



Nnaji, Chioma Frances Agatha (2015) *Modeling bacterial dynamics in chemostats*. PhD thesis.

<https://theses.gla.ac.uk/7164/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk

Modeling Bacterial Dynamics in Chemostats.

Chioma Frances Agatha Nnaji

Submitted in fulfillment of the requirements for the Degree of
Doctor of Philosophy in Civil Engineering

University of Glasgow
Department of Civil Engineering

August 2015

To my dearest Parents...

ABSTRACT

Clean water is a vital resource, which climate change and population growth conspire to make increasingly scarce. Thus it is imperative that we maintain the quality of our watercourses and recycle polluted waters. Biological treatment of wastewater is at the forefront of current strategies employed to treat domestic wastewater. The transformation of wastewater into less harmful products is performed by complex naturally forming microbial communities. Waste treatment processes are essentially a product of the ecology of these communities and yet we have a poor understanding of some of the most basic ecological processes. In particular, the coming together, or assembly, of the community. Thus when failure occurs it can be baffling. In the last decade, there has been growing evidence that stochastic processes are an important component of microbial community assembly. In open biological communities, if species are functionally equivalent and it is only stochasticity that shapes the community then the dynamics are said to be neutral. It is then important to ascertain the relative importance of stochasticity in engineered systems and on that basis formulate theories to guide the successful design and operation of wastewater treatment plants.

Neutral community assembly theory has been in the limelight for more than a decade because of its ability to predict species area relationship and species abundance distribution. Its underlying assumption of equivalent specific growth rates is controversial and has led to much debate, which is muddied by sampling issues, parameter estimation techniques and lack of data. In this

thesis, I attempt to address these confounding factors and properly parameterize the stochastic model originally postulated by Hubbell (Hubbell, 2001) and subsequently modified by Sloan (Sloan *et al.*, 2006) to suit microbes. This is done by conducting detailed experiments in parallel chemostats to give time series of the abundance of organisms that have been engineered to have highly tunable kinetics. Calibration using time series data affords the first opportunity to validate neutral (and near neutral) dynamics. Previous studies only use stationary abundance distributions that could have emerged from a variety of alternative mechanisms.

The organisms with tunable kinetics were engineered by a genetic recombination technique, which generated two strains of *E.coli* with different antibiotic resistance genes placed at the same location in otherwise identical genomes. When no antibiotics are present these strains have identical growth characteristics and hence are neutral with respect to one another. In the presence of low concentrations of antibiotics the strain with the appropriate resistance gene has an advantage.

Time series abundance of the two strains were obtained under three different experimental setups that were devised to give differing weight to neutral and selective processes. A model that incorporates selective and neutral effects could be calibrated in all cases, but the match between experimental and theoretical parameters could only be achieved if an 'effective community size' that is smaller than the real community size is used. Evidence is given for this being a phenomenon associated with spatial correlation in the demographics

of the community. The consequences of this are profound. It means that even in very large microbial populations random drift will affect community composition and identically engineered systems will yield differing population dynamics.

Table of Contents

1	Introduction	1
2	Literature Review	6
2.1	Wastewater Treatment	6
2.2	Ecology	7
3	Conceptual Design of Experiments	19
	Chemostat	23
4	Genetically Identical Organisms	28
4.1	Genetic Recombination	28
4.1.1	Recombination-mediated Genetic Engineering	29
4.1.2	Result of Recombination-mediated Genetic Engineering	40
4.2	Confirming the deletion of <i>pepA</i> Gene	45
4.3	Copy Number	48
4.4	Real Time-Polymerase Chain Reaction	54
4.5	Pseudo Quantification Experiment	59
5	Characterization of Strains	62
5.1	Generation of process parameters in the absence of antibiotics	62
5.1.1	Batch Growth Rate	63
5.1.2	Monod's Kinetics	67
5.1.3	Growth Yield Coefficient	71
5.2	Generation of process parameters in the presence of antibiotics	77
6	Chemostat Experiments	85
6.1	Neutral Dynamics with Immigration	85
6.1.1	Deterministic Model for Neutral Dynamics with Immigration Experiment	86
6.1.2	Experimental Setup for Neutral Dynamics with Immigration Experiment	95
6.1.3	Experimental Result for Neutral Dynamics with Immigration	102
6.2	Non-neutral Dynamics in a Completely Insular Community	106
6.2.1	Deterministic Model for Non-neutral Dynamics in a Completely Insular Community	107
6.2.2	Experimental Setup for Non-neutral Dynamics in a Completely Insular Community	109
6.2.3	Experimental Result for Non-neutral dynamics in a completely insular community	111
6.3	Non-neutral dynamics with immigration	113
6.3.1	Deterministic Model for Non-neutral dynamics with immigration	115
6.3.2	Experimental Setup for Non-neutral dynamics with immigration	118
6.3.3	Experimental Result for Non-neutral dynamics with immigration	122
6.4	Discussion	125
7	Stochastic Modeling	127
7.1	Stochastic Differential Equations (SDE)	127
7.2	Model Calibration	131
7.3	Simulating generic behaviour of the system and validating the calibration method	132
7.3.1	Neutral model with Migration	134
7.3.2	Non-Neutral model with immigration	140
7.3.3	Completely insular neutral community ($m=0$)	142
7.3.4	Testing the Calibration Technique: Retrieving Process Parameters	144
7.4	Calibration of Chemostat Experiments	146

7.4.1	Neutral dynamics with immigration.....	146
7.4.2	Non-neutral dynamics in a completely insular community	149
7.4.3	Non-neutral dynamics with immigration	151
8	Discussion and Conclusion.....	155
9	Further Research	166
10	Bibliography.....	168

LIST OF FIGURES

- Figure 2-1: Schematic representation of the unified theory of biodiversity and biogeography. The local community represents a subset of the species found in the source community. Here they are made up of 16 individuals from 6 species. (a) Random deaths occur in the local community creating space. (b) The spaces are filled randomly either by local birth or (c) random dispersal from the source community with a probability m . Random speciation can also occur leading to the formation of new species. These processes occur in each time-step bringing about changes in species abundance over time.....12
- Figure 3-1: Schematic representation of a chemostat showing all three chambers – the feed reservoir, growth chamber and effluent reservoir along with accessories such as pumps and connecting tubes.....24
- Figure 3-2: Conceptual design of the chemostat experiments used in this study.....27
- Figure 4-1: Illustration showing the target region for genetic recombination in the E.coli genome. I) The *pepA* gene is flanked to the left by *yjgP* gene and to the right by *holC* gene. II) The expected outcome of the recombination. III) Generation of linear dsDNA molecules carrying the antibiotic resistance gene, one for each organism via PCR.....33
- Figure 4-2: I) 1% Agarose electrophoresis gel showing final yield of PCR products. II) Schematic representation of PCR products showing the desired gene (antibiotic resistance) at the center flanked by the target genes.....35
- Figure 4-3: Purified PCR products without primer dimers.36
- Figure 4-4: Schematic representation of chemical transformation of cells via (CaCl_2).....38

Figure 4-5: A schematic representation showing the summary of the recombineering reaction used in this project. I) PCR setup with forward primer consisting of 50bp from yjgP and 22bp from Cm/Kan resistance gene plasmids in the 5'→3' and reverse primer consisting of 21bp from Cm/Kan resistance gene plasmids, 19bp from holC gene and 31bp from pepA gene in the 3'→ 5' direction. The sequences from Cm/Kan resistance gene plasmids on the primers are complementary to those before and after the Cm/Kan resistance plasmid orf. The PCR yields (II), which is transformed into competent cells containing pKOBEG-A that has the recombinase λ Red system. Red mediated recombination occurs yielding the recombinant product (III) where Cm/Kan resistance gene replaces pepA in the genome of DS941 cell.44

Figure 4-6: An illustration showing site-specific recombination at cer and psi sites and the genes that facilitates the recombination.....46

Figure 4-7: 1% agarose electrophoresis gel showing extracted plasmid DNA fragments from the strains and DS941 cells characterized according to sizes.....47

Figure 4-8: Diagram showing the primers for copy number verification, their target regions in the genome of the modified strains and expected products.....50

Figure 4-9: Result showing the verification of copy numbers using primer pair ABc.....51

Figure 4-10: Result showing the verification of copy numbers using primer pair AB_K.....52

Figure 4-11: Result showing the verification of copy numbers using primer pair AC.....53

Figure 4-12: An illustration of the mode of action of TaqMan probe chemistry during DNA amplification and detection. The diagram shows the position of each oligonucleotide sequence on the target DNA sequence prior to annealing along with reporter quencher dyes. It also shows the extension of the forward and reverse primers and subsequent displacement of the fluorophore resulting in its fluorescence. The intensity of the fluorescence indicates the amount of target DNA sequences detected and amplified.58

Figure 4-13: Result showing goodness of fit for both strains in terms of ratio added and ratio observed.....61

Figure 5-1: Graphs a, b and c show batch growth curves of the wild type and other strains on three different days and graph d, an average of the three curves plotted together with standard deviation as error bars.....65

Figure 5-2: Monod Kinetics result of Graph showing actual experimental data points and data points derived by substituting growth parameters from the regression analysis a) – the chloramphenicol resistant strain and b) – the kanamycin resistant strain in the absence of antibiotics.....69

Figure 5-3: Estimates of growth yield of the strains when glucose is the growth-limiting substrate: a) when calculated normally, b) when calculated from the initial weight of substrate.....74

Figure 5-4: Result of growth yield of a) the chloramphenicol resistant strain and b) the kanamycin resistant strain in the absence of antibiotics.75

Figure 5-5: Preliminary growth assessment showing OD measurements at a time point of both strains in different concentrations of a) chloramphenicol and b) kanamycin.81

Figure 5-6: Detailed growth curve of both strains in different concentrations of antibiotics predetermined from the preliminary test. A) Cm+ strains in different concentrations of Kanamycin and B) Kan+ strains in different concentrations of chloramphenicol.....82

Figure 5-7: Monod Kinetics result of Kan+ strain in 0.5µg/ml Chloramphenicol84

Figure 6-1: Schematic representation of mass balance equations around a bioreactor, where V = Volume of reactor, F = flow rate in and out of reactor as indicated by arrows, x = biomass and $[S]_0$ = original concentration of growth limiting substrate.87

Figure 6-2: Schematic representation of mass balance equations around the local community. The subscripts i and j represent the Cm+ and Kan+ strains. In this case $X_{0i} = X_{0j}$ and will subsequently be written as X_0 90

Figure 6-3: Simulation results for growth in both source communities using a dilution rate of 0.5hr^{-1} a) and b) and c) that for the local communities with a dilution rate of 0.33hr^{-1} when no advantage is applied.....92

Figure 6-4: Conceptual chemostat design for neutral dynamics experiment showing arrangement of bioreactors, pumps and connecting tubes.....95

Figure 6-5: Figure showing individual pieces of bioreactors used in the continuous reaction experiments. a) The bioreactors used for the pure cultures of the strains in the source communities, b) The manifold used to distribute migrants and c) The reactors that hold the mixed culture of both strains that represent the local communities.96

Figure 6-6: Picture showing assembled units of apparatus for the chemostat experiment.....98

Figure 6-7: Diagram showing growth of both strain in the source community with a dilution rate of 0.5hr^{-1} in a) absolute abundance, b) relative abundance and c) diagram showing variability between reactors in the local community from sampling times 22 – 34 hours. Dilution rate in the local community was 0.33hr^{-1} 104

Figure 6-8: Growth of both strains in the 10 local communities when no selective advantage exists. Growth was achieved using a dilution rate of 0.33hr^{-1} . The legend shown for reactors 1 and 2 applies to all reactors – the red data points depict the Cm+ strain and the blue data points the Kan+ strain. The missing data points depict nighttime when sampling could not be conducted..... 105

Figure 6-9: Schematic showing conceptual arrangement of chemostat for the Non-neutral dynamics in a completely insular community' experiment.....107

Figure 6-10: Simulation result of the local community when selective advantage is added to the system..... 108

Figure 6-11: Growth of both strains in the 10 local communities when only selective advantage exists in the system. Growth was achieved using a dilution rate of 0.17hr^{-1} . The legend shown for reactors 1 and 2 applies to all reactors – the red data points depict the Cm+ strain and the blue data points the Kan+ strain. The missing data points depict nighttime when sampling could not be conducted. 112

Figure 6-12: Diagram showing chemostat design for the 'Non-neutral dynamics with immigration' experiment..... 115

Figure 6-13: Simulation results for the growth of both strains as pure cultures in their respective source communities using a dilution rate of 0.5hr^{-1} (a and b) and c) in the local communities using a dilution rate of 0.42hr^{-1} when both selection and migration experiment. 116

Figure 6-14: Diagram showing growth of both strain as pure cultures in the source community in terms of a) absolute abundance and b) relative abundance with a dilution rate of 0.5hr^{-1} 123

Figure 6-15: Growth of both strains in the local communities when both migration and selection exists with dilution rate 0.42hr^{-1} . The legend shown for reactors 1 and 2 applies to all reactors – the red data points depict the Cm+ strain and the blue data points the Kan+ strain. The missing data points depict nighttime when sampling could not be conducted..... 124

Figure 7-1: Simulation results from the stochastic model with 0.1 as migration rate. Graph a) shows the relative abundance of both species as the simulation runs while b) shows the relative occurrence of each species at specified relative abundances. 137

Figure 7-2: Simulation results from the stochastic model with 0.4 as migration rate. Graph a) shows the relative abundance of both species as the simulation runs while b) shows the relative occurrence of each species at specified relative abundances. 138

Figure 7-3: Simulation results from the stochastic model with 0.8 as migration rate. Graph a) shows the relative abundance of both species as the simulation runs while b) shows the relative occurrence of each species at specified relative abundances. 139

Figure 7-4: Simulation results from the stochastic model with 0.75 selective advantage and no migration into the community. Graph a) shows the relative abundance of both species as the simulation runs while b) shows the relative occurrence of species at specified relative abundances. 141

Figure 7-5: Results from the stochastic model when drift alone occurs (no migration rate and selective advantage). Graph a represents two realizations, showing that each species can be the dominant one and this is solely based on chance, while graphs b & c show the relative occurrence of each species at specified relative abundances for both realizations..... 143

Figure 7-6: Histogram of residuals for the neutral model with immigration from both the experimental chemostats (blue) and a long simulated time series (red)..... 147

Figure 7-7: Graphical result of simulation of the dynamics of both species in the local community modeled using the stochastic process with calibrated parameters..... 148

Figure 7-8: Graph showing that the experimental data from the ‘Non-neutral dynamics in a completely insular community’ experiment fall within the upper and lower limits set by the standard deviations of the calibrated mean. 150

Figure 7-9: Graphs showing the stochastic modeled dynamics of both species in the local community for the ‘Non-neutral dynamics in a completely insular community’ experiment. 151

Figure 7-10: Graph showing that the experimental data from the ‘Non-neutral dynamics with immigration’ experiment fall within the upper and lower limits set by the standard deviations of the calibrated mean. 153

Figure 7-11: Graph showing the stochastic modeled dynamics of both species in the local community for the ‘Non-neutral dynamics with migration’ experiment. 154

LIST OF TABLES

Table 4-1: Table showing the sequences for the primers used in the knockout PCR reactions.	34
Table 4-2: Concentration of DNA in the samples.	36
Table 4-3: Result of recombination reaction after 36 hours at 30°C showing the performance of the PCR products using different concentrations of antibiotics.	40
Table 4-4: Result of strains streaked on different L-agar containing 100µg/ml ampicillin, 10µg/ml chloramphenicol and 20µg/ml kanamycin and grown at 42°C.....	43
Table 4-5: Table showing sequences of the primers used for the copy number analysis.....	50
Table 4-6: Expected PCR product sizes for each primer pair combination in each cell.	51
Table 4-7: Table showing the sequences of the probes, forward and reverse primers in the 5'3' direction of both strains used in the qPCR quantification analysis.	57
Table 4-8: Table showing the ratio in which both bacterial strains were mixed.	59
Table 4-9: Table showing result of RT-PCR for both strains.	60
Table 5-1: Result of growth rates and doubling times of the three strains on three different days.....	66
Table 5-2: Summary of process parameters in the absence of antibiotics generated via Monod Kinetics and growth yield experiments for both strains.	77
Table 5-3: Summary of process parameters in the presence of antibiotics generated via Monod Kinetics and growth yield experiments for both strains.	84
Table 7-1: Table showing simulation results from the calibration models	145
Table 7-2: Resulting process parameters from the 'Neutral dynamics with immigration' calibration model.	146

Table 7-3: Resulting process parameters from the ‘Non-neutral dynamics in an insular community’ calibration model..... 149

Table 7-4: Resulting process parameters from the ‘Non-neutral dynamics with immigration’ calibration model..... 152

Table 7-5: Summary of fitted model parameters for each chemostat experiment using corresponding calibration models. 154

Table 8-1: The key variables measured in the local communities in the three different experiments. 159

Table 8-2: Summary of fitted model parameters for each chemostat experiment using corresponding calibration models. 160

Table 8-3: The number of bacterial clumps that would yield the observed dynamics and an estimate of the number of bacteria per clump. 163

TABLE OF SYMBOLS

N_i	The number of individuals of species i in the local community.
N_T	The total number of individuals in the local community.
x_i	The relative abundance of species i in the local community.
θ	The fundamental biodiversity number, a scaling parameter for the log series distribution.
v	Rate of speciation.
J_m	The total number of individuals in the source community.
P_i	The relative abundance of species i in the source community.
m	Rate of migration in the system.
α	Advantage parameter of the system.
a	Mean replacement time between individual events in the population.
Cm^+	Chloramphenicol resistant strain.
Kan^+	Kanamycin resistant strain.
t_d	Doubling time.
μ	Growth rate.
μ_{max}	Maximum growth rate.
K_s	Half Saturation Constant.
Y	Cell growth yield.

ACKNOWLEDGEMENT

My profound gratitude goes to God Almighty for His sustaining grace, mercy, provision and guidance throughout my stay at Glasgow and the good health provided despite several health scares such as carpal tunnel syndrome and inguinal hernia that required immediate surgery.

