $\delta \leq 0.1$, α fluctuates about 1.0 but then shoots up to $\alpha = 1.35$ for $\delta = 1$. Moreover, the δ dependence profile of both the parameters predicted for different growth cases are starkly different.

To probe this further, simulations were performed δ varying between 10^{-4} and 10^{-2} . The results from the subsequent segmented window MSD analysis and diffusion equation curve fitting are shown in figure 4.9. Even for relatively

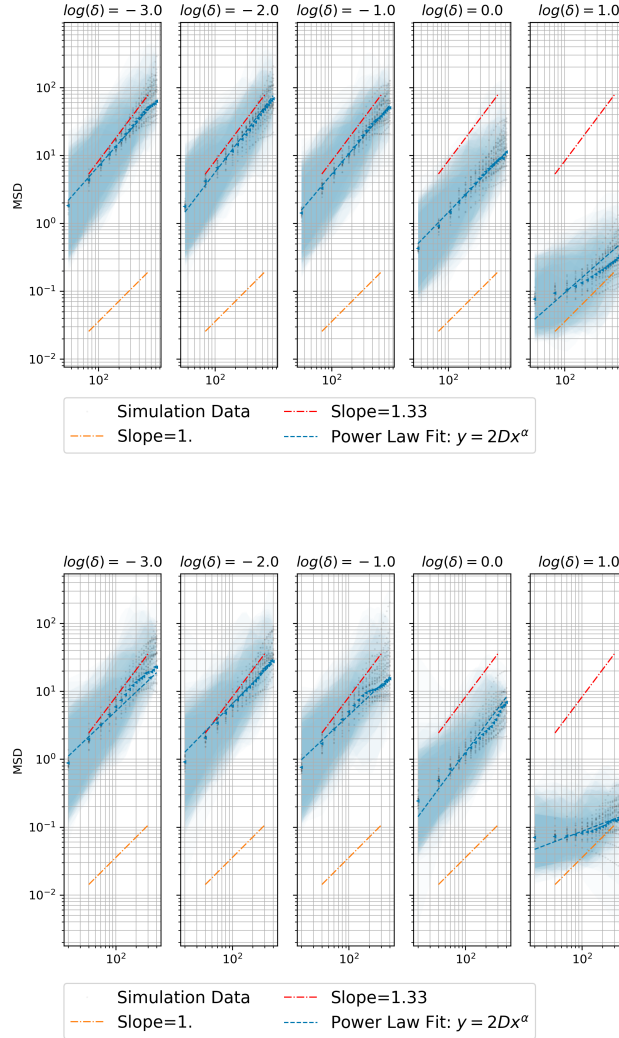


Figure 4.7: Subfigure (a): **Segmented window MSD plots for $\mathcal{L}\mathcal{S}_{\text{init}}$ with growth delay for varying δ from 10^{-3} to 10.** Each plot utilizes single domain boundary data from 25 different simulation runs. The segment window sizes ranged from 36 to 720 with increments of 36 while each total boundary lengths were ≈ 1000 . Subfigure (b): **Segmented window MSD plots for $\mathcal{S}\mathcal{S}_{\text{init}}$ initializations with growth delay for varying δ from 10^{-3} to 10.** Each plot utilizes single domain boundary data from 25 different simulation runs. The segment window sizes ranged from 20 to 400 with increments of 20 while each total boundary lengths were ≈ 600 . The axis are plotted in log-log scale. The dashed blue line indicates the fitted $y = 2Dx^\alpha$ curve. Black dots are data points from individual runs and blue triangles are indicate the mean MSD over 25 runs. The light blue regions indicate the spread of the data. Representative lines of fixed slope 1.33 and slope 1 are drawn in red and orange respectively.⁴²

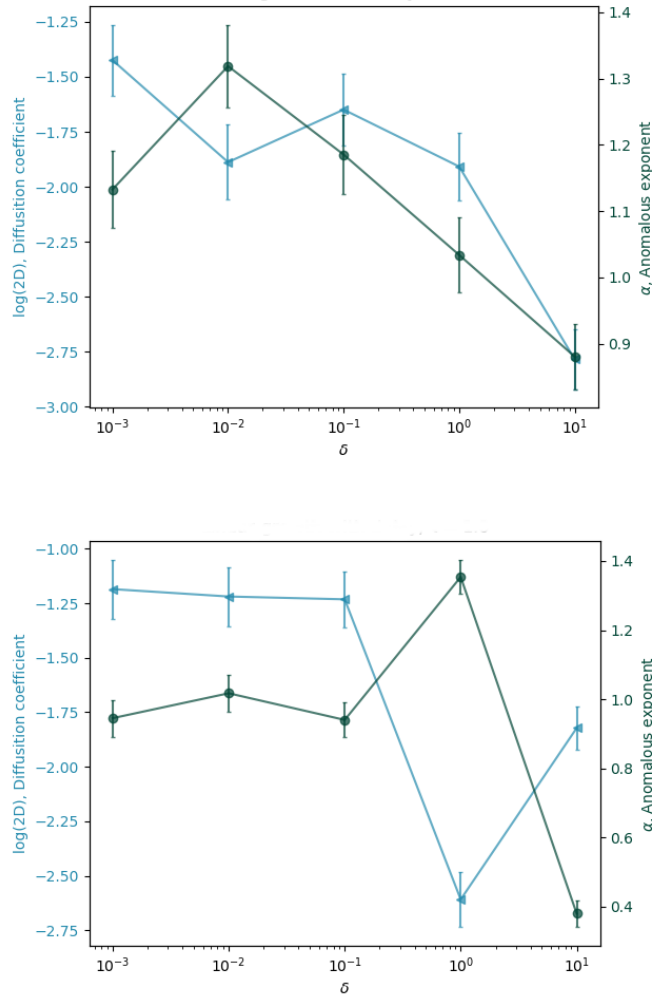


Figure 4.8: Subfigure (a): **Fit values of $\log(2D)$ and α as function of δ as obtained from data from $\mathcal{LS}_{\text{init}}$.** Both variables are plotted over a common x-axis. Subfigure (b): **Fit values of $\log(2D)$ and α as function of δ as obtained from data from $\mathcal{SS}_{\text{init}}$.** Both variables are plotted over a common x-axis. The blue curve and axis (left) corresponds to $\log(2D)$ where $2D$ is the diffusion coefficient. The green curve and axis (right) corresponds to α , the diffusion exponent. See figure 4.7(b) for the fitted curves.

low δ values the diffusion parameters fluctuate dramatically upon varying δ . α value, in fact, goes above 1.4 at $\log(\delta) = -2.8$ which shows a higher degree of superdiffusive character than the original EM. Another peculiar observation to be made is that the fluctuation pattern for both $\log(2D)$ and α between $10^{-4} \leq \delta \leq 10^{-3}$ and $10^{-3} \leq \delta \leq 10^{-2}$ are remarkably similar except the fluctuations are much more pronounced in the latter segment.

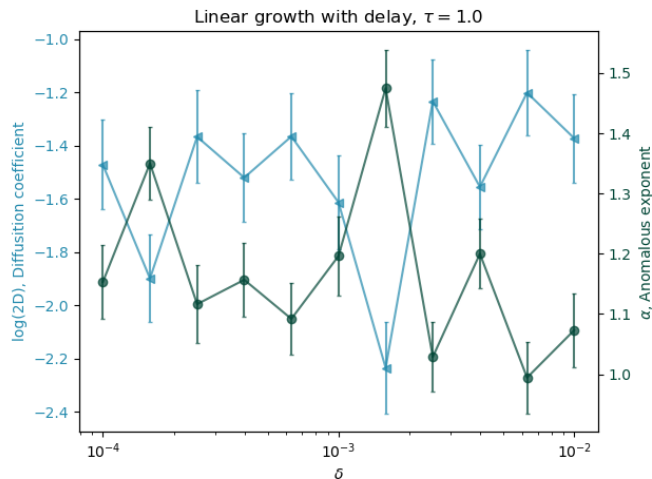


Figure 4.9: **Fit values of $\log(2D)$ and α as function of δ as obtained from data from $\mathcal{LS}_{\text{init}}$ for finer δ sampling.** Both variables are plotted over a common x-axis. The blue curve and axis (left) corresponds to $\log(2D)$ where $2D$ is the diffusion coefficient. The green curve and axis (right) corresponds to α , the diffusion exponent.

Also, for almost all consecutive δ values, a decrease in α is accompanied by an increase in $\log(2D)$ and vice versa. Since, both parameters drive the boundary wandering, they seem to fluctuate in a compensatory manner. To look at the combined effect of these parameters on the overall boundary wandering, the fitted parameters were used to compute the predicted MSD of the domain boundary for fixed propagation length r , and plotted over varying δ .