I am uttermosty grateful to my supervisors William Sloan and Susan Rosser for letting me on board with the project, their immense guidance through it all and also EPSRC, who provided the funding for this research work. A special thanks to Susan's group – Arnold Lab and other affiliated laboratories and individuals such as Prof Stark, Dr. Colloms, Dr. Olorunniji, Stuart, Craig, and Aileen for their massive help with the molecular biology aspect of this work.

I am grateful to teaching laboratory at Boyd Orr Building, University of Glasgow and the Civil Engineering laboratory at University of Strathclyde for granting me access to their laboratories when ours was under renovation to carry out the chemostat experiments of this study, which involved the continuous growth of the strains.

Thanks to the departmental secretaries for assisting with the various orders on laboratory equipment and the mechanical workshop for assisting with the construction of the experimental setup including John Liddell at the glass blowing workshop.

To my family and dearest friends, thank you all for your patience, good will, moral and financial support. Without you all, this work would have been impossible.

AUTHOR'S DECLARATION

I declare that this thesis is a record of the original work carried out solely by myself in the Department of Civil Engineering at the University of Glasgow, United Kingdom, during the period of March 2011 to August 2015. The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright acts. Due acknowledgement must always be made of use of any material contained in, or derived from, this thesis. This thesis has not been presented elsewhere in consideration for a higher degree.

Chioma Frances Agatha Nnaji

August 2015

1 Introduction

Naturally occurring communities of microorganisms are omnipresent catalysts in nature. They help, amongst other things, to metabolise our food, degrade pollutants and provide nutrients to plants, and experience has taught us how to harness their power in a variety of different technologies in medicine, agriculture and engineering. They are the principal contributors in the biological treatment of wastewater; which is important since water is becoming increasingly scarce due to population growth, climate and land use change and access to clean water is difficult in some regions of the world (Faures, 2006). Therefore, it is imperative that we increase our efforts to keep water clean in a cost effective and sustainable way as possible (Simon, 2003). In wastewater treatment plants, microorganisms consume waste, converting it to CO₂, biomass and other less harmful products (Davies, 2005). Wastewater treatment is the most economically important and extensively used environmental biotechnology technique.

The competitive interactions between populations of bacteria that consume the waste are crucial to the functioning of wastewater treatment plants and the quality of the treated water produced (Saiklay & Oerther, 2004). Yet our understanding of their population dynamics is limited (Curtis et al, 2003). Part of the reason that the population dynamics of microbes have remained opaque was that for most of the last century, when wastewater treatment technologies were being developed, microbiologists relied on culture-based techniques and only a very small proportion of bacterial species have been

cultured. As a consequence the focus was on the physiology of culturable microorganisms that might only have a relatively small role in the function of a treatment plant rather than the ecology of the mixed diverse consortium of organisms. This was not only true for wastewater treatment systems, it was also the case in human health related microbial systems and agriculture. Thus, a body of knowledge upon which to develop fundamental theories of microbial ecology did not emerge. Therefore the coming together, interaction and functioning of these communities, or in other words their ecology, is still poorly understood. Indeed, there is an absence knowledge of core ecological principles and thus we lack the tools and expertise to control the environment to our benefit. A good knowledge of ecological axioms is essential in developing fundamental rules for the population dynamics of complex microbial communities (Prosser, 2007).

Microbial ecologists need not start from an entirely blank canvas in conceiving plausible mathematical descriptions. There is a wealth of literature on the theoretical ecology of larger organisms (from both deterministic and stochastic views) that might be drawn upon. These are not without their own problems and there is still debate, for macro-organisms, on the interplay of factors responsible for community structure and assembly. However, we are faced with the challenge of adapting ecological theories of higher organisms to the microbial scale, taking into account the differences that exist between microbes and higher organisms; for example, their size and reproduction rate (Horner-Devine, 2003). Microbial ecologists' ambitions for the mathematical models they develop have been tempered by the almost overwhelming diversity and complexity of naturally occurring microbial consortia. Thus there

are few who attempt to deploy very detailed parameter rich models like those that often appear in mainstream ecology.

Therefore, for the most part, models of microbial communities are parsimonious and employ prudent simplification of a few key processes that might ultimately be manipulated to effect change. There is much quibbling over semantics in describing ecological processes but, as Mark Vellend points out, for all their perceived complexity, ecological patterns are influenced by only four broad classes of process: selection, drift, dispersal (or immigration) and speciation (mutation) (Vellend, 2010). Teasing out the relative importance of these will have a profound influence on the way microbial communities are manipulated in a wide range of different applications, from engineering to medicine. Previous studies assert strong evidence that microbial communities maybe stochastically gathered, thus models that encapsulate drift or demographic stochasticity and dispersal have been important in this respect (Hubbell, 2001). The most widely studied of these stochastic community assembly models adds the additional, and controversial, constraint of neutrality, where all species in the community have the same specific birth rate. Neutral models are often described as 'null-models'; there to be rejected if a sign of selection can be identified in the data. However, there is a growing consensus that demographic stochasticity cannot be neglected even when selection is at play. Yet nobody has actually observed and quantified demographic stochasticity in the population dynamics of microbes, or any other organism. Demographic stochasticity has only ever been inferred by fitting a model to stationary distributions of taxon abundance at one (Volkov,

2003) or multiple sites (Woodcock, 2007). The inference is not robust, which feeds back into the debate on different processes leading to the same patterns (Rosindell, 2012; Clark, 2007). If we are to verify that stochastic drift and immigration can have a significant effect on diversity and we want to modify it to effect change then we need to observe it occurring, rather than infer it.

The aim of this thesis is to see whether in carefully designed chemostats containing two strains of *E.coli*, neutral and stochastic dynamics can be observed and quantified. The strains have been engineered to have a tunable selective advantage in the presence of two different antibiotics, thus in the chemostats, the importance of stochastic demography when there is very well defined selection is also tested. To give the statistical power required to calibrate the stochastic process model, ten identical chemostats are operated in parallel. In each set of experiments, the chemostats are fed with an identical mix of the two strains and growth media and the population density of the strains in the chemostats are recorded through time. Where the effects of selection are being tested then a constant flow of antibiotic conveys an advantage on one of the two strains. The uniqueness and flexibility of this system to test for both these ecological processes at once would offer a complete and comprehensive insight into the dominant factors responsible for microbial community assembly and structure for organisms in the same functional group and possibly the interplay between them. This type of experimentation could potentially move microbial ecology onto a more predictive basis. The scope of this work is divided into the following chapters:

Chapter 2 details the background literature for this thesis, showing the potential contribution of this work to the field of microbial ecology. In chapter 3, the conceptual design of the chemostat setup for the experiments is presented. Chapter 4 takes an in-depth look into the genetic recombination techniques used to modify the wild type *E.coli* species used in this study, chapter 5 tackles the growth characterization of the modified species and the parameterization thereof for the chemostat equations, which provides a blueprint for the experimental design. Three different experimental setups are detailed in Chapter 6. In each of these the population dynamics of the species are driven by different combinations of ecological processes. The results of these experiments are described but only a brief discussion of the results is given in Chapter 6. Chapter 7 deals with the stochastic modeling and calibration technique to describe the population dynamics of both species under the three different conditions. Lastly, chapter 8 pulls together the experimental and modeling results from the previous two chapters and discusses the biological and engineering implications and chapter 9 proffers opinion on potentially fruitful avenues for further research.

2 Literature Review

This chapter delves into microbial ecology; a fundamental discipline central to the biological treatment of wastewater. It reviews ecological principles that are thought to explain microbial community assembly and structure and thus, in principal, provide the basis for quantification of ecological processes. Lastly, the approach used in this study in attempting to understand and quantify the stochastic effects in microbial assembly is introduced.

2.1 Wastewater Treatment

Human sustenance is dependent on water; an essential commodity for many activities. The soaring increase in the human population is posing a challenge for accessibility to clean and safe drinking water, which water engineers are redressing (Tchobanoglous, 2003). A report given by Addams et al. (2009) suggests that by 2030, in some parts of the world, the demand for water could exceed its supply by about 50%. A key aspect of remedying this global issue is the treatment and recycling of wastewater.

Biological treatment of wastewater has been at the forefront of biotechnological processes for over a century, and microbes are the key drivers of this process (Martin, 1927). The activities of mixed communities of microbes in wastewater treatment plants are responsible for purifying wastewater by transforming the organic waste into biomass, CO₂ and other,

less harmful products. Importantly, treatment processes can be engineered so that the microbes remove nutrients like nitrogen and phosphorus that can be harmful to the environment if released into rivers, lakes or the sea. However, it is also known that microbes cause some of the crucial problems that arise during the operation of the treatment plants. For instance nitrifying bacteria, which are difficult to cultivate, and sensitive to pH and temperature changes have been reported to disrupt the nitrification process in the wastewater treatment plant (Wagner, 2002). Thus, the abundance and species composition of resident microbial communities and the quality of the final effluent produced by individual treatment plants can vary based on different physical designs and operating conditions employed at each wastewater treatment plant. Central to the proper functioning and smooth operation of treatment plants is a good understanding of the diverse communities of microbes that exist in them, their assemblage, interactions with each other and behaviour; in other words, their ecology (Graham & Smith, 2004).

2.2 Ecology

Ecology has spawned a suite of different sub-disciplines, many of which claim distinctive viewpoints. However, for the purposes of this review, I will broadly divide ecology into two perspectives: the deterministic and the stochastic. These are broad and not necessarily mutually exclusive, but there have been times in the history of theoretical ecology where the two perspectives have clashed, resulting in highly polarizing and heated debates; not least after the

publication of Hubbell's 'Unified Theory of Biodiversity and Biogeography' (Hubbell, 2001).

The deterministic perspective has a much longer pedigree and is typified by the classical ecological niche theory, formulated by Gause (1934). It states that two species competing for the same limited resource cannot coexist if other ecological factors remain constant; over time one will out compete the other (Gause; 1934). Extending the deterministic perspective, most theories assert that resource partitioning, niche overlap and competition among species of organisms are the principal forces driving community assembly and structure (Sinclair; 2006).

On the other hand, the stochastic, or neutral, perspective proposes that within broad clades of organisms, competition for resources is less important and can sometimes be completely ignored. Instead, random immigration, stochastic demography and neutral speciation shape the community structure. This perspective ultimately has its roots in Kimura's neutral theory in population genetics (Alonso; 2006). It is also the underlying principle of McArthur and Wilson's (McArthur & Wilson; 1956) theory of Island Biogeography, which predicts species richness on islands as a balance between immigration and extinction. In essence, Hubbell's theory (Hubbell, 2001) merely pulls together the concepts of island biogeography with some of the mathematical rigour of Kimura in a single unified theory of biodiversity and biogeography (UNTB) to predict not only richness but also the relative abundance of species in natural communities. He successfully applied this to

abundance data of tropical forest trees and shrubs on Barro Colorado Island in Panama. Hubbell does not suggest that species differences do not exist but in leaving these differences out of his theory he argues that they are not always the driving forces that shape a community. Rather he postulates random processes as having the upper hand in shaping a community, should both deterministic and stochastic factors be present (Hubbell, 2001).

Surprisingly, UNTB has provided as good a fit to real-life observations as the niche theories despite its controversial, simplifying assumption that species in the same trophic level are ecologically equivalent (Bell, 2001). This aspect has attracted its main criticism from contemporary researchers (Chave, 2004; Dornelas, 2006). However, given its success in predicting relative species abundance, there have been calls for the unification of the UNTB and niche theories, rather than a polarized dialogue on which of the two theories best explains biodiversity. Researchers such as W. Stan Harpole have called for the merging of both theories by saying “It’s not niche or neutral – both things are happening. It’s determining the relative importance of the two” (Harpole & Tilman, 2006). McPeck has also been quoted saying “What’s missing from each theory is the other” (Leibold & McPeck, 2006). Merging of the models is already underway; for instance, Tilman extended the niche theory to include some stochastic elements. His model combines the effect of demographic stochasticity and resource competition on community structure (Tilman, 2004). Mouquet added dispersion elements to a niche-based model to account for the relationship between local and regional species richness (N. Mouquet & M. Loreau, 2003).

Both Tilman and Mouquet adopt a niche perspective and add some stochasticity to their models. This means that their models rely on a huge number of parameters that I would argue cannot be measured, especially for microbial communities. This over parameterization of models is a major concern to ecology as a discipline; uncertainties in immeasurable parameters lead to an inadequacy of its predictive powers (Palmeri, 2013; Jorgensen, 2001; Jorgensen, 2009). Most parameters are immeasurable because many of the ecological processes, or at least the conceptual models of them, are nebulous for most ecosystems. This contrast with physical and chemical processes that are typically more keenly observed and thus lend themselves to rigorous mathematical description and precise quantification. In ecology, the uncertainties in the process are often compensated for by giving 'measured' parameters as intervals rather than fixed values (Jorgensen, 2001). More often than not the parameters are not measured but are fitted and, with over-parameterised models, the likelihood of the calibrated parameters being representative of any physical/biological reality diminishes as the number of free parameters increases.

Both the deterministic and stochastic approaches have been applied to microbial communities. The results, like those for larger organisms, suggest a combination of forces at play. However, there does appear to be stronger evidence for demographic stochasticity being one of the driving forces in microbial community assembly in many environments including: desert regions (Caruso, 2011), soil (Ferrenberg, 2013; Beck, 2015) and wastewater treatment plants (Ofiteru et al, 2010). From the study carried out by Caruso et

al on desert microbial communities, stochastic processes had the most effect on the assemblage, though deterministic processes were also present.

Rampal Etienne, who has been at the forefront of theoretical developments in stochastic ecology for more than a decade, suggested that the assumption of ecological equivalence be discarded from UNTB when combining both models but elements of stochasticity and dispersion should be retained (Etienne and Olf, 2004). The implication of this idea is that that even when species differences are accounted for then stochasticity demography would play a significant role in shaping the structure. The first attempt to demonstrate this was by Sloan et al (2006) and aimed at microbial communities. Not only did they present a unification of both models as suggested by Etienne, they also extended it to suit microbial communities by making provisions for their large population size. This is the model being presented here - the assumption of ecological equivalence among species has been relaxed. In its place species are allowed to have competitive advantage (α) over each other – there species differences are the basic element of the niche theory, while also retaining elements of stochasticity such as migration, random births and deaths (Sloan *et al.*, 2006). The model also features the concept of island biogeography comprising of a large source community and local communities. Throughout the thesis the terms source community and meta community are used interchangeably. This reflects the fact that for microbes any source community is likely to be spatially distributed and thus subject to similar ecological forces as those that shape the local communities that are being explicitly modeled. The model is described in more detail below.

The local community is assumed to be saturated with a total of N_T individuals. For the dynamics of the local community to change an individual leaves the system either by random death or emigration. Since the zero sum assumption is maintained, the individual that left the system is immediately replaced by immigration from the meta community with a probability m , or a random birth within the local community, which could be from the same species or another species with a probability $1 - m$. Thus the local community is continually maintained through the processes of death, migration and birth. Figure 2-1 illustrates this principle.

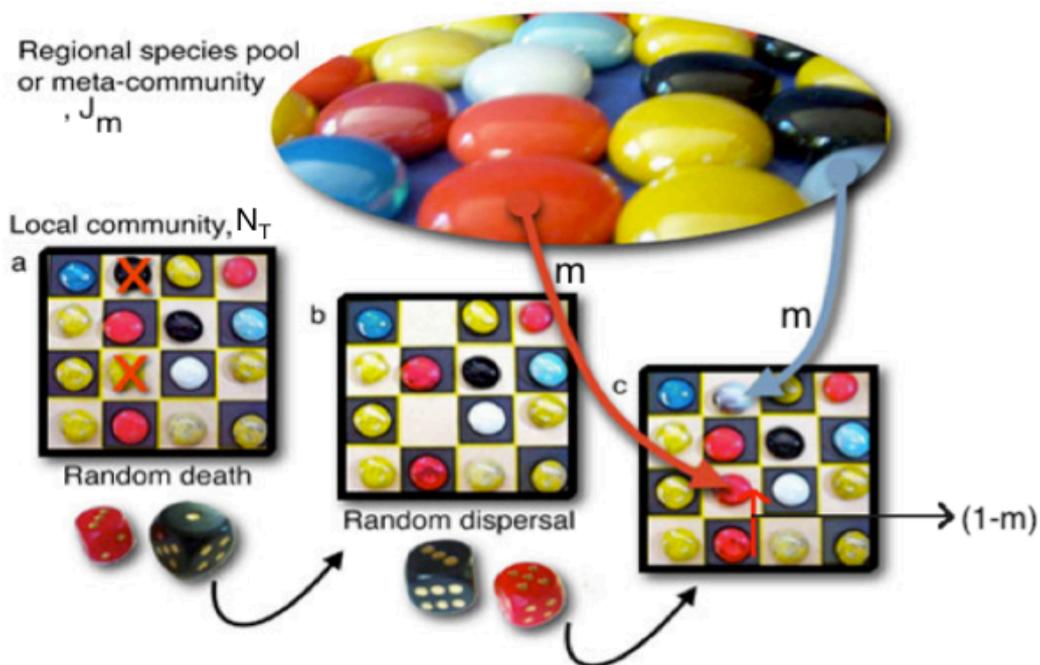


Figure 2-1: Schematic representation of the unified theory of biodiversity and biogeography.¹ The local community represents a subset of the species found in the source community. Here they are made up of 16 individuals from 6 species. (a) Random deaths occur in the local community creating space. (b) The spaces are filled randomly either by local birth or (c) random dispersal from the source community with a probability m . Random speciation can also occur leading to the formation of new species. These processes occur in each time-step bringing about changes in species abundance over time.

¹ cf. Harpole, W. S. (2010).

² 1.0 dsDNA \approx 50 μ g/ml at 260nm.

³ 1L basis of Tryptone broth is made of 10g tryptone and 5g sodium chloride.

Considering the transition probabilities at a single time-step for the i th species with abundance N_i in the local community and relative abundance p_i in the meta community, the probability for species i to lose an individual in the local community is given by equation (2:1)

$$Pr \{N_i - 1/N_i\} = \frac{N_i}{N_T} \left[m(1 - p_i) + (1 - \alpha_i)(1 - m) \left(\frac{N_T - N_i}{N_T - 1} \right) \right] \quad (2:1)$$

where

$Pr\{N_i - 1/N_i\}$ = the probability that species i decreases by an individual.

$\frac{N_i}{N_T}$ = the proportion of species i in the local community.

$m(1 - P_i)$ = the probability of the migration of another species other than species i .

$(1 - \alpha_i)$ = the decrease in the probability of birth in the i th species due to some competitive advantage of other species over species i .

$(1 - m) \left(\frac{N_T - N_i}{N_T - 1} \right)$ = the probability that a migration event does not occur and there is a birth from a species other than species i .

For species i to gain an individual in the local community there must be a death/emigration of some other species followed by either an immigration of the i^{th} species or a local birth in the i^{th} species. Thus the transition probability is given by equation (2:2)

$$Pr \{N_i + 1/N_i\} = \left(\frac{N_T - N_i}{N_T}\right) \left[mP_i + (1 + \alpha_i)(1 - m) \left(\frac{N_i}{N_T - 1}\right) \right] \quad (2:2)$$

For the abundance of species i to remain unchanged after a single time-step then the birth, death or immigration does not involve an individual from species i . Thus

$$Pr \{N_i/N_i\} = \frac{N_i}{N_T} \left[mP_i + (1 - m) \left(\frac{N_i - 1}{N_T - 1}\right) \right] + \left(\frac{N_T - N_i}{N_T}\right) \left[m(1 - P_i) + (1 - m) \left(\frac{N_T - N_i - 1}{N_T - 1}\right) \right] \quad (2:3)$$

In addition, Sloan transformed Hubbell's UNTB model from a discrete to a continuous form to accommodate the large population size of microbes (Sloan; 2006). This was done by letting

$$x_i = \frac{N_i}{N_T} \quad (2:4)$$

So that x_i is the relative abundance of the i th species in the local community. Then using methods developed by Kimura and Ohta (1971), originally for population genetics, it is possible to show (Sloan et al., 2006) that the probability distribution for the abundance of the i^{th} species can be approximately described by a partial differential equation,

$$\frac{\partial \phi(x_i, t)}{\partial t} = \frac{\partial (M\phi(x_i, t))}{\partial x_i} + \frac{1}{2} \frac{\partial^2 (V\phi(x_i, t))}{\partial x_i^2}, \quad (2:5)$$

where $\phi(x_i, t)$, is the probability density function and t is time with 1 unit of time corresponding to one replacement in the community,

$$M = \frac{m(p_i - x_i)}{N_T}, \quad (2:6)$$

and

$$V = \frac{2x_i(1 - x_i) + m(p_i - x_i)(1 - 2x_i)}{N_T}, \quad (2:7)$$

The case where species have an advantage is not given here; it is rederived specifically for the experiment conducted in this thesis in Chapter 7. Sloan et al (2007) extended the neutral case to the Fokker-Planck equation to describe the joint probability density function for all species in the local community. This led to them to show that, at steady state, the taxon abundances are distributed according to a Dirichlet distribution, which yielded a method of fitting the model to observed data. Subsequently Holmes et al. built on the fact that neutrally assembled communities are Dirichlet distributed to derive a Bayesian approach to fitting neutral community models (Holmes et al, 2012). Whilst the review above describes the dynamics involved in the assemblage and structure of the local community and successful attempts to adapt them to microbial communities, in his UNTB postulations, Hubbell also postulates the

dynamics for the meta community where he supposes that it does not differ much from that of the local community, except that the replacement of species is brought about by randomly occurring speciation events (Hubbell, 2001). Hence, the diversity within the source community occurs as a rare point mutation, where species appear spontaneously at a constant average rate. By assuming small speciation rates and taking the limit as the population size and age of species in the community tends to infinity, he shows that the species abundance follows a log series distribution. In a source community of size J_m and speciation rate v , the probability density function of a randomly selected species i , with relative abundance p_i , is given by,

$$f(p_i) = \frac{\theta}{J_M} p_i^{\frac{\theta}{J_M}-1} (1 - p_i)^{\theta - \frac{\theta}{J_M}-1} \quad (2:8)$$

where θ , the fundamental biodiversity number, which describes the sampling distribution of the species abundance, is given as

$$\theta = J_M \frac{v}{1 - v} \quad (2:9)$$

As an approximation for small speciation rates, this is often written as

$$\theta = J_M v \quad (2:10)$$

The fundamental biodiversity number (θ) is often estimated in many research studies along with other parameters such as migration rate (Etienne; 2005) but for the purpose of this thesis, the experiments presented in the next chapters are designed such that what happens in the source communities can be dictated, thus providing an accurate representation of the population dynamics in the community so that θ becomes irrelevant for the purposes of this study.

In all but one case neutral models have been calibrated using stationary taxon abundance distributions. The only exception is for microbes in a wastewater treatment plant (Ofiteru *et al.*, 2010). Rosindell *et al.* (2011) points out that observing neutral dynamics will be the ultimate test of whether neutrality is important. It is also critical to determine the speed of neutral dynamics in contexts like engineering where the ultimate goal is to manipulate communities to effect change. As Rosindell pointed out, UNTB has the potential to influence research in various fields such as conservation, paleontology and phylogenetics. The combination of the temporal dynamics of species abundance distribution from the UNTB and coalescence techniques could shed more light on taxonomy, which would be beneficial to understanding fossil records while its combination with phylogenetic data could provide information on the evolutionary development of species and the abundances of the species involved. Also, by extending UNTB to predict species abundance distributions alongside species richness rather than just species richness alone, new meanings to MacArthur and Wilson's theory of

island biogeography could be unearthed (Rosindell, 2011). The controversy around the fitting of neutral models to data is discussed in the next chapter.

Using this unified model, in the case where species have different growth rates, as is the case in nature, could allow for the prediction of microbial population dynamics in open systems such as wastewater treatment plants. Thus inoculation strategies could be determined through predictive modeling rather than trial and error. The type of detailed experimental study combined with modeling that is described in the remainder of this thesis has never previously been done and will be critical in pinning down aspects such as the rate of stochastic dynamics and the relative importance of stochasticity. The experimental results on species dynamics can be compared with the model and thus an assessment can be made of the model's ability to accurately predict microbial population dynamics, or its limitations in doing so.

The results of this thesis can also be invaluable in the engineering practice of bioaugmentation, predicting the fate of new species introduced to improve a process against resident communities in a system, for instance the introduction of *Nitrosomonas* to out perform other nitrifiers and hence enhance the nitrification process in a wastewater treatment plant. The detailed description of stochastic model calibration used in this work can be found in subsequent chapters.

3 Conceptual Design of Experiments

This chapter lays out the rationale for the experiments that are conducted in the remainder of the thesis. It also gives details of the equipment and the experimental setup that have been devised to test our hypotheses.

From the literature review it is apparent that there is a growing acceptance that randomness has a role to play in shaping biological communities. Yet there still seems to be confusion around the source of randomness and that becomes manifest in the language used in academic papers. As outlined in the literature review, neutrality in population ecology means that the specific birth and death rates across all taxa in a system are the same. The consequences of this, when built into a stochastic birth-death-immigration model (Hubbell, 2001), is the emergence of neutral dynamics. In toned-down versions of this where the neutral assumption is relaxed, the stochastic element, known as demographic stochasticity, can still be strong. The critical factor that is often neglected is that demographic stochasticity produces serial correlations in the dynamics of populations that can in principal be related back to key properties of the system. In most of the published studies that try to tease apart the relative importance of ‘Niche .v. Neutral’ processes there is actually no test of neutral dynamics. Indeed, there is often the erroneous assumption that neutrality will lead to a random distribution of taxa in space and time and that if this is not observed, it is inferred that non-neutral assembly dominates (Dornelas; 2006). In studies that specifically aim to test whether neutral process have shaped the community, then a suite of

sophisticated calibration tools have emerged that fit neutral taxa-abundance distributions to the observed data; a good fit is used to suggest that the community has been neutrally assembled (Etienne *et al*, 2005; Harris *et al*, 2015). But this inference is not entirely satisfactory; other processes may well yield the same, or very similar taxa abundance distributions and the underlying stochastic dynamics has not been observed. Part of the problem is that for macroscopic creatures, generation times are too long to realistically conduct experiments to observe stochastic dynamics. Indeed, from Nee's theoretical study of the time scale for the observance of UNTB in a source community of tree species, he postulates that UNTB would occur over an inordinate length of time (Nee, 2005). He arrived at this conclusion by drawing parallels between alleles and tree species and using the equation describing the time scale of generations for the average age of alleles to reach a given relative abundance derived by Kimura for population genetics. He suggests that for UNTB to be realistic, then the total number of species in the source community N has to be lower than what is actually observed (effective population size) by assuming it is a harmonic mean, which fluctuates as expected for stochastic models. Or a case whereby species have their specific niches but the dynamics within each niche is purely neutral as this lowers N while increasing diversity among species (Nee, 2005). The time scale for UNTB to occur in microbial communities would be verified in this study.

The beauty of working with fast growing microbes is that the dynamics over short periods of time may well reveal whether or not neutral dynamics are shaping the abundance patterns. In the one study where a neutral model was calibrated against the dynamics of microbes in a Californian sewage works, it appeared that neutral dynamics did play a role in microbial community assembly especially migration driven drift as it accounted for much of the variance in the observed dynamics of two functional groups (AOB and heterotrophs) but that difference in organisms' specific growth rates as a function of environmental variables were also apparent (Ofiteru *et al.*, 2010). They arrived this conclusion by examining time series data of both functional groups in addition to their taxa abundance distribution and used a model whose core was that of neutral dynamics layered with environmental effects as opposed to deterministic based models which were extended to include migration or stochasticity and hence provide ambiguous fit to data. Given the size of a sewage works, with a population of perhaps 10^{18} individual bacteria it is perhaps surprising that what appeared to be neutral dynamics were detected.

All of these warrant experiments that are set up with the explicit aim of observing neutral dynamics in a simplified system where, if they do occur, they should be easily identified. In this study, a system monitoring the dynamics of a very simple community comprising just two species – strains from a single wild type *Escherichia coli* K-12 species is set up. One of the strains has been made resistant to the antibiotic chloramphenicol (Cm+) while the other is resistant to kanamycin (Kan+). When no antibiotics are present, then both strains should have exactly the same growth kinetics. In addition,

however, the capacity of the resistant strains to function normally and reproduce in the presence of their respective antibiotics gives a competitive advantage in terms of birth rate when that antibiotic is present in a mixed culture. This allows a high degree of tuning in the competitive advantage that one strain has over another and hence defines the deterministic factor controlling the community dynamics. Thus one can look at the role of stochastic dynamics when species are competing.

The aim is to emulate the island biogeography model, with large, pure cultures of the two strains serving as source communities. The cultures are mixed in known proportions en route to a suite of smaller local communities (islands) where they deliver a constant feed of immigrants. The flow rates from the pure cultures to the smaller mixed growth cultures define the migration rates of species from the source communities to the local communities. The growing evidence that demographic stochasticity (i.e. randomness in the birth, death and immigration processes) plays a significant role in shaping the community composition means that the trajectories of the population dynamics in any two local communities are very unlikely to be identical. Thus it is imperative to have the statistical power afforded by many repeat experiments. Therefore ten local communities were constructed.

This is the first set of highly replicated simple experiments where one can expect to be able to quantify stochastic dynamics should they occur. Importantly, the strains, when grown in a non-selective media, should be neutral with respect to each other, but they can be selectively excluded when

grown in media to which either chloramphenicol or kanamycin has been added. In doing so, characteristics of the niche (differences in species) and neutral theory (ecological equivalence), have been preserved in the model. This robustness makes this system appealing for testing the ecological theories discussed above.

In order to incorporate the migration element of the neutral theory, an apparatus appropriate for bacterial growth, which also ensures continuous introduction of species to its growth chamber, is required. The chemostat is a classical device that has been used on numerous occasions to study the dynamics of microbes as it achieves and maintains steady state conditions which ensures the same environmental conditions throughout the duration of the experiment (de Crecy et al; 2007).

Chemostat

Monod (1950) and Novick and Szilard (1950) were pioneers in the development of microbial continuous culture devices. The chemostat, an example of a continuous culture device, enables the growth kinetics of microbes to be studied in great detail. This is because it provides a continuous sterile condition for the cultivation of microbes by offering an optimum environment (temperature, oxygen level, pH, and nutrients) where microorganisms can grow at a constant or near constant rate and population density. The volume of the culture in the growth chamber is maintained at a desirable constant level by a continuous inflow and outflow of media through the chamber at a constant flow rate (de Crecy et al, 2007). Nutrients are fed to

the organisms in the growth chamber through the inlet tube while organisms and feed leave the chamber through the outlet tubing into a collecting waste vessel where the organisms can be analyzed. The chemostat is preferred to batch-wise cultivation of microbes because it offers its operator complete control over the growth kinetics of the organisms. The flow rate controls the growth rate of the microbes in the growth chamber while the concentration of the growth-limiting nutrient determines their population density (Smith & Waltman, 1995).

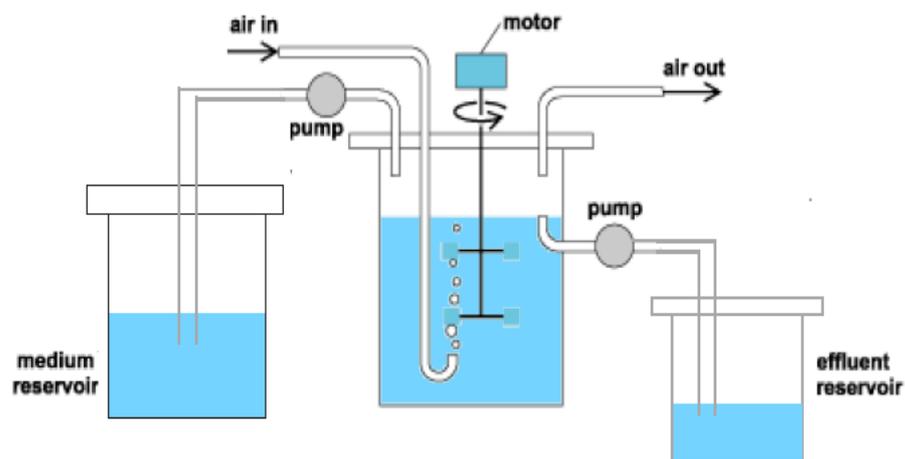


Figure 3-1: Schematic representation of a chemostat showing all three chambers – the feed reservoir, growth chamber and effluent reservoir along with accessories such as pumps and connecting tubes.

The chemostat has a central role in many industrial applications such as the food industry, pharmaceuticals, manufacturing and research. It is a useful tool in ecological studies because, if suitably controlled, it can provide a laboratory based model of real environmental conditions. It is used extensively to model

wastewater treatment systems and more generally it offers simplified controllable conditions to study the interaction between components of complex ecosystems (Smith & Waltman; 1995) and evolution of populations (Lenski; 1986). The importance of the chemostat for ecological studies is echoed in the words of Williams below.

“It is a dynamic system with continuous material inputs and outputs, thus modeling the open system character and temporal continuity of nature. The input and removal of nutrients analogs the continuous turnover of nutrients in nature. The washout of organisms is equivalent to non-age specific death, predation or emigration, which always occurs in nature”. (Williams, F. M; 1971).

The chemostat can be realistically divided into three chambers – the feed nutrient reservoir, the growth chamber and the collecting reservoir – all connected by tubing as shown in Figure 3-1. Microbial reproduction and growth occur in the growth chamber due to the continuous supply of nutrients from the reservoir. The nutrients supplied to the organisms are in sufficient concentrations except one that is in limited supply, often described as the growth-limiting nutrient, which ultimately influences the growth rate of the organism. The collecting reservoir ensures that the volume of culture in the growth chamber is constant, implying that the flow rate, a crucial growth parameter, can be varied (Andrews, 1997). Hence, the chemostat is an attractive piece of equipment for studying the growth of microbial population even for long periods of time such as Lenski, who monitored the growth of 12 identical populations of *E.coli* in a chemostat for 60,000 generations and saw evolution of genetic changes such as improved fitness, defects in DNA repair

leading to mutator phenotypes (Lenski). Detailed mathematical description of bacterial growth in a chemostat is presented in Chapter 6 of this thesis.

Figure 3-2 shows the framework for the design of the multiple chemostat experiments introduced above that deploy to emulate island biogeography. In the experiments there are two large identical chemostats, where pure cultures of both strains are cultivated. These are labeled Cm⁺ and Kan⁺ reactors in the diagram. There are also ten small identical chemostats for repeatability containing a mixture of both strains, which are supplied by dispensing pumps at chosen flow rates. The growth rates achieved in the large reactors (source communities) are controlled by the flow rates of the pumps connecting them to the feed and the cell density is controlled by the concentration of the growth-limiting nutrient in the feed. To maintain constant volume and population density in the local communities, as required for neutral theory, there are outlet tubes from the local communities to a waste vessel, which is aided by a pump. Lastly, when introducing selectivity into the system, the desired antibiotic concentration is added to the feed, which is then be distributed evenly to the reactors by the pumps rather than adding the antibiotics individually to each reactor. This helps minimize human error and also any variations in concentration or volume of antibiotics added will be the same across all reactors.

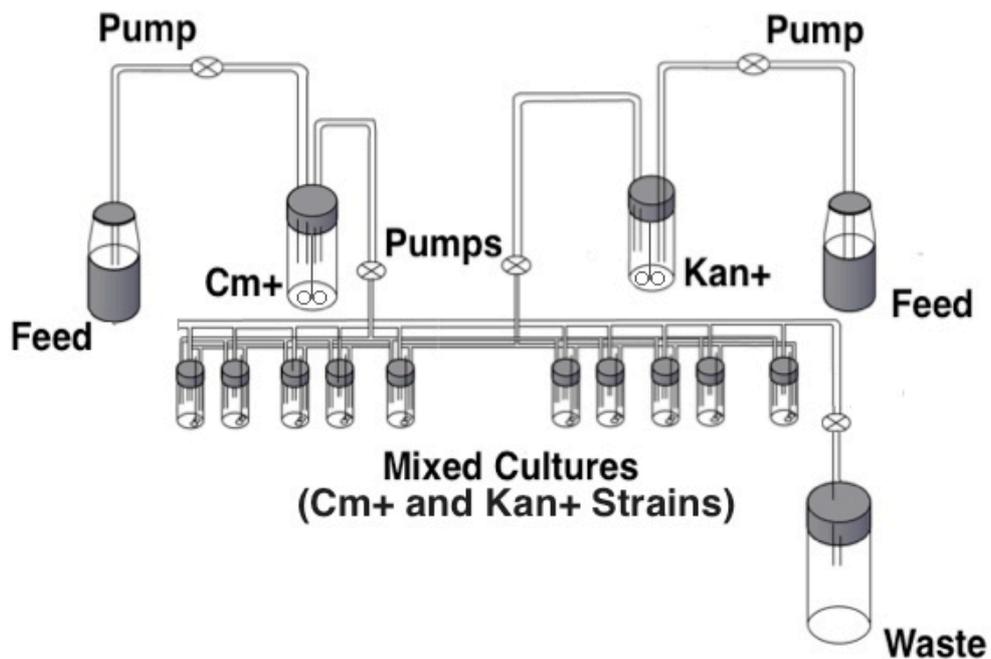


Figure 3-2: Conceptual design of the chemostat experiments used in this study.