Figures 4.10(a) and 4.10(a) both show that predicted MSD vs δ curves seem to be stable for $\delta < 0.1$ around the values which are fairly close to the predicted MSD values from the EM, $\text{MSD}_{\text{fit}}(r) = 2.0, 0.7, -0.6$ for $r = 1000, 100, 10$ respectively, followed by a decrease in predicted MSD for all r when δ increases beyond 0.1. In figure 4.11, for lower δ below 0.01, the predicted MSD values remain fairly constant and consistent with the pre-

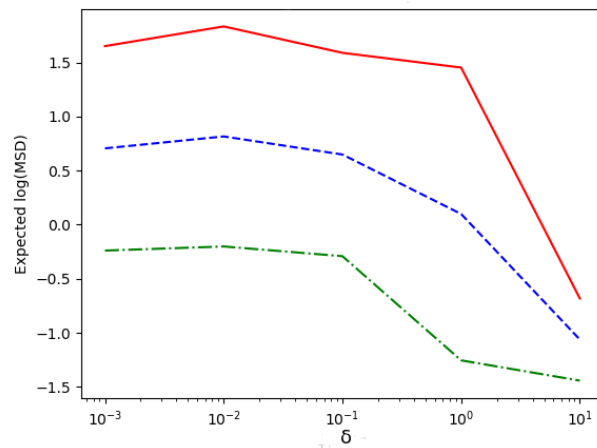
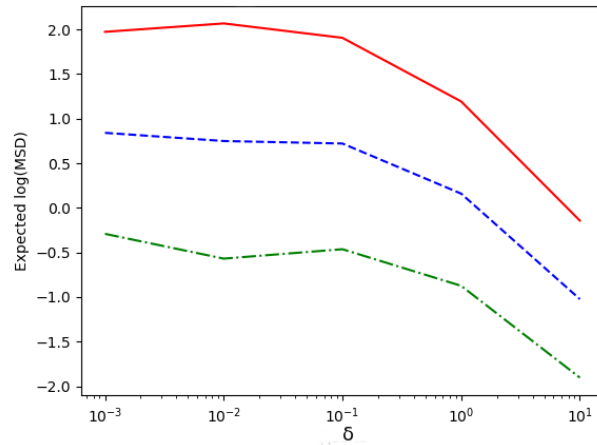


Figure 4.10: Subfigure (a): **Expected log MSD(r) = $\log(2D)_{\text{fit}} + \alpha_{\text{fit}} \cdot \log(r)$ vs δ for $\mathcal{LS}_{\text{init}}$ data** Subfigure (b): **Expected log MSD(r) = $\log(2D)_{\text{fit}} + \alpha_{\text{fit}} \cdot \log(r)$ vs δ for $\mathcal{SS}_{\text{init}}$ data** Red, blue and green lines denote $r = 1000, 100$ and 10 respectively.

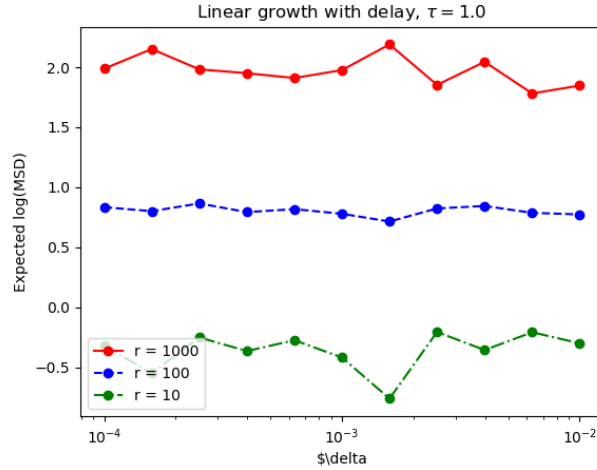


Figure 4.11: **Expected** $\log \text{MSD}(r) = \log(2D)_{\text{fit}} + \alpha_{\text{fit}} \cdot \log(r)$ vs δ for $\mathcal{LS}_{\text{init}}$ data Red, blue and green lines denote $r = 1000, 100$ and 10 respectively.

dicted MSD for the EM. Thus, the variation in D and α as seen in figure 4.9 seem to be almost definitely compensatory to give a constant boundary wandering.

Taken altogether, the results from the simulations of EMwD show that the model can show a wide range of behaviour with respect to the domain boundary wandering phenomenon going from superdiffusive to diffusive and even subdiffusive character under certain values of δ . Moreover, we can identify $0.1 < \frac{\delta}{\tau} < 10$ as the critical range in which the domain boundary behavior goes from fairly pronounced wandering like the EM to an almost ballistic propagation.