For the experimental design to work is critical that the antibiotic resistant *E.coli* strains are engineered to have identical growth characteristics when they are not exposed to the antibiotics. In addition the growth characteristics of the modified strains under different conditions need to be characterized, so as to generate process parameters that enable the appropriate pumping rates and reactor sizes in the experiments to be determined in advance.

4 Genetically Identical Organisms

In order to track the dynamics of neutral populations and to tune the fitness of bacterial strains then it is imperative to create two populations that can be distinguished and whose fitness can be modified in a controlled manner. This chapter describes the molecular microbiology deployed in generating these populations and the molecular analysis used to subsequently identify and quantify the abundance of each strain at any given time in each reactor during the chemostat experiments.

4.1 Genetic Recombination

A community of microorganisms where two populations of species can be exposed to conditions that will induce one to have an advantage over the other or alternatively, in benign environments, have equivalent specific growth rates (neutral) is required. To achieve this, a wild type strain of *Escherichia coli* (DS941) was used and a non-essential gene in its genome replaced with two different antibiotic resistance genes, one for each species. This implies that a site in the genome of the wild type organism was deleted and an antibiotic resistance gene incorporated in its place for each species. This process is called recombination. A genetic recombination process whereby, based on similar nucleotide sequences on the same double stranded DNA molecule, a DNA exchange can be carried out between them is termed homologous recombination.

4.1.1 Recombination-mediated Genetic Engineering

Recombination-mediated genetic engineering (recombineering) is a type of in vivo homologous recombination, which is mediated by phage-encoded lambda (λ) recombination enzymes found in *E.coli*. This technology allows any nucleotide to be inserted or deleted at any location on the DNA sequence, as opposed to classical in vitro genetic engineering, which uses restriction enzymes and requires restriction sites. Thus recombineering is not limited by the location of restriction sites, restriction enzymes or DNA ligases (Sawitze *et al.*, 2007). Recombineering requires bacteriophage recombination lambda (λ) proteins and relatively short DNA homologies approximately 50 bases to bring about effective recombination. It is regarded as a simple, precise, fast, inexpensive and highly efficient method of manipulating chromosomal DNA in *E.coli* (Datsenko *et al.*, 2000).

For the recombineering technology carried out in this project, the wild type organism used here is called DS941, a strain of *Escherichia coli* (*E.coli*) K-12 that is commonly used in research involving DNA recombination. It is also considered to be commercially safe and has no known adverse effect on microorganisms or plants (Casali, 2003). The *pepA* gene of DS941 served as the target gene (the gene to be replaced) in this work because its presence or absence in the genome has no effect on the growth rate of the organism and its function is well known and can be verified, which would be the basis of characterization for the populations and communities. The *pepA* gene encodes aminopeptidase A, a 330 kDa homohexameric multifunctional protein with aminopeptidase and DNA-binding activities, the latter being associated

with mechanisms of transcriptional control and DNA site-specific recombination (Colloms; 2003). Chloramphenicol (Cm) and kanamycin (Kan) antibiotic resistance genes were introduced into the genome of DS941, to create two distinct strains. Both strains are similar to the parent strain and only differ when the antibiotic resistance genes in each are expressed and in turn, they are only expressed in the presence of the antibiotic. Both of these antibiotic resistance genes were used as markers as they are strong bacteriostatics antibiotics, which slow down the growth rate of bacteria or completely inhibit growth depending on the concentration added. Placing the antibiotic resistance genes in the chromosome of the organism offers more stability and ensures that they are always expressed as opposed to having them in a plasmid inserted into the organism.

The recombination was mediated by means of a vector plasmid pKOBEG A, which has an ampicillin resistance gene as a marker. This plasmid carries the phage-encoded recombination enzyme (a recombinase λ Red) that effectively cuts the target gene (pepA) and replaces it with one of interest (Cm⁺ and Kan⁺). Typically during in vivo recombination in *E.coli*, linear DNA introduced is degenerated by endonuclease RecBCD, but the recombinase λ red proteins have properties that allow recombination to occur at very high frequencies especially when short homologies are involved. (Yu, 2000) The bacteriophage λ Red system contains Gam, Exo and Beta proteins.

- Gam protein prevents the degradation of the linear double stranded DNA (dsDNA) that carries the antibiotic resistance gene by inhibiting *E.coli*'s exonucleases, RecBCD and SbcCD, thereby preserving it as a substrate for recombination.
- Exo degrades the linear dsDNA in a 5'→ 3' creating single stranded DNA (ssDNA) with a 3' overhangs.
- The Beta protein binds to the ssDNA produced by exo protein and drives the recombination by promoting annealing of the complementary ssDNA to the homologous genomic target site. It also protects the DNA from single strand nuclease attack (Sawitze *et al.*, 2007).

Thus these enzymes protect the linear dsDNA, cleave it into ssDNA and produce complementary strands to the ssDNA, thereby yielding recombinant molecules. In effect they catalyze the homologous exchange of Cm and Kan resistance genes with the *pepA* gene. The following steps were carried out for the recombination reaction.

- I. Creating a linear dsDNA carrying the desired antibiotic resistance gene and a short flanking oligonucleotide sequence homologous to the target gene (*pepA*) via Polymerase Chain Reaction (PCR).
- II. Introducing the vector pKOBEG A into DS941 cell.
- III. Finally, introducing the linear DNA from PCR into the transformed DS941/pKOBEG-A cells.

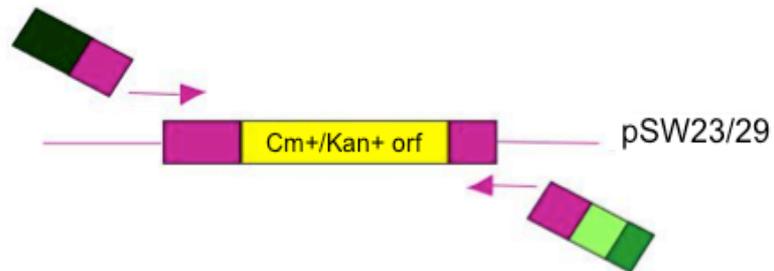
4.1.1.1 Polymerase Chain Reaction (PCR)

Using PCR, the linear DNA sequence of interest can be replicated. A forward and reverse primer sequence with the 5' ends having homology to 50bp of the *pepA* gene and the 3' primer ends having homology to the antibiotic resistance plasmid were designed to be complementary to the chloramphenicol (pSW23) and kanamycin (pSW29) plasmid templates. This was done to yield a linear DNA fragment containing the desired antibiotic resistance genes with short homologies to the target gene (*pepA*) at the flanks. The PCR product having the *pepA* gene at the flanks would enable the homologous recombination machinery (pKOBEG-A) to recognize the homologous sequence to *pepA* and effectively cut the *pepA* gene in DS941 and replace it with the PCR product.

DNA sequence as it appears in DS941 genome



Desired sequence region in DS941 genome



Legend

-  Forward primer in 5' 3' direction
-  Reverse primer in 3' 5' direction
-  Cm+ /Kan+ plasmid open reading frame (orf)
-  Sequences before and after Cm+ /Kan+ orf

Figure 4-1: Illustration showing the target region for genetic recombination in the E.coli genome. I) The *pepA* gene is flanked to the left by *yigP* gene and to the right by *holC* gene. II) The expected outcome of the recombination. III) Generation of linear dsDNA molecules carrying the antibiotic resistance gene, one for each organism via PCR.

Table 4-1: Table showing the sequences for the primers used in the knockout PCR reactions.

Primer	Sequence in 5' 3' direction	Length (bp)
Forward (KF)	attctatctgtagccaccgcccgttgccttaagattcaggagccgtagtgc aagagggtccaacttcacat	72 (50+22)
Forward (nKF)	attctatctgtagccaccgcccgttgccttaagattcaggagccgtagtgc cgagatttcaggagcctaagga	72 (50+22)
Reverse (RF)	acagcctgcctgacgcaattactctcgcggttaaacccagcgcggtt tttctaggcaccaataactgc	71 (50+21)
Control (F)	aagagggtccaacttcacat	22
Control (R)	tttctaggcaccaataactgc	21

The PCR reactions were set up using Taq polymerase, a thermostable DNA polymerase that catalyzes the template-directed polymerization of deoxyribonucleotide triphosphate (dNTPs) at high temperatures, and three sets of primer pairs – two for the main (knockout) reaction and one as a control reaction. The primer pair for the control reaction was designed to be complementary to the antibiotic plasmid templates while the primer pairs for the main reaction (knockout) were also complementary to the antibiotic plasmid templates, with additional sequences incorporated on the flanks that were complementary to the *pepA* gene. The knockout primer pairs had two forward primer variant; Knockout forward (KF), which was further away from the start codon, had a promoter sequence that would drive the transcription of the antibiotic genes once they were integrated into the chromosome, while the other forward primer (new knockout forward – nKF) was without a promoter but was closer to the start codon. This was done to achieve effective PCR synthesis and expression of the antibiotic resistance genes. This implies that

five PCR reactions were set up: two each for the Cm and Kan antibiotic resistance genes and one control reaction. The result of the reaction can be seen below as viewed on a 1% agarose electrophoresis gel. The four PCR generated targeting products from the two sets of primer pairs (Cm/KFR, Cm/nKFR, Kan/KFR and Kan/nKFR) were used for the subsequent experiments.

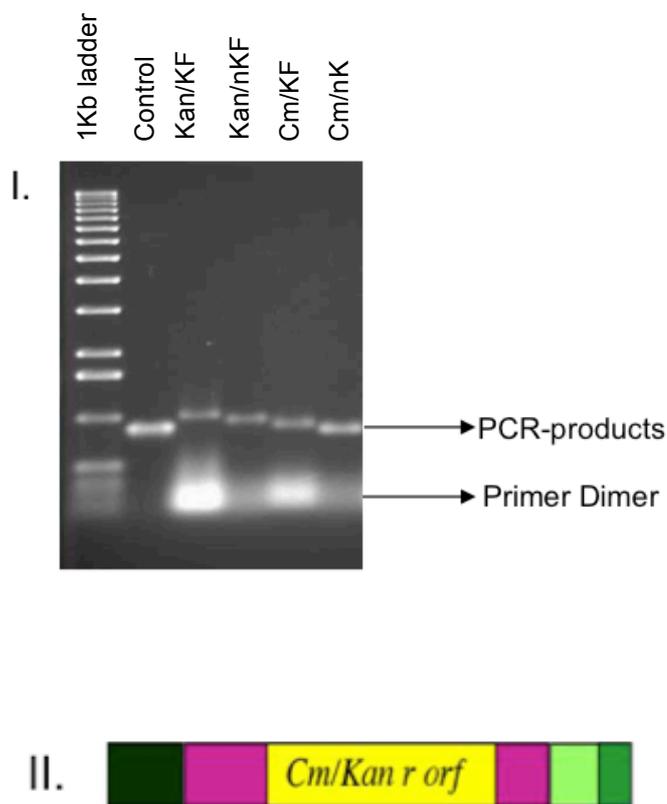


Figure 4-2: I) 1% Agarose electrophoresis gel showing final yield of PCR products. II) Schematic representation of PCR products showing the desired gene (antibiotic resistance) at the center flanked by the target genes.

The PCR products were purified using Qiagen's QIAquick gel extraction kit to remove impurities that might affect the recombination reaction such as nucleotides, unused PCR primers, primer dimers - a by-product of PCR reaction, which results, for example, when self-annealed primers extend. The concentration of the DNA in the four purified samples was determined by measuring the absorbance at 260nm wavelength using a Helios Zeta UV-Vis spectrophotometer from Fisher Scientific. For the analysis, 2 μ l of DNA solution was added to 998 μ l double distilled water (ddH₂O), giving a dilution factor of 500.² The concentrations were between 7.5 μ g/ml – 15 μ g/ml as shown in Table 4-2, while the purified DNA samples are shown in Figure 4-3.

Table 4-2: Concentration of DNA in the samples.

Samples	Absorbance Reading	Converting to μ g/ml at 260nm (x50)	Final concentration in (μ g/ml) (x500)
Kan/KF & KR	0.0004	0.02	10
Cm/nKF & KR	0.0006	0.03	15
Cm/KF & KR	0.0003	0.015	7.5
Kan/nKF& KR	0.0004	0.02	10

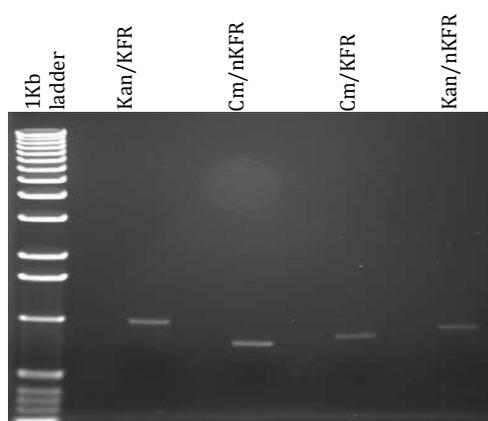


Figure 4-3: Purified PCR products without primer dimers.

² 1.0 dsDNA \approx 50 μ g/ml at 260nm.

4.1.1.2 Transformation of Competent DS941 cell via Calcium Chloride (CaCl₂)

The vector plasmid, pKOBEG-A, encoding the enzymes which will catalyze the homologous exchange of the antibiotic resistance gene with the *pepA* gene, is introduced into the DS941 cell because DS941 does not express any bacteriophage recombination system itself. The process of introducing exogenous genetic material into a cell is called transformation. For cells to undergo transformation, the cell membrane has to be made permeable and the cells are then said to be competent. There are a few organisms that naturally allow exogenous material into them. These are said to be naturally competent (e.g. *B.subtilis*) but the transformation process for these cells is typically slow (Primrose et al, 2006). Artificial competence is a laboratory procedure in which cells are passively made permeable to exogenous DNA, using either chemical treatment or electroporation. Cells that are undergoing very rapid growth are made competent more readily than cells at other stages of growth because they are viable and actively metabolizing. Such cells can be found in the mid-log phase of the growth curve and can tolerate the necessary treatment, as the procedures can be very harsh on the cells (Primrose, 2006).

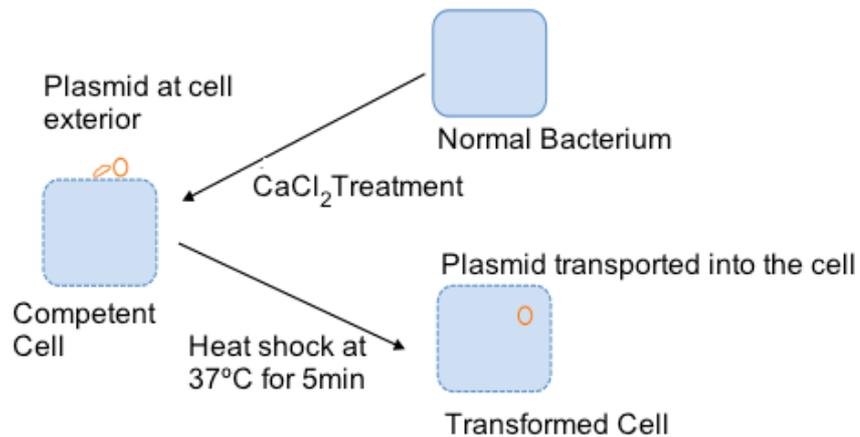


Figure 4-4: Schematic representation of chemical transformation of cells via (CaCl₂)

DS941 cells were chemically transformed by first suspending them in 50mM ice-cold Calcium Chloride solution to make them permeable to plasmid DNA. This makes them competent. Then they were incubated with pKOBEG-A on ice for 20mins and heat shocked at 37°C for 5mins for the plasmid vector to enter the cells and bind to them. They were returned on ice to discontinue the effect of the heat shock. Thereafter, they were grown at 30°C for 30mins to allow the expression of the ampicillin resistance gene prior to plating on agar media containing 100µg/ml ampicillin. This ensures that only the transformed DS941 cells containing the plasmid DNA - pKOBEG-A grows, as it has ampicillin as its marker. After 16 hours there were colonies on the ampicillin agar plate indicating successful transformation as opposed to the control plate consisting of non-transformed DS941 cells streaked out on agar containing 100µg/ml ampicillin, which had no colonies. Thus DS941 now has the vector plasmid and can take up the linear dsDNA constructed via PCR in section 4.1.1.1 above.

4.1.1.3 Transformation of DS941 cells containing pKOBEG A plasmid with PCR products via electroporation

A single colony was taken from the transformant DS941 cell carrying pKOBEG-A and grown overnight at 30°C shaking at 225 rpm. An aliquot of the resulting over night culture was added to 10ml of broth to which 100µg/ml ampicillin had been added and cultured at 30°C shaking at 300rpm to mid-log phase for about 3 hours. Ampicillin was added to prevent the loss of pKOBEG-A plasmid during cell division and to ensure that culture contains cells that have the ampicillin resistance gene. L-arabinose, a five-carbon sugar produced in plants, was added to the growing culture to induce the expression of the λ Red recombinase in pKOBEG-A whose function is to catalyze the recombination. Plasmid pKOBEG-A has the araC-PBAD repressor-promoter assemblage. When L-arabinose binds to AraC protein, a positive and negative regulator, it turns on the transcription from the arabinose promoter PBAD, which enables the expression of the λ Red protein system that facilitates recombination (Guzman et al, 1995).

Due to the higher efficiency of taking up exogenous material in electroporation compared to chemical transformation, electroporation was used for this stage of the process. The cells were made electrocompetent by washing extensively and repeatedly with a mixture of ice cold 10% glycerol solution and centrifuged in a cold rotor at 4°C for 15mins at 4000g to remove all salts. To introduce the PCR products into the cells, they were mixed with the electrocompetent DS941 cells containing pKOBEG-A and then pipetted into a cuvette. They were subsequently placed in the electroporation apparatus,

which had aluminum electrodes, where an electric pulse of 2400volts/cm was applied to the cuvette and its contents. This causes a disruption in their lipid bilayer resulting in the formation of pores in the membrane through which the PCR product enters. LB broth was immediately added to the cells and incubated at 30°C for the antibiotic resistance genes to be expressed before plating. Duplicate cultures were prepared for each of the PCR products (two for each antibiotic resistance gene), thus there were four cultures in all, and each was plated. Cells with the chloramphenicol resistance gene were spread on agar plates containing 5 and 10µg/ml chloramphenicol while those with the kanamycin resistance gene were spread on 10 and 15µg/ml kanamycin. Electroporated DS941 cells with double distilled water in place of PCR products were setup as control. The results obtained after 36 hours are shown in Table 4-3.

4.1.2 Result of Recombination-mediated Genetic Engineering

Table 4-3: Result of recombination reaction after 36 hours at 30°C showing the performance of the PCR products using different concentrations of antibiotics.

Antibiotics resistance gene	Concentration (µg/ml)	Name	No of cells
Chloramphenicol	5	Control	0
		Cm/nKFR	6
		Cm/KFR	9
	10	Control	0
		Cm/nKFR	1
		Cm/KFR	2
Kanamycin	10	Control	>500
		Kan/nKFR	>500
		Kan/KFR	>500
	15	Control	5
		Kan/nKFR	611
		Kan/KFR	178

The table above shows explicitly the performance of the recombination reaction on different concentration of antibiotics for cells using both primers: nKFR (new forward primer), KFR (the previous forward primer) and double distilled water as control. The results from the cells selected for chloramphenicol resistance show that primer KFR yielded higher recombination performance than nKFR since it yielded more colony forming units for both concentrations of antibiotics tested. As expected, more colonies were seen when selected with lower concentrations of the antibiotics. For the cells selected for kanamycin resistance, 10µg/ml seems to be an unsuitable concentration of the antibiotics to draw conclusive results because of the multiple colonies in the control to which no PCR product was added. With the 15µg/ml concentration, the numbers of colonies in the control samples are less compared to the samples with PCR products, with the new forward primer pair - nKFR having more colonies compared to those of KFR. Perhaps primer nKFR expresses the kanamycin resistance gene better.

As a result of growth in the control samples, single colonies from each plate were streaked again on higher antibiotic concentrations to further validate the results. For the re-streaking, higher concentrations of the antibiotics were used - 10µg/ml for chloramphenicol and 20µg/ml for kanamycin. Various colonies were picked at random from the same plate and streaked out on different plates with the same concentration of antibiotics. For instance the six colonies from the Cm/nKFR plate were picked and streaked on six different plates containing 10µg/ml chloramphenicol. This was done to ensure the colonies were not artefacts. The plates were incubated at 37°C for 24hrs to cure the strains of plasmid pKOBEG-A, which is temperature sensitive and

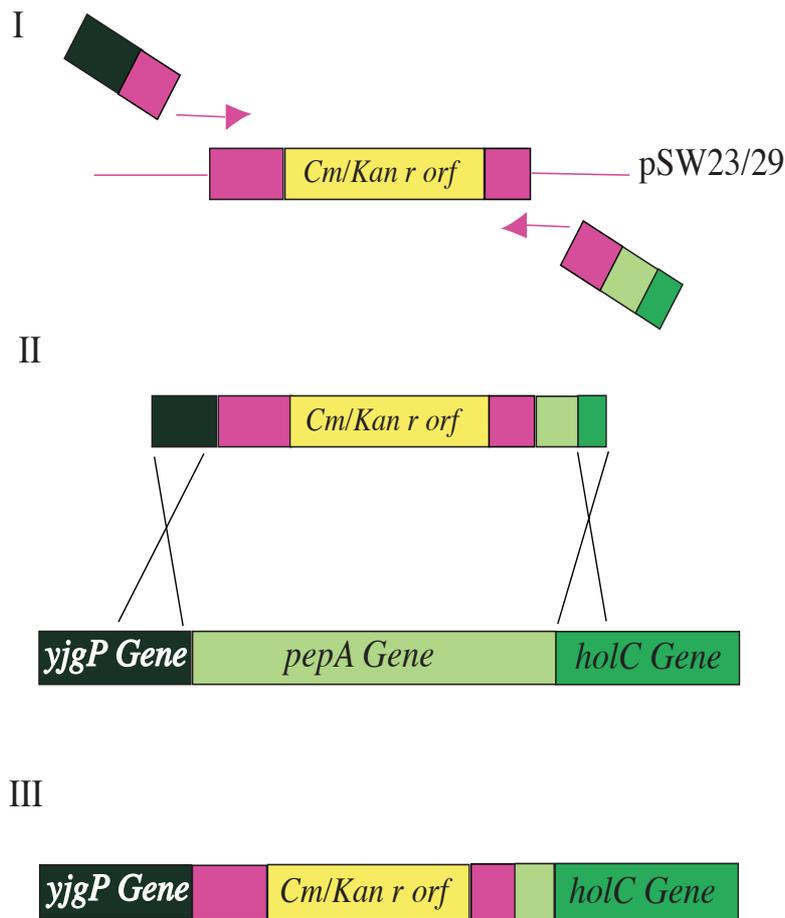
should be lost at 37°C, in an attempt to show that the incorporated antibiotic resistance genes are not on the plasmid. For the control experiment, three colonies from the chemically transformed DS941/pKOBEG-A, which were not electroporated were streaked on L-agar, 10µg/ml chloramphenicol and 20µg/ml kanamycin plates each.

More colonies were recorded on the plates that originated from the KFR primer design compared to the nKFR plates for both antibiotics. This is suspected to be as a result of the KF forward primer having a promoter – a region that facilitates the transcription of a particular gene – of its own which drives the transcription of the resistance gene compared to nKF primer which had to rely solely on the promoter of *pepA* to drive the transcription of the resistance gene. The control (DS941/pKOBEGA) grown on 10µg/ml chloramphenicol and 20µg/ml kanamycin antibiotic plates had no growth while the control on L-agar plates had colonies. This is as expected because the control was not treated with the linear dsDNA carrying the antibiotic resistance genes. The strains were further grown to 42°C to ensure they were completely cured of pKOBEG-A. A subsequent assay was performed to test for the presence of pKOBEG-A by streaking the cells on 100µg/ml ampicillin, 10µg/ml chloramphenicol and 20µg/ml kanamycin plates and growing again at 42°C. It is expedient to lose pKOBEG-A after the recombination, as it is no longer needed. The results are shown below.

Table 4-4: Result of strains streaked on different L-agar containing 100µg/ml ampicillin, 10µg/ml chloramphenicol and 20µg/ml kanamycin and grown at 42°C

Cells	Amp Plates 100µg/ml	Cm Plates 10µg/ml	Kan Plates 20µg/ml
Cm2	-	+	-
Cm3	-	+	-
Cm7	-	+	-
Cm8	-	+	-
Cm9	-	+	-
Kan7	-	-	+
Kan9	-	-	+
DS941/pKOBEG-A	+	-	-

Cm2, 3, 7, 8 and 9 cells have the chloramphenicol resistance gene while cells Kan7 and 9 have the kanamycin resistance gene. From the table above, it can be deduced that pKOBEG A has been cured from the cells and most importantly that the antibiotic resistance genes are not on the plasmid but in the chromosome as the cells only grew on the plates containing the antibiotics they are resistant to. The recombination process used in this study is summarized in Figure 4-5 below.



Legend

- Cm/Kan* resistance plasmid open reading frame (orf)
- Sequences before & after *Cm/Kan r* plasmid orf
- Forward primer in 5' → 3' direction
- Reverse primer in 3' → 5' direction

Figure 4-5: A Schematic representation showing the summary of the recombineering reaction used in this project. I) PCR setup with forward primer consisting of 50bp from *yjgP* and 22bp from *Cm/Kan* resistance gene plasmids in the 5' → 3' and reverse primer consisting of 21bp from *Cm/Kan* resistance gene plasmids, 19bp from *holC* gene and 31bp from *pepA* gene in the 3' → 5' direction. The sequences from *Cm/Kan* resistance gene plasmids on the primers are complementary to those before and after the *Cm/Kan* resistance plasmid orf. The PCR yields (II), which is transformed into competent cells containing pKOBEG-A that has the recombinase λ Red system. Red mediated recombination occurs yielding the recombinant product (III) where *Cm/Kan* resistance gene replaces *pepA* in the genome of DS941 cell.

Having confirmed that the presence of antibiotic resistant inserts in the modified strains, it is expedient to verify that the *PepA* gene has indeed been knocked out and the antibiotic inserts are actually in its place in the DS941 chromosome.

4.2 Confirming the deletion of *pepA* Gene

The *pepA* gene is one of the proteins in the *E.coli* genome responsible for DNA site-specific recombination at *cer* and *psi* sites for plasmid dimer resolution. In the presence of *pepA*, plasmid dimers with two or more *cer* or *psi* sites get resolved into single plasmids, thus they recombine to sizes smaller than the dimers (Schumann, 2006). In this manner, the activity of the *pepA* gene can be verified.

Plasmid dimers are formed when homologous recombination or rolling circle replication occurs between multiple copies of a plasmid in a cell. This is an undesired phenomenon that occurs frequently in cells with high copy number of plasmids during plasmid replication. It leads to unstable maintenance and distribution of bacterial plasmids in a cell. However, many plasmids have plasmid dimer resolution systems that encodes a site-specific recombination system that rapidly resolves any plasmid dimers formed (Schumann, 2006).

Site-specific recombination systems at *cer* and *psi* sites found in multicopy natural plasmids are usually catalyzed by the recombinase proteins XerC, and XerD. Sometimes the accessory protein pairs *PepA* & *ArgR* and *PepA* & *ArcA* are required for the recombination mechanism at the *cer* and *psi* sites

respectively. The accessory protein pair binds to the specific site (cer or psi site) to form a protein-DNA complex that activates the recombinase proteins to carry out strand cleavage and rejoining reactions (Schumann, 2006). Recombination at the psi site can still proceed in the absence of PepA & ArcA proteins as depicted in Figure 4-6 below.

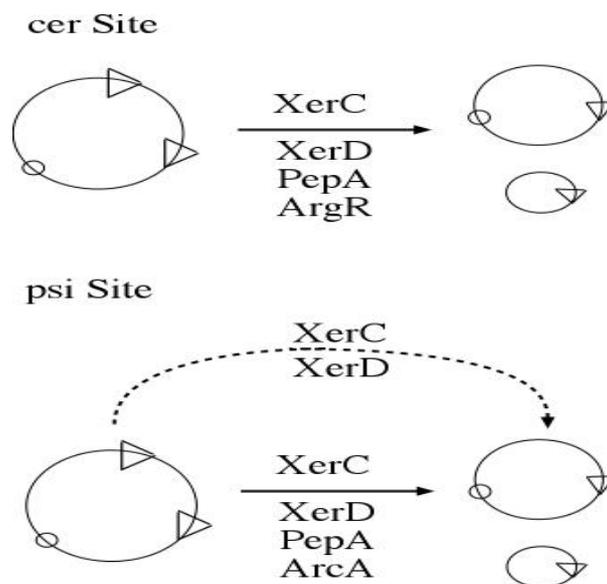


Figure 4-6: An illustration showing site-specific recombination at cer and psi sites and the genes that facilitates the recombination.

E. coli plasmids pJB43 and pSDC134 having two cer and psi sites respectively were used for this experiment. The modified strains together with the wild type DS941 cells were chemically transformed with pJB43 and pSDC134 using calcium chloride (CaCl₂) and their plasmid DNA extracted using the QIAprep Spin Miniprep Kit from Qiagen. The extracted DNAs were run on an agarose electrophoresis gel to check their sizes. Plasmids pJB43 and pSDC134 were also run on the gel to indicate their original sizes. The results shown below.

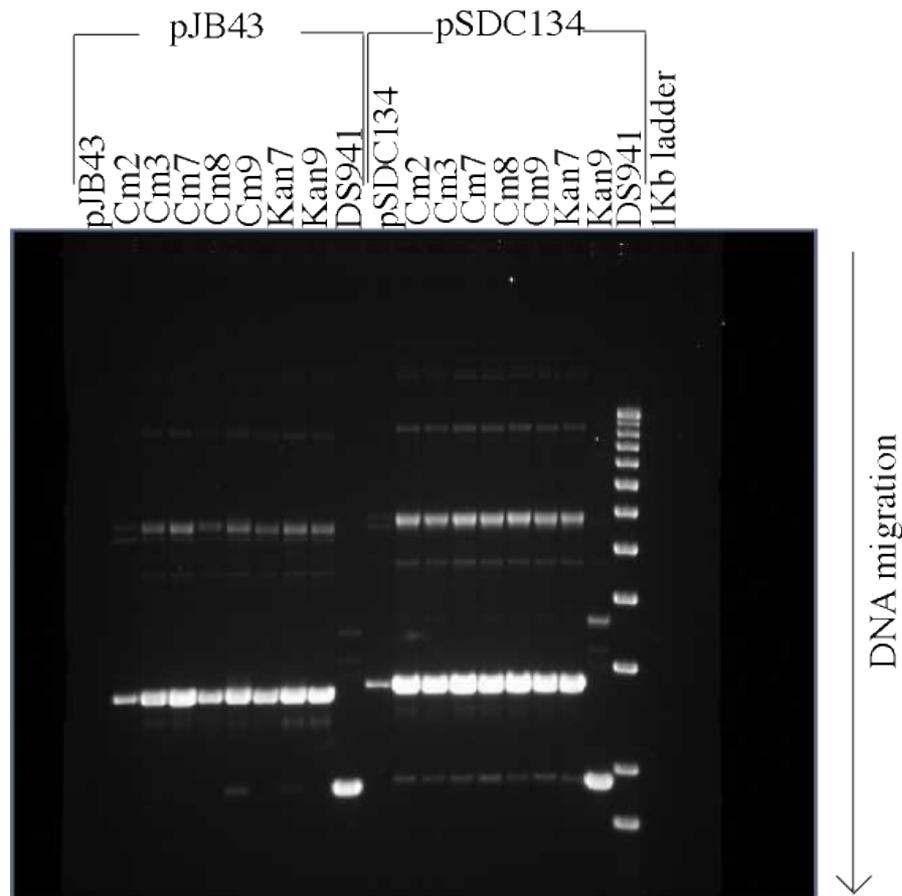


Figure 4-7: 1% agarose electrophoresis gel showing extracted plasmid DNA fragments from the strains and DS941 cells characterized according to sizes.

In the gel above, the first nine lanes are associated with the pJB43 plasmid. pJB43 on the first lane is an indicator of the original size of the plasmid. The next eight lanes represent cells from which the pJB43 plasmid has been extracted. It is clear to see that the plasmids extracted from the cells with antibiotic resistance gene are the same size as the original plasmid. The ninth lane corresponds to DS941 cells from which pJB43 has been extracted. Its band is further down the lane; this shows that its size is smaller than the original plasmid, as smaller DNA fragments travel through the pores of the gel faster than larger fragments when electric current is applied. This is the principle of an agarose electrophoresis gel. The reduction in size of the plasmid extracted from DS941 occurs due to the presence of *pepA* in its

genome. *pepA* within the genome of the transformant DS941 cell carrying pJB43 causes recombination at the *cer* sites of the plasmid, resulting in two smaller sized fragments.

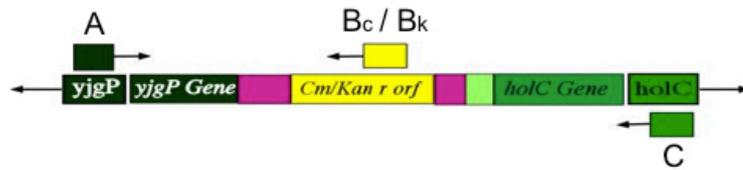
The second part of the gel corresponds to cells from which the pSDC134 plasmid has been extracted and pSDC134 plasmid denotes the original size of the plasmid DNA. The plasmid DNAs extracted from the strains are of similar sizes to that from pSDC134 plasmid, while that from DS941 is smaller. The explanation given for pJB43 holds here also. However, bright bands can be seen lying above the plasmid DNA fragments. This is because pSDC134, which has *psi* sites, can perform site-specific recombination even in the absence of *pepA*. This in effect verifies that the *pepA* gene has indeed been knocked out of the genome of the recombinant strains.

4.3 Copy Number

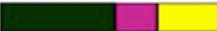
Having verified the presence and activity of the antibiotic resistance genes and confirmed the deletion of the *pepA* gene from each of the modified strains, it is imperative to determine the number of copies of each antibiotic resistance gene insert in each of the strains, though this method of homologous recombination should ensure only one copy of the insert in one cell (Costantino, 2003). It is necessary to know the number of copies of the inserts in each strain because this determines the type of identification and quantification analysis to be used when measuring for the population sizes of both strains when grown in pure and mixed cultures during the chemostat

experiments. It is desired that each modified strain possesses just one antibiotic resistance gene insert as this makes for easy quantification, thus one antibiotic resistant gene identified would be equivalent to one cell.

The verification of the number of copies of each antibiotic insert in each strain was done using PCR analysis. The primers for this analysis were designed to be complementary to both the sequences in the middle of the insertion and those at the ends of the insertion, thus targeting the exact sequences of the antibiotic resistance genes and the whole sequence of the region of the genome where recombination occurred. A total of four primers each with 20bp in length was designed: one forward primer and three reverse primers making for three primer pairs. Two primer pairs were for the identification of the sequence in the region of insert in the genome, one for each strain, and the final primer pair targets the size of the insert that replaces the *pepA* gene. The PCR primer pair targeting the whole sequence of the region of the genome where recombination occurred (Primer pair AC) determines the size of the inserts and checks for duplication while the primer pairs targeting the middle of the insertion (Primer pairs AB_C and AB_K), identify the sequences in the insert and thus the identity of the gene in the insert. The analysis was only conducted for the region in the genome where the recombination occurred because the recombineering technique has been shown to be a highly precise and exact method for DNA recombination (Sawitze et al, 2007). The primer pairs used for this test and their sequences are shown in Figure 4-8 below while their sequences are shown in Table 4-5 below.