Chapter 5

Discussion

The microbial colony growth in two-dimensions is subject to strong genetic drift resulting in spatial demixing of genotypes in the growing population and establishment of monoallelic domains separated by a sharp domain boundary. As part of this study, it is demonstrated that a simple agent-based model on a lattice such as the Eden Model is sufficient to capture the salient behaviour of genetic demixing and sectoring and even match the boundary wandering exponent of the experimentally observed *E. coli* colonies. Nevertheless, this is not enough to emulate the range of genetic sectoring patterns seen in nature, let alone give us insight as to what factors drive and control this behaviour. In this project, I outline a bottom-up approach to model this phenomenon, in which we start with the existing Eden Model and incorporate a biologically suggestive detail while maintaining the simplicity of the original.

The formulation of EMwD expands the range of behavior as far as the genetic sectoring is concerned, while the EM remains as a special subcase of the new model. The incorporation of a nascent cell state and the maturation time (δ) in the growth process, provides us with a tunable parameter to control the behavior of the system. The results from analysing the singular domain boundary for various cases, suggests that increasing δ from 0 to the values comparable to the mean division time dampens the domain boundary wandering, leading to a more stable sectoring pattern and reduced genetic drift. This makes good on the promise of giving us a possible avenue through which we can go from the sectoring pattern of, say, *E. coli* colony to that of yeast colonies. In fact, experiments of observing the sectoring pattern of *E. coli* colony at 21° C and 37° C, show that the domain boundary wandering in colonies grown at lower temperature show less genetic drift (Gralka *et al.*, 2016). This further corroborates with our model in which slower cell growth leads to less genetic drift.

In our EMwD model, the nascent cell state is comparable to a cell which

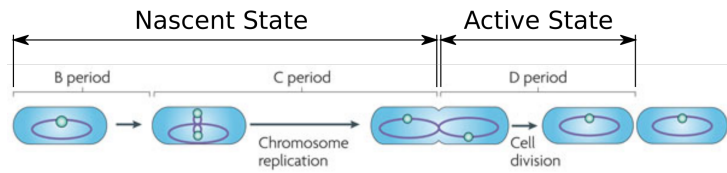


Figure 5.1: Figure adapted from Wang and Levin 2009. The illustration depicts various stages of a traditional bacterial cell cycle. Here the B period is the period between cell birth and the initiation of chromosomal replication. The C period depicts the stage in which chromosomal replication takes place and finally, the D period is the time between completion of chromosomal replication and the completion of cell division. We indicate that B/C period corresponds to the nascent state in our EMwD model whereas the D period corresponds to the active state in the same.

is doubling its size and undergoing chromosomal replication. The active cell state, then, would correspond to a cell which has undergone chromosomal doubling and can undergo cell division. Specifically in the case of bacterial cell cycle, the nascent cell state and the active cell state can be thought of analogous to B/C periods and D period of the bacterial cell cycle respectively.

This mapping is not accurate as the cell division in EMwD is asymmetric in the sense the two daughter cells do not start at the same cell state after division. It would be interesting to simulate EMwD in which an active cell reverts to the nascent state after reproducing a daughter cell at a neighbouring site. In that formulation, perhaps we can map fast-growing bacteria and slow-growing bacteria on a $\frac{\delta}{\tau}$ axis and see if the sectoring patterns compare to what the model would predict.

In the case of a eukaryotic microbe, such as *S. cerevisiae*, the mapping of the nascent and active states described in EMwD will correspond to the interphase and M phase in the eukaryotic cell cycle respectively. Although, this mapping is not comprehensive at the moment and a more elaborate modelling of the cell cycle stages in further development of EMwD would be required to have a more faithful modelling of various cell cycle stages, their regulation and how the time period with each are distributed.

Comparison of the effect of cell growth cycle between *E. coli* and *S. cerevisiae* colony genetic sectoring pattern cannot be made conclusively as there are other large scale differences in how the cells spatially organize in colonies between the two organisms. The cell size and shape difference between a rod-shaped bacteria and a larger ellipsoidal yeast cell would certainly lead to difference in the spatial organization of the cells which our grid-based EMwD

model is not fit to address. Nonetheless, if quantitative description of spatiogenetic pattern and the domain boundaries in *S. cerevisiae* colonies are provided, our expectation is that the diffusive parameters of the boundary wandering measured from experiments would correlate with the coefficient of variation of the time period from birth to the subsequent division of the cell. In our model, this coefficient of variation would have the value $\frac{\tau}{\tau+\delta}$. Expectation from our model is that boundary wandering decreases with increase in the C_V of the birth to division time. Experiments can be performed by measuring C_V for *E. coli* and *S. cerevisiae* and performing fluorescence imaging experiments as described in the introduction. The outcomes from these experiments can support or oppose the predictions from our model but cannot wholly rule out the effect of any other factor independently influencing the spatiogenetic pattern.