Expected Results

- i Primer pair AB_c / AB_k = 
- ii Primer pair AC = 

Legend

-  A Forward Primer
-  B_c / B_k Reverse Primer
-  C Reverse Primer

Figure 4-8: Diagram showing the primers for copy number verification, their target regions in the genome of the modified strains and expected products.

Table 4-5: Table showing sequences of the primers used for the copy number analysis

Primer	Sequence in 5' 3' direction
A	acacgaagtcacgcaacag
B _c	gggcgaagaagttgtccata
B _k	aatatcacggtagccaacg
C	tcatctcacagggcgaatgag

Primer pair AB_c is expected to amplify only the Cm⁺ strain, as the reverse primer B_c is complementary to only sequences of the chloramphenicol resistance gene, thus identifying the Cm⁺ strain. Likewise the reverse primer B_k is complementary to only the kanamycin antibiotic resistance gene and identifies the Kan⁺ strain. PCR products are expected in all cells when using primer pair AC as the primer sequences occur in all cells though different

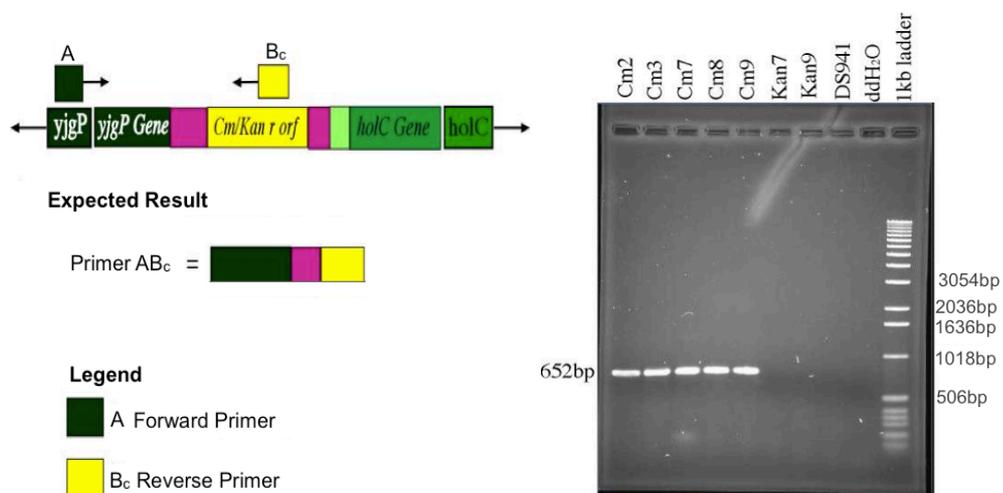
sizes of PCR products would be achieved with each cell. These different scenarios for the expected results for each cell type using the various primer pair combinations have been summarized in Table 4-6 below.

Table 4-6: Expected PCR product sizes for each primer pair combination in each cell.

Cells	Primer Pairs and expected product sizes (bp)		
	AB _C	AB _K	AC
Cm+	+ (652)	-	+ (1165)
Kan+	-	+ (830)	+ (1308)
DS941	-	-	+ (1863)

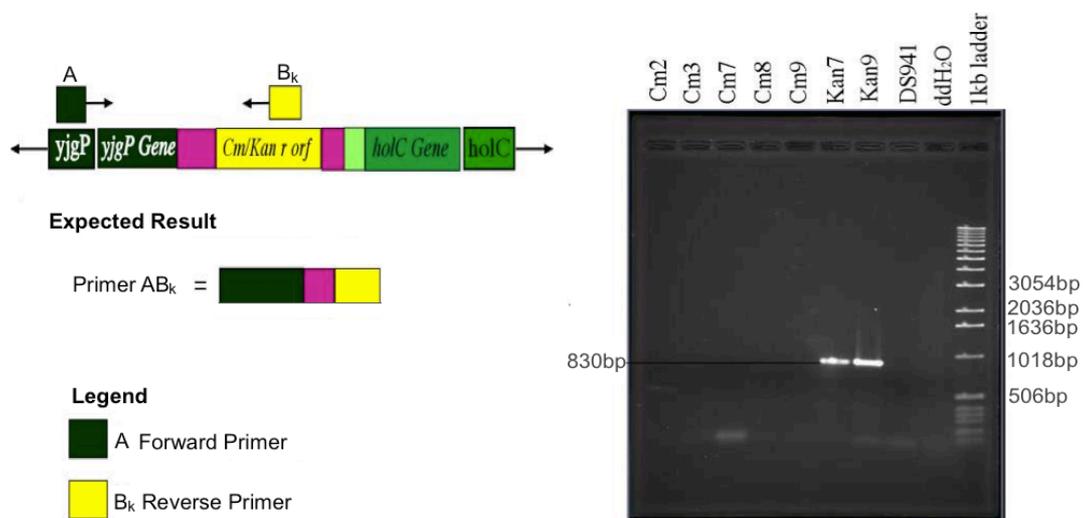
DS941 cells are used as the control experiment. Using this analysis, the exact size of the insert in each cell type can be verified and thus the number of copies in the insert. The results of the copy number experiment using all primer pairs are shown below.

1. Primer Pair ABC



PCR products were realized for the chloramphenicol resistant strains (Cm2, 3, 7, 8 & 9) as expected. The size of the products corresponds with the known size of the sequences targeted for, when measured against the 1Kb DNA ladder. This confirms that the chloramphenicol resistance gene replaces the deleted *pepA* gene in the Cm+ strains and just one copy exists.

2. Primer Pair AB_K



The size of the PCR products obtained using primer pair AB_K lies in the appropriate region of the 1KB ladder confirming the presence of only one copy of the kanamycin resistant gene in the Kan+ strains.

3. Primer Pair AC

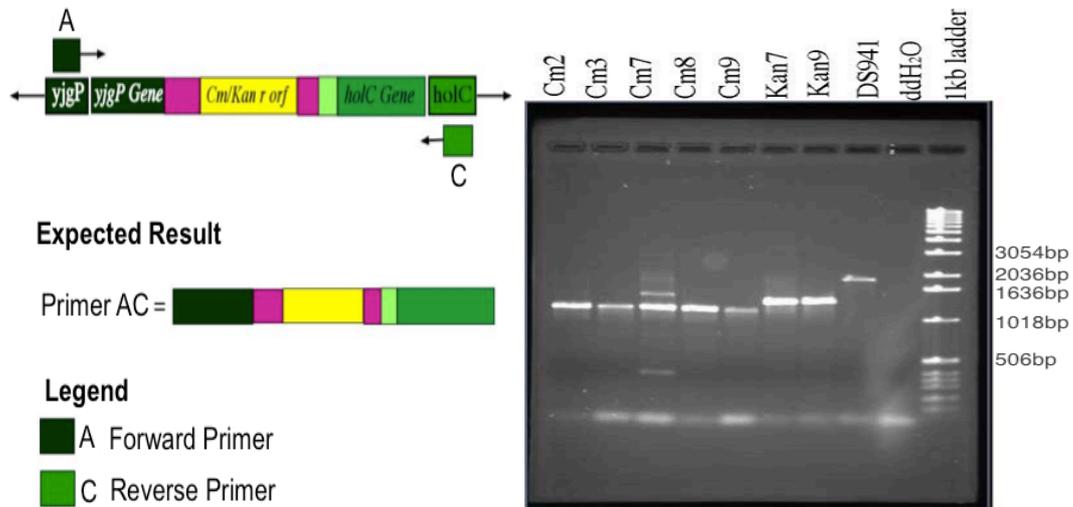


Figure 4-11: Result showing the verification of copy numbers using primer pair AC

The expected sizes of the PCR products obtained by counting the base pair sequences for the Cm⁺, Kan⁺ and DS941 strains are 1165, 1308 and 1863bp respectively. Since DNA fragments migrate through the electrophoresis gel according to their sizes, the Cm⁺ strains being the smallest in size (1165bp) are further down the lanes than the rest of the cells. The realized sizes from the experiment correspond to the expected sizes when measured against the 1KB DNA ladder. This experiment confirms that there is one copy of each antibiotic insert per cell. Thus using the antibiotic resistance gene as a marker to identify each strain makes for easy quantification in the bioreactors when running the chemostat experiments.

4.4 Real Time-Polymerase Chain Reaction

Accurately differentiating, monitoring and quantifying both strains at each time point in the bioreactors is of paramount importance in understanding the forces at work in shaping the microbial community. In this study, real time-polymerase chain reaction (RT-PCR), a culture independent technique, was used to distinguish and monitor both strains as opposed to a culture dependent technique such as plating due to the amount of samples to be analyzed.

RT-PCR is a molecular method used to amplify DNA sequences. It differs from traditional PCR in that the products of amplification during each phase of the PCR cycle can be detected, measured and visualized in real time. For traditional PCR, the products are only quantified at the end of the process (an average of 40 PCR cycles) but for RT-PCR, the detection and quantification of products happens at each PCR cycle with the help of a fluorescent molecule, which is integral to RT-PCR technique because it measures the amount of DNA after each cycle (Logan, 2009). Monitoring the reaction during the exponential phase enables the initial amount of the target DNA in the sample to be determined with great precision. An increase in fluorescent signal corresponds to an increase in the amount of PCR product generated in the exponential phase of the reaction. Plotting the fluorescence signal detected against the cycle number in which it occurs gives an amplification plot that represents the accumulation of products during the entire PCR reaction. A

sample with a large amount of target DNA would be detected earlier than that with a smaller amount of starting DNA (Logan, 2009).

DNA binding dyes and fluorescently labeled sequences of primers or probes are among the fluorescence chemistries usually used in RT-PCR (Logan, 2009). RT-PCR reactions are carried out in specialized thermal cyclers that can detect fluorescence. In this respect, RT-PCR offers a robust, highly reproducible and sensitive method to quantitatively track genes across temporal scales in complex environmental samples. It is a powerful tool for quantifying bacterial abundance under varying experimental conditions down to genus and species level. It also provides information on both the relative and absolute abundance of a species by targeting and amplifying a specific sequence of the DNA of the target species and gives an accurate determination of the number of individuals present (Zhang, 2006).

In this study, dual labeled fluorescent DNA probe sequences were used as the fluorescent chemistry (TaqMan probe), because of their increased specificity to the target sequences. TaqMan probes use a fluorophore, a fluorescent chemical compound, attached at the 5' end of a sequence and a quencher at the 3' end of the same sequence. The sequence anneals to the corresponding sequences within the target DNA template, which a specific set of forward and reverse primers amplify. The fluorophore and quencher act as a donor-acceptor FRET pair where, when the fluorophore is excited by the light source from the thermal cycler, it transmits the energy to the quencher. The quencher regulates the fluorescence signal emitted by the fluorophore by absorbing and dissipating its energy depending on their proximity. When both

are in close proximity, no fluorescence signal is emitted. Hence, fluorescence can only be detected when the fluorophore has been cleaved from its sequence by the exonuclease activity of the polymerase, which extends the forward primer during DNA replication. When the forward primer is extended on its complementary sequence of the target DNA sample. In this way, the fluorophore is away from the quencher and can fluoresce. The efficiency of the process is dependent on the overlap of the fluorescence emission of the fluorophore and quencher absorption spectra (Smith, 2009). The fluorophore used in this study was hexachlorofluoresceine (HEX) while the quencher was black hole quencher 1 (BHQ 1). This pair was chosen because black hole quenchers do not have native fluorescence and provide better resolution of the fluorescence spectrum; Hex was used as it is cheap and has a good spectral overlap with BHQ1 (Zhang, 2006). The target DNA sequences for the chloramphenicol species are from the chloramphenicol resistance gene while those for the kanamycin species are from the kanamycin resistance gene, thus genomic DNA was the template for the analysis.

The amount of fluorescent signal detected for any species above the ambient background signals (threshold frequencies) corresponds to the amount of DNA of that species in the sample. The first cycle number at which detection is recorded is called the Ct (threshold cycle number) value. An early cycle number signifies that the sample contains a high amount of the starting DNA template while a later Ct value denotes that the sample has less starting template. Thus the Ct value is the inverse of the starting amount of DNA in any given sample, which reflects the number of individuals of the species in

the sample since there is only one copy of the target gene in an individual for each species.

Samples of culture would be taken from each reactor and genomic DNA (gDNA) extracted from each. To relate the amount of gDNA to the number of cells in the sample, a standard curve is required. The standard curve for this study was constructed using pure cultures of each strain. Both strains were grown under the same conditions but in different flasks to stationary phase, gDNA was extracted and the gDNA concentration was determined by measuring the absorbance at 260nm (see page 36) while the respective copy numbers from each gDNA were calculated from Avogadro's number and type of DNA. Standard curves were then generated by preparing serial dilutions of the extracted gDNA for each strain and conducting RT-pcr on each dilution. The standard curve should give a linear correlation between DNA concentration and copy number. Using this, the number of Cm⁺ and Kan⁺ cells (copies) in the unknown samples can be determined as they should lie within the standard curve, since from the construct, one target gene (chloramphenicol or kanamycin resistant gene) is equivalent to one cell. The forward, reverse and probes sequences used in this study are shown below.

Table 4-7: Table showing the sequences of the probes, forward and reverse primers in the 5'3' direction of both strains used in the qPCR quantification analysis.

	Sequences of strains in the 5' 3' direction	
Primers	Cm⁺	Kan⁺
Forward	AGACGGTGAGCTGGTGATATG	TCTGGATTCATCGACTGTGG
Reverse	GTAGAAACTGCCGAAATCG	GATACCGTAAAGCACGAGGAAG
Probes	TCATCGCTCTGGAGTGAATACCACGA	CGTTGGCTACCCGTGATATTGCTGA

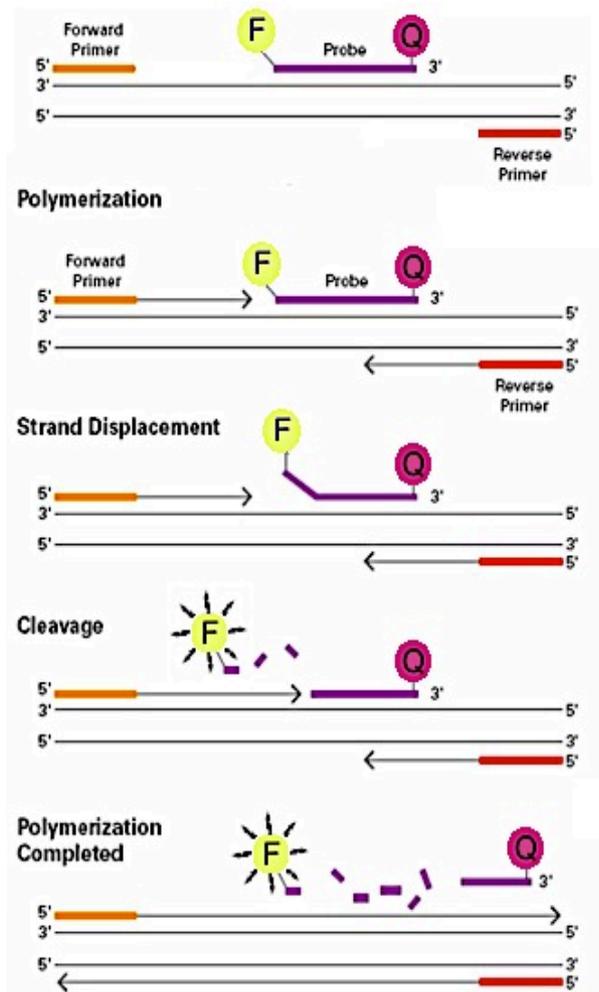


Figure 4-12: An illustration of the mode of action of TaqMan probe chemistry during DNA amplification and detection. The diagram shows the position of each oligonucleotide sequence on the target DNA sequence prior to annealing along with reporter quencher dyes. It also shows the extension of the forward and reverse primers and subsequent displacement of the fluorophore resulting in its fluorescence. The intensity of the fluorescence indicates the amount of target DNA sequences detected and amplified.

4.5 Pseudo Quantification Experiment

To establish the applicability of RT-PCR for differentiating and quantifying the bacterial species used in this study, a prior mixing and quantification test was carried out where known volumes of both strains were mixed together in different proportions and analyzed to verify if the resulting proportions obtained by RT-PCR were consistent with the volumes mixed.

Table 4-8: Table showing the ratio in which both bacterial strains were mixed.

Sample	Cm+	Kan+	Cm+	Kan+
	Volume added (µl)		Ratio	
1	0	2000	0	1
2	400	1600	0.2	0.8
3	800	1200	0.4	0.6
4	1000	1000	0.5	0.5
5	1200	800	0.6	0.4
6	1600	400	0.8	0.2
7	2000	0	1	0
8	500	1500	0.25	0.75
9	1000	500	0.67	0.33

Pure cultures of both strains were grown overnight to stationary phase and mixed in the proportions shown in the table above. The samples were spun down, the supernatants decanted and DNA from each sample was extracted from the resulting pellets using QIAamp DNA mini kit from Qiagen. RT-PCR reactions were set up using the extracted genomic DNA as template and the analysis was done in triplicate for the nine samples giving a total of twenty-seven reactions for each strain. The Cm+ strains in each sample were quantified using primers and probes specific to the chloramphenicol

resistance gene. Using fresh DNA templates from the nine samples, the analysis was repeated for the kanamycin strain but with primers and probes set specific for the kanamycin resistant gene. Standards were run alongside sample unknowns for both cases. The results of the RT-PCR reactions are shown below.

Table 4-9: Table showing result of RT-PCR for both strains.

RT-PCR Efficiency (%)			93.1	94.6		
	Volume mixed (ul)		Resulting Mean Copy Number		Resulting Abundance Ratio	
Sample	Cm+	Kan+	Cm+	Kan+	Cm+	Kan+
1	0 (0)	2000 (1)	7.48E-01	2.40E+05	0.000	1.000
2	400 (0.2)	1600 (0.8)	4.83E+04	1.85E+05	0.207	0.793
3	800 (0.4)	1200 (0.6)	1.20E+05	2.02E+05	0.373	0.627
4	1000 (0.5)	1000 (0.5)	1.39E+05	1.56E+05	0.471	0.529
5	1200 (0.6)	800 (0.4)	1.69E+05	1.31E+05	0.563	0.437
6	1600 (0.8)	400 (0.2)	2.01E+05	5.77E+04	0.777	0.223
7	2000 (1.0)	0 (0.0)	2.59E+05	4.40E-01	1.000	0.000
8	500 (0.25)	1500 (0.75)	7.04E+04	2.18E+05	0.244	0.756
9	1000(0.67)	500 (0.33)	2.07E+05	1.13E+05	0.647	0.353

From the table above, the RT-PCR efficiency of the reactions carried out for chloramphenicol based reactions is 93% and 95% for those of kanamycin. The mean of the copy number of the triplicate runs for each sample was calculated and from this the abundance ratio of each strain in each sample was determined. The abundance ratio of the copy number of both strains in each sample shows a tight fit to the ratio of the volume in which they were mixed. Also, graphs showing the goodness of fit between the ratio of the volume of both strains added in each sample and the ratio of the copy number recovered for both strains are shown below.

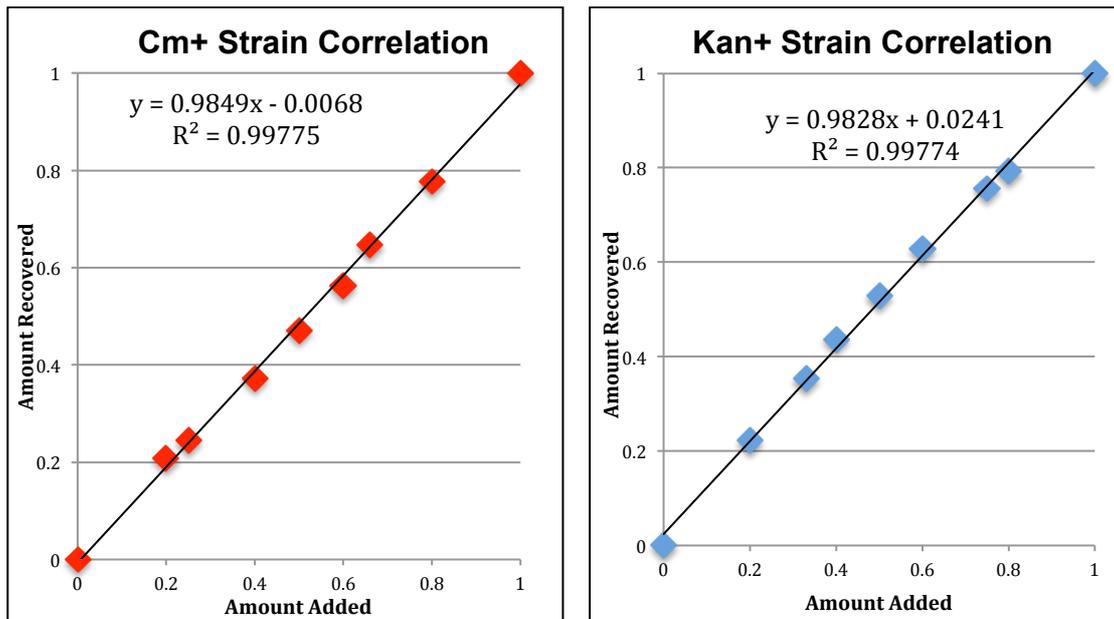


Figure 4-13: Result showing goodness of fit for both strains in terms of ratio added and ratio observed.

The goodness of fit for both strains shows an R² of 1, implying the ratio of copy number of both strains recovered from the RT-PCR reactions are exactly the same ratio in which they were added in the samples. This validates RT-PCR as an accurate and reliable method for quantifying both bacterial strains in the mixed bioreactors during the chemostat experiments.

5 Characterization of Strains

In order to rationally design the chemostat experiments, it is imperative that the growth kinetics of the modified strains of *E.coli* are determined. This chapter describes how the strains were characterized in terms of their growth rates and yields both in the presence and absence of antibiotics. This knowledge allowed us to parameterize a mathematical model that solve the classical deterministic models of a chemostat and ascertain the most appropriate operating parameters for the competition experiments described in subsequent chapters. Whilst the aim of the thesis is ultimately to quantify the demographic stochasticity manifest in population dynamics, the rationale for designing the experiments on the basis of a deterministic model is that the long-term mean behavior of the stochastic dynamics should converge on that predicted by the deterministic model.

5.1 Generation of process parameters in the absence of antibiotics

To ensure that the strains had identical birth rates, one of the assumptions of the neutral theory in its original form, a batch growth rate analysis was conducted without antibiotics. The growth rate was investigated for the modified strains and the wild type DS941 cell independently, to assess the similarity of their growth rates.

5.1.1 Batch Growth Rate

Nutrients and physiochemical conditions such as temperature, pH and aeration are critical factors that influence the growth rate of *E.coli*. Depending on the nature of the research, various combinations of these factors affect the optimal doubling time of *E.coli*. When all factors are present in sufficient amounts, the growth rate of *E.coli* is estimated to be approximately 20 minutes (Archunan, 2004). In this study, all critical factors responsible for *E.coli* growth (temperature, pH and aeration) were present in adequate concentrations except the growth nutrients, which were allowed to vary. This was because the analysis of the solution of standard chemostat equations is easiest when the bacterial growth rate is a function of a single growth-limiting nutrient. Thus a single growth-limiting nutrient culture media was used that was sufficiently rich to ensure a high population density in the community. Minimal media was not used in this study because it does not ensure high population density in the community as a rich media would.

Tryptone broth³ was used as the culture media in this study as opposed to Luria-Bertani (LB) broth⁴, a more commonly used rich bacterial culture medium, because although both broths have complex undefined components, tryptone broth comprises just two components, tryptone and sodium chloride (Miller; 1995). Tryptone, a casein digest that is comprised of many proteins, serves as the nitrogen & carbon source and can be regarded as the growth-limiting nutrient while sodium chloride is present for osmotic balance. LB broth

³ 1L basis of Tryptone broth is made of 10g tryptone and 5g sodium chloride.

⁴ 1L basis of Luria-Bertani (LB) broth is made of 10g tryptone, 5g yeast extract, and 5g sodium chloride. (LB Miller's recipe).

contains yeast extract, a complex blend of amino acids, vitamins, some trace elements, nucleotides and some fermentable sugars, in addition to tryptone and sodium chloride (Sezonov, 2007). Though tryptone, a single nutrient made up of complex proteins, is the growth-limiting nutrient, it is still a complex nutrient and its molecular and atomic compositions are not easy to analyze (Bronzino, 2000).

Three overnight cultures were setup for the three bacterial strains from their respective glycerol stock stored at -70°C by streaking the stocks on LB agar plates containing $10\mu\text{g/ml}$ chloramphenicol, $20\mu\text{g/ml}$ kanamycin, and $50\mu\text{g/ml}$ streptomycin (selection for Cm^+ , Kan + and DS941 strains respectively) and incubating at 37°C for 20 hours for colonies to form. A single colony was picked from each plate and grown overnight for 16 hours in a culture tube containing 10ml LB broth at 37°C shaking at 200rpm for each strain. $200\mu\text{l}$ of the overnight cultures were used to inoculate 50ml of tryptone broth in a 250ml conical flask, which was preheated at 37°C , 200rpm.

Samples from the growing cultures were taken every 30mins for 10 hours and their absorbance of light at 600nm wavelength recorded with a spectrophotometer. According to Beer Lambert's law, the absorbance of an incident ray of light by a solution is proportional to the concentration of the solution and path length of light (Sharma et al, 2009). When visible light is passed through a suspension of cells, the amount of light scattered as a result is called the optical density (OD). The OD is a direct measure of the concentration of bacterial cells present in a sample. The results of the growth curve for the three strains on three different days are presented below.

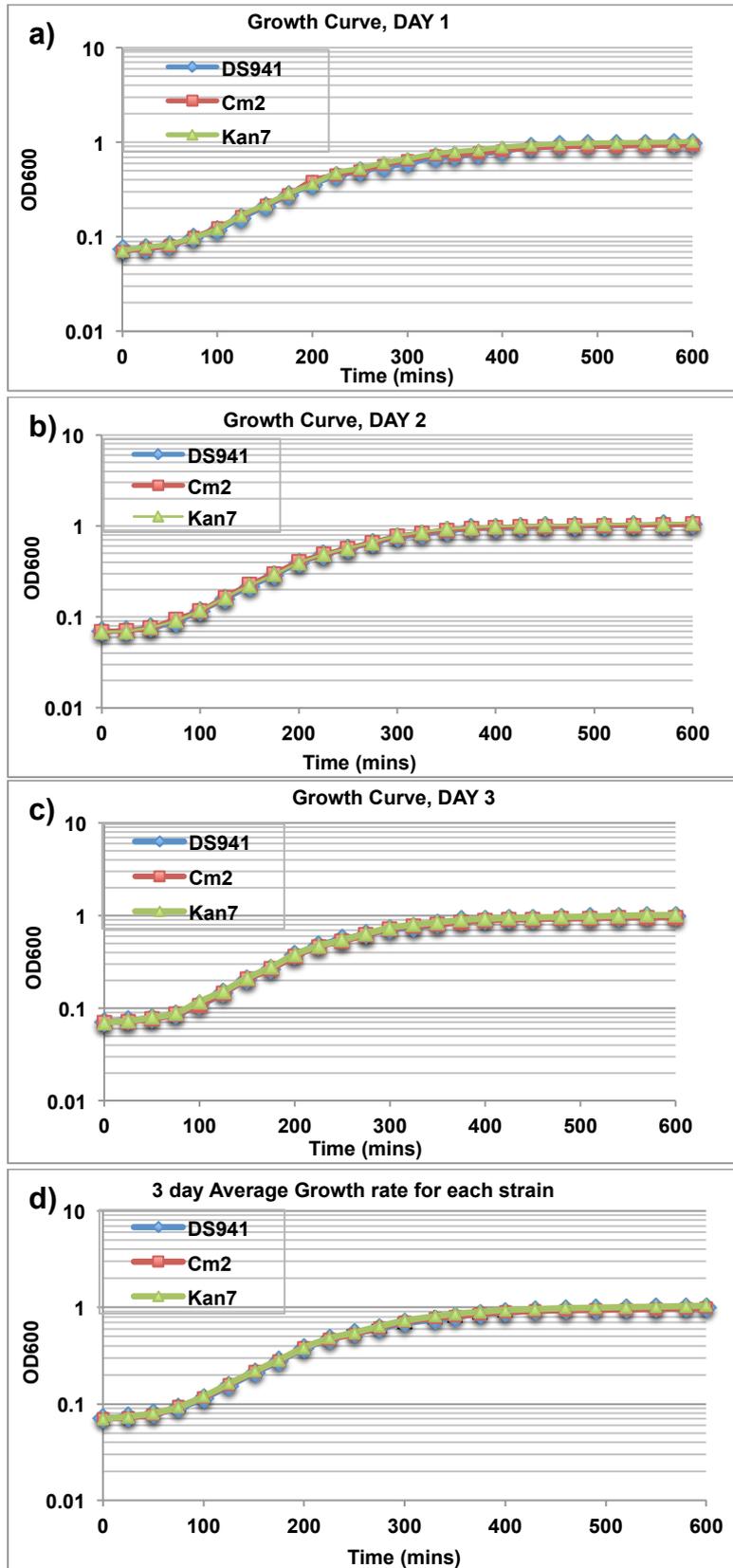


Figure 5-1: Graphs a, b and c show batch growth curves of the wild type and other strains on three different days and graph d, an average of the three curves plotted together with standard deviation as error bars.

Table 5-1: Result of growth rates and doubling times of the three strains on three different days.

	Day 1			Day 2			Day 3		
Cells	DS941	Cm+	Kan+	DS941	Cm+	Kan+	DS941	Cm+	Kan+
μ (hr ⁻¹)	0.6	0.61	0.62	0.67	0.67	0.67	0.73	0.70	0.68
td (hr)	1.16	1.13	1.12	1.03	1.04	1.04	0.96	0.99	1.01

The absorbance at 600nm is plotted against time on a semi-logarithmic graph (Figure 5-1). Each growth curve follows the standard bacterial growth curve: an initial lag phase, an exponential growth phase and a stationary phase. By preheating the media at the optimal temperature for *E.coli*, 37°C, a short lag phase was attained. The growth rate (μ) of the three strains was calculated from the exponential phase of the individual growth curves to be approximately 0.6hr⁻¹ while the doubling time (t_d), the quotient of the natural logarithm of 2 and μ , is roughly 1 hour as shown in Table 5-1 above. Such a slow doubling time could be advantageous in this study in that it permits better monitoring of the dynamics that may exist between the strains and thus the demographic stochasticity in the system. The growth rates of the strains are very close from day to day and on any given day the growth rates of the three strains are very similar. The error bars (2 times the standard deviation) are very small. Thus, there is little or no difference between the growth rates for the three days implying that by using recombineering, almost identical strains with reproducible growth rates can be created.

5.1.2 Monod's Kinetics

For the continuous growth of bacteria in the chemostat, a suite of process equations, which rely on knowledge of the growth rate at different substrate concentrations, have been shown to characterize the mean behavior of the system. To approximate the dynamics without demographic stochasticity, and to rationally design the system, it is imperative to characterize the mean specific growth rate. Monod investigated the effect of a growth-limiting nutrient on bacteria growth in a chemostat and postulated the relationship,

$$\mu = \frac{\mu_{max} S}{K_s + S} \quad (5:1)$$

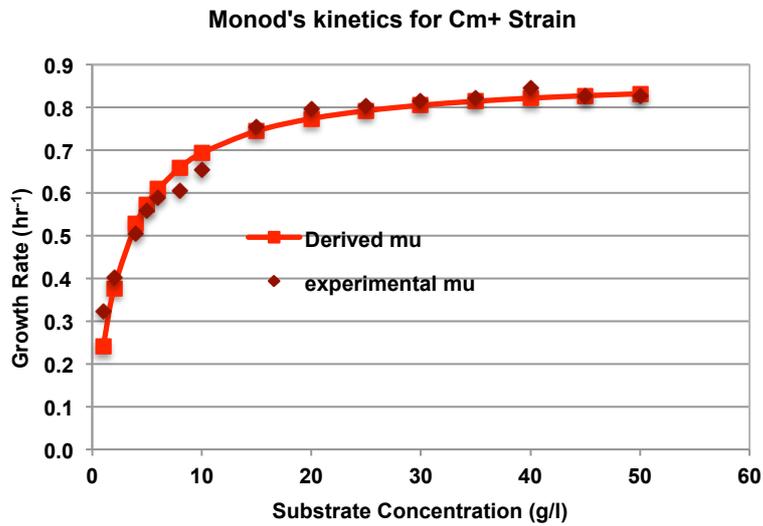
where μ is the specific growth rate, μ_{max} the maximum specific growth rate, S the substrate concentration and K_s the half-saturation constant for the growth limiting nutrient. This form of relationship is referred to as Monod Kinetics. From equation (2:1), an organism's specific growth rate increases with the concentration of the growth-limiting nutrient then approaches an asymptote at μ_{max} , beyond which additional substrate has no effect on the growth rate. The parameter values vary between different organisms and growth-limiting nutrients. K_s is the substrate concentration at half μ_{max} and it determines the affinity of an organism to the substrate. The higher an organism's K_s value in a particular substrate, the higher substrate concentration required for μ to approach μ_{max} (Grady et al, 2011).

To parameterize the Monod Kinetics, it is necessary to generate growth curves at different concentrations of substrate and estimate the growth rate for the strains. As above, tryptone broth was used at varying concentrations of tryptone while the concentration of sodium chloride remained the same in all cases. The following tryptone concentration ratios were used: 0, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5 and 5. A graph of the optical density (OD) at 600nm was plotted against sampling time for each substrate concentration and the specific growth rate calculated. Subsequently, a graph of the different specific growth rates was plotted against their corresponding substrate concentration to replicate Monod's growth kinetics model for limiting substrate concentration. The graphs of specific growth rate against substrate concentration for both strains are shown in Figure 5-2.

The result of Figure 5-2 shows that both strains appear to follow Monod Kinetics for microbes when a single nutrient is limiting, whereby the growth rate of an organism increases with increasing substrate concentration, until a maximum growth rate is reached where further increase in substrate concentration has no influence on the growth rate. Thus Monod's kinetics is a combination of first (substrate dependent) and zero (substrate independent) order growth kinetics, providing a platform of growth kinetics one can choose to operate in with regards to bacterial batch growth. The value of K_s determines the region of operation: first or zero order growth kinetics. Thus the correct estimation of these parameters from experiment is vital to the design of the continuous system. To determine the maximum specific growth rate of both strains, the Monod kinetics equation was linearized to

$$\frac{1}{\mu} = \frac{1}{S} \frac{K_s}{\mu_{max}} + \frac{1}{\mu_{max}} \quad (5:2)$$

a)



b)

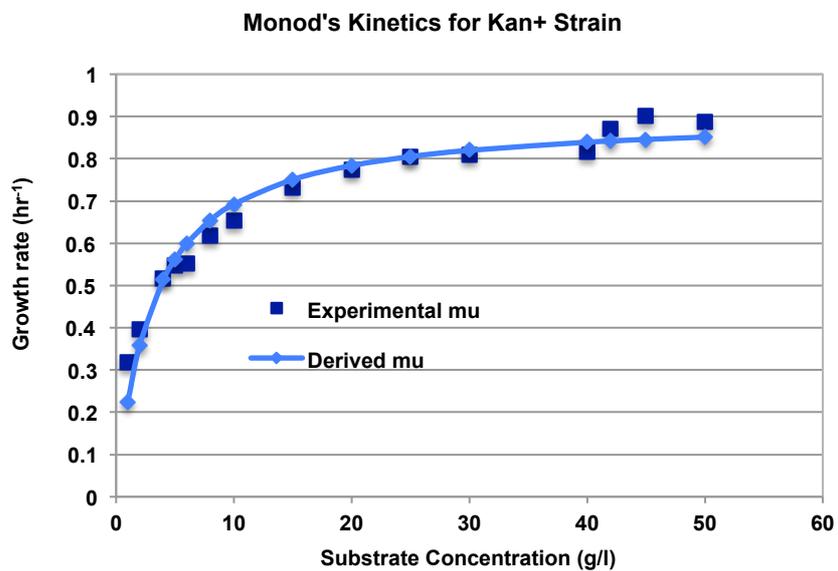


Figure 5-2: Monod Kinetics result of Graph showing actual experimental data points and data points derived by substituting growth parameters from the regression analysis a) – the chloramphenicol resistant strain and b) – the kanamycin resistant strain in the absence of antibiotics.

Thus plotting a graph of $1/\mu$ against $1/S$ yields a straight line, which for enzyme kinetics is called a Lineweaver-Burk plot. Following this linearization, the inverse of the y-intercept gives μ_{max} and from the slope of the graph, K_s can be determined, as μ_{max} is already known. The inverse of the x-intercept further confirms the value of K_s . The slope and intercept of this graph were estimated by linear regression. These estimated growth parameters were also further validated using nonlinear regression. Figure 5-2 shows that the experimental and modeled values of the specific growth rate (both linear and nonlinear) are a good fit and gives a value of 0.89hr^{-1} and 0.90hr^{-1} as the μ_{max} and as 2.9g/l and 3.0g/l the K_s for Cm^+ and Kan^+ strains respectively. The Monod Kinetics constants for both strains are summarized in Table 5-2.

Since the substrate concentration (10g/l ; from tryptone broth) to be used in this experiment is greater than the K_s of both strains ($\sim 3\text{g/l}$), the growth dynamics of community would be of first order growth kinetics. Besides, Monod's kinetics has shown to have limited applicability at such low substrate concentration (Bekins et al, 1998). These parameters along with the growth yield provide the basis for the design of the chemostat in terms of process flow rate, culture volume and desired process dilution rate.

5.1.3 Growth Yield Coefficient

In the chemostat growth chamber, depletion and replenishment of nutrients occur concurrently until a steady state condition is reached. Substrate depletion, which occurs as a result of consumption by microbes, can be correlated to microbial growth via a parameter called the growth yield (Y). The growth yield is the quantity of biomass synthesized per unit of substrate consumed, reflecting the conversion efficiency of substrate to biomass by the microbes (Lee, 2006). The yield, like μ_{max} and K_s varies between species, or even between the strains, and is a function of the substrate. Knowledge of the values of these growth parameters - μ_{max} , K_s , and Y – is the basis for designing bioreactors, optimization of culture process and determination of control strategy (Dunn, 2003).

Empirical investigation of the growth yield for both strains in different substrate concentrations continues from the set up used for the Monod kinetics. After the Monod kinetics experiment, 10ml was taken from each of the culture solutions, which corresponded to the different substrate concentrations listed above, and filtered onto a 0.2 μ m cellulose nitrate filter paper from Whatmann, for which the dry weight had previously been measured. The cells that were trapped on the filter paper were dried in an oven at 105°C and weighed continuously until there was no further change in weight for each filter paper and hence the dry weight of bacteria and filter paper was calculated. The actual weight of the cells (biomass) was determined by subtracting the initial weight of the filter paper from the final weight of the filter paper with biomass.

Usually, an analytic assay is conducted to determine the substrate concentration before the start of the experiment and at its end. Thus the amount of substrate consumed by the organism can be calculated. This is subsequently plotted against the biomass trapped on the filter paper, whose slope then gives the growth yield. But in this study, as mentioned earlier, tryptone, our growth-limiting nutrient is made up of complex proteins. To date, the analysis available for tryptone only verifies its presence or absence and not its concentration, thus it is difficult to determine the concentration of tryptone in the growth media at any time (MacWilliams, 2009). This is not an uncommon occurrence when a rich media (usually contains complex nutrients) is used in studies involving cell growth. Rich media is preferred when rapid growth rates are desired (Bronzino, 2000). Ultimately, the interest here is in the stochasticity in the dynamics rather than the depletion of resources, thus the yield is of secondary importance; nonetheless getting a 'ball park' estimate helps us to determine the substrate concentrations that should be used in the experiments to achieve the desired growth rates.

To get around this, a test growth yield analysis for both strains using glucose as the growth-limiting nutrient was conducted. A Glucose Assay Kit (GO) from Sigma Aldrich was used to quantify the actual amount of glucose consumed by the strains. The analysis was conducted as follows

$$S_s - S_E = YB \quad (5:3)$$

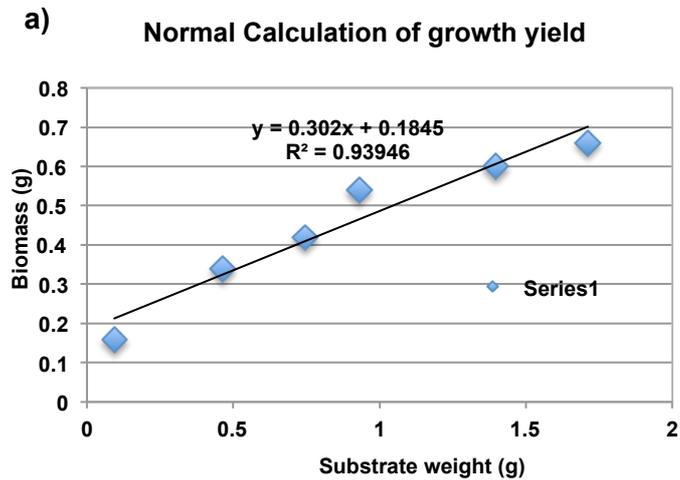
Where S_s is the weight of substrate at the start, S_E , the weight at the end, Y , the yield and B , the biomass produced. Suppose that only a small proportion, γ , of the initial substrate is used. Then

$$S_E = \gamma S_s \quad (5:4)$$

and so substituting equation (5:4) into (5:3), gives

$$S_s(1 - \gamma) = YB \quad (5:5)$$

Then plotting S_s against B will give an estimate of Y if γ is small. It is shown in Figure 5-3 that a rough value of Y can be obtained for glucose where S_E can be measured.



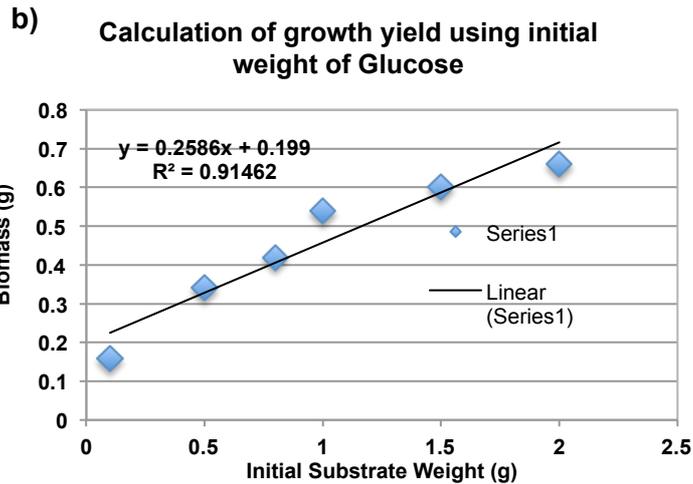


Figure 5-3: Estimates of growth yield of the strains when glucose is the growth-limiting substrate: a) when calculated normally, b) when calculated from the initial weight of substrate

For tryptone, it is expected that γ would be small and hence a crude estimate of Y can be obtained using only the weight of substrate at the start. Therefore, a straight-line relationship between the weight of substrate at the start and biomass produced would give a good estimate of the yield. Using the initial weight of tryptone at the start for our growth yield calculations is inconsequential to the stochastic modeling, which is based on growth rates and not knowledge of the substrate consumed. A graph of the dry cell weight was plotted against the corresponding substrate concentrations, for which the slope of the graph gives the growth yield coefficient of the strains. The graphs of the growth yields for both strains are shown below.

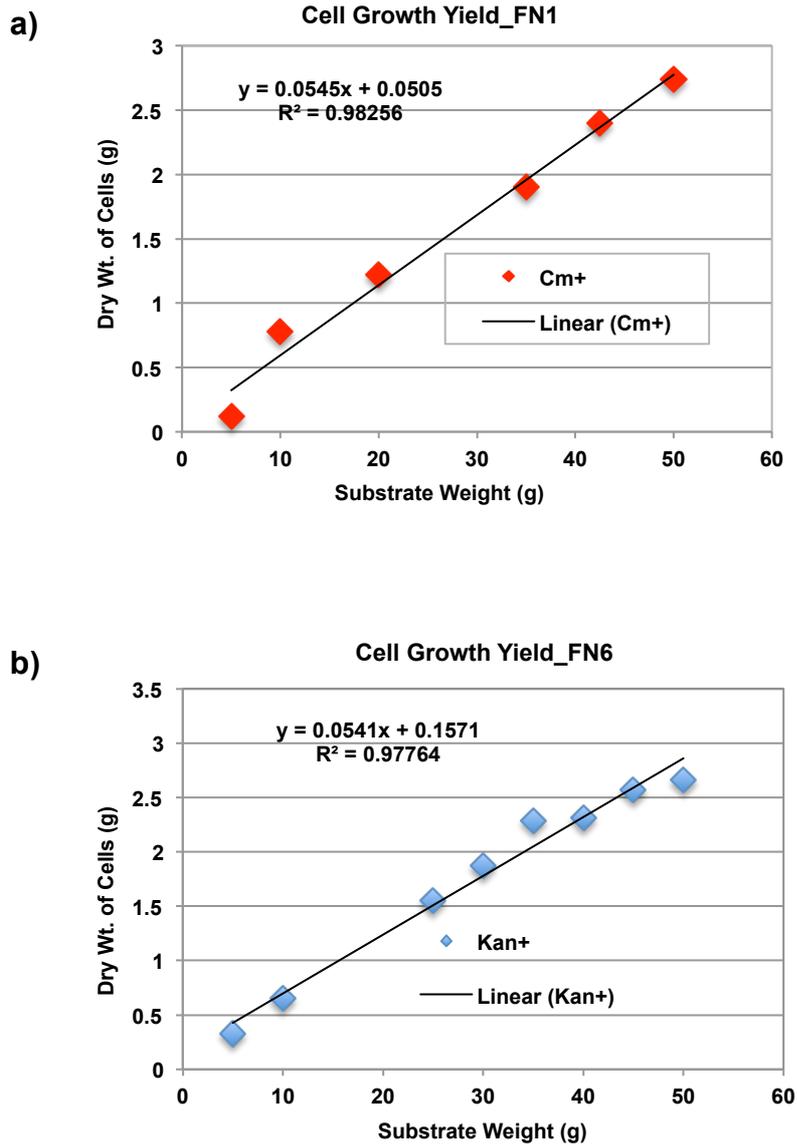


Figure 5-4: Result of growth yield of a) the chloramphenicol resistant strain and b) the kanamycin resistant strain in the absence of antibiotics.

The growth yield is approximately 0.05 for both strains in all replicas, further confirming the similarity between the strains. However, this value is low considering that typical bacterial growth yield in carbon sources such as glucose and lactose is about 0.1 – 0.6 (Heijnen, 1981). This implies that the production of cells per unit tryptone consumed is low. Thus in order to

maintain a high population density in a continuous culture, there will be a significant depletion of resources. This could potentially have a knock on effect on subsequent experiments in this thesis where two pure cultured strains are fed into chemostat; the growth media emanating from the outlet tubes of the pure cultures might not be sufficient to sustain growth of the mixed cultures. This low growth yield could be attributed to using nitrogenous base as the limiting nutrient, serving as the sole energy source, rather than using it in conjunction with a carbon source, which *E.coli* rapidly metabolizes. When nitrogenous source is the growth-limiting nutrient, the growth yield is low, as energy and electron imbalance arise as a result of nitrogen reduction during nutrient assimilation (Xiao, 2005). Nitrogenous-based growth-limiting nutrient was the media of choice in this study because of its high growth rates despite having low growth yields compared to carbon sources. Achieving such high growth rates with carbon sources implies an increase in the concentration of the carbon source, which causes a considerable build up of acetate when *E.coli* is cultured, a toxic biosynthetic product that lowers recombinant protein production (Han, 2002). In this thesis, the interest is to monitor and determine the interplay between stochastic and deterministic forces in shaping microbial communities and quicker growth will manifest more rapid dynamics, which will be easier to observe and hence the selection of tryptone broth as the growth media. The growth parameters measured for both strains in the absence of antibiotics are shown below. Thus the two strains have very similar growth characteristics. Given that the only difference is in the insertion of different antibiotic resistance genes in the genomes of the strains there should indeed be no difference in their growth characteristics.

Table 5-2: Summary of process parameters in the absence of antibiotics generated via Monod Kinetics and growth yield experiments for both strains.

Process Parameters (in the absence of antibiotics)	Cm+			Kan+		
	Value	LCL	UCL	Value	LCL	UCL
μ_{\max} (hr ⁻¹)	0.8571	0.817	0.9006	0.8675	0.8193	0.9216
K_S (g/l)	2.4587	2.1774	2.7400	2.6000	2.260	2.9401
Yield	0.053	0.0428	0.0651	0.0515	0.0403	0.0627

5.2 Generation of process parameters in the presence of antibiotics

The parameters measured in the previous section were for growth of the strains in the absence of antibiotics and implied that both strains had equal opportunity and access to nutrients in the environment they were cultivated. The strains should thus have almost identical specific growth rates and hence can be considered to be functionally neutral and used to test neutral theory.

For many systems of practical importance, it would be beneficial to predict the microbial populations dynamics where a species has an advantage over other species present in the environment (Stroo, 2013). For example, in bioaugmentation a species, natural or engineered, is introduced into an environment to improve a process because it is better performer of a desired function than the resident species. For instance, the widespread use of *Dehalococcoides spp.* to completely degrade chlorinated solvents, a prominent groundwater contaminant from industrial sites to ethene (Hendrickson, 2002). Due to the specialized function this organism performs and due to it being an immigrant species, not in its natural habitat, it remains

unclear whether it can ward off the competition from resident species and sustain itself in this new environment (Stroo, 2013).

Using the stochastic model and the experimental chemostat, one can predict and then test for non-neutral dynamics. In the models, a selective advantage is represented by increasing the probability of birth of a species relative to all others in the system and conversely a disadvantaged organism has a decreased probability of birth. In this experimental system, a selective advantage or disadvantage is conveyed by introducing a low dose of antibiotics; either chloramphenicol or kanamycin. The organism not resistant to chloramphenicol will be disadvantaged in the presence of chloramphenicol and whilst the one that is resistant will continue to grow unfettered and vice versa. Chloramphenicol is a bacteriostatic antibiotic, effective against a large number of gram-negative and gram-positive aerobic bacteria, which can be bactericidal at higher concentrations. Its effectiveness is attributed to inhibiting protein synthesis. Protein is a molecule responsible for regulating metabolism, reproduction, growth and repair. It does so by binding to a receptor site on the 50S subunit of the 70S bacterial ribosome thereby preventing the binding of aminoacyl-tRNA to the active site of peptidyl transferase, an enzyme that catalyzes peptide bond formation (Scholar, 2000). The mechanism of resistance against chloramphenicol is two fold, one mechanism involves altering the 50S subunit of the 70S ribosome, thereby decreasing the binding affinity of the drug to its receptor site and the other, a high resistance mechanism involves the action of the cat gene. The cat gene codes for the enzyme chloramphenicol acetyltransferase and causes the acetylation of the

drug by introducing acetyl groups from acetyl-S-coenzyme A, which covalently links to hydroxyl groups of the chloramphenicol structure forming 1,3-diacetylchloramphenicol. In this way, the structure of the drug is transformed thereby rendering it inactive (Roettig, 2013).

Kanamycin, a member of the group of antibiotics called aminoglycosides, is a bactericidal antibiotics active against aerobic, gram-positive and gram-negative bacteria. It is widely used against mycobacteria tuberculosis. Generally, it irreversibly binds to the 30S ribosomal subunit and 16S rRNA at the aminoacyl site thereby interfering with the translation initiation complex which causes misreading of the codons along the mRNA by the tRNA carrying the anticodons which, then codes for the wrong amino acids. Thus abnormal proteins are synthesized from the genetic information contained on the bacterial DNA and it is unable to make the exact proteins necessary for its growth and maintenance (Clark, 2013). Also, the interference of the tRNA binding on the aminoacyl site by kanamycin prevents the movement of tRNA from the aminoacyl to the peptidyl site, hence inhibiting the elongation of the polypeptide chain (Clark, 2013). This implies that folding of protein cannot take place, as the polypeptide chain is not linking the amino acids produced together, thus protein is not synthesized. Similar to chloramphenicol, bacteria can become resistant to kanamycin by decreasing permeability or by the activity of aminoglycoside phosphotransferase, an aminoglycoside-modifying enzyme. This inactivates the kanamycin molecule by transferring the gamma phosphate group from ATP to the 3' hydroxyl group of the kanamycin forming kanamycin (3')-phosphate, thereby modifying the molecule (Wright, 1999).

The selective advantage imposed here does not completely exclude the other strain, but rather provides an environment where both strains can grow. However, the growth rate of one strain will be slightly suppressed due to the concentration of antibiotics present in the nutrient media. This requires fine-tuning the concentration of both antibiotics to find a concentration that permits the growth of both strains, while favouring one species.

Prior to these experiments, knowledge of the concentration of antibiotics, which would inhibit growth but not stop it altogether, and hence convey a quantifiable disadvantage on the probability of the birth of the strains, was not available. Thus a preliminary set of growth experiments were conducted to determine thresholds in the concentration where significant effects on growth were apparent. In these initial experiments, detailed time dependent growth curves were not monitored. Instead, OD measurements of both strains growing in media containing different concentrations of antibiotics were recorded after 360 minutes, which lies in the stationary phase where growth has ceased. These are shown in the diagram below. Concentrations above $0.2\mu\text{g/ml}$ for chloramphenicol and $3\mu\text{g/ml}$ kanamycin seem to have a significant effect on growth in the strains without halting it. Thus concentrations of antibiotics above these thresholds were considered in more detail to determine the growth parameters from closely monitored