The within species comparative experiments where only cell growth cycle stages are regulated and the resultant spatiogenetic pattern are compared as outcomes would be more promising test of the EMwD model. For such an experiment for a particular microbial species or strain, any retardation of chromosomal replication or, separately, the division of the cell by the means of addition of metabolic agents which are known to inhibit one of these processes can be used to create two or more types of populations which show differing ratio of time period of B/C phase and D phase, which in EMwD corresponds to $\frac{\delta}{\tau}$. Moreover, temperature can be used to control the cell cycle kinetics. These time periods can be measured by flow cytometry. After having measured those, the two population with differing cell cycle behavior can be inoculated separately and their spatiogenetic patterns can be analyzed. From the model, I hypothesize that higher $\frac{\delta}{\tau}$ ratio will lead to lesser domain boundary wandering and higher number of surviving sectors for the same amount of colony propagation.

The results from the EM with growth delay, also poses interesting questions as to how the incorporation of nascent cell state and why the diffusive boundary parameters behave in such an unintuitive manner for varying δ . Analysing the growth frontier, particularly in terms of its roughness and roundness, in case of circular inoculation, might give insight as to the interaction of growth frontier and domain boundaries.

The deviation from the circularity for colonies is an artefact of the lattice, for large $\frac{\delta}{\tau}$ ratio, which is at odds with how microbial colonies spread. I speculate that the cells in the model seem to go from asynchronous stochastic division events to synchronous divisions, for large δ , and the colony cluster tend towards the diamond shape similar to that of the Von Neumann neighbourhood. This can be alleviated trying out the model implementation over a hexagonal lattice or a square lattice with Moore neighbourhood.

So far, we have only characterized the domain boundary wandering. Statistical characterization the genetic sectoring pattern at large is required in terms genetic diversity measures such as number of surviving sectors, local spatiogenetic correlations.

References

- Eden M (1961). A two-dimensional growth process. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Contributions to Biology and Problems of Medicine, 223–239. University of California Press, Berkeley, Calif.
- Farrell FD, Gralka M, Hallatschek O, Waclaw B (2017). Mechanical interactions in bacterial colonies and the surfing probability of beneficial mutations. *Journal of The Royal Society Interface* 14(131), 20170073.
- Gralka M, Stiewe F, Farrell F, Möbius W, Waclaw B, Hallatschek O (2016). Allele surfing promotes microbial adaptation from standing variation. *Ecology Letters* 19(8), 889–898.
- Hallatschek O, Hersen P, Ramanathan S, Nelson DR (2007). Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences* 104(50), 19926–19930.
- Hallatschek O, Nelson DR (2008). Gene surfing in expanding populations. *Theoretical Population Biology* 73(1), 158 – 170.
- Jullien R, Botet R (1985). Scaling properties of the surface of the eden model in $d=2, 3, 4$. *Journal of Physics A: Mathematical and General* 18(12), 2279–2287.
- Korolev K, Xavaier J, Nelson D, Foster K (2011). A quantitative test of population genetics using spatio-genetic patterns in bacterial colonies. *The American Naturalist* 178(4), 538.
- Kuhr JT, Leisner M, Frey E (2011). Range expansion with mutation and selection: dynamical phase transition in a two-species Eden model. *New Journal of Physics* 13(11), 113013.
- Matsuyama T, Matsushita M (1993). Fractal morphogenesis by a bacterial cell population. *Critical Reviews in Microbiology* 19(2), 117–135.

McFarlane A (2016). Periodic boundary conditions for lattices in python.

Rudge TJ, Federici F, Steiner PJ, Kan A, Haseloff J (2013). Cell polarity-driven instability generates self-organized, fractal patterning of cell layers. *ACS Synthetic Biology* 2(12), 705–714. PMID: 23688051.

Saito Y, Müller-Krumbhaar H (1995). Critical phenomena in morphology transitions of growth models with competition. *Phys Rev Lett* 74, 4325–4328.

Appendix A

Source Code

Source code of the work used in the results described in this text are stored at <https://github.com/CyCelsLab/microbialDemix>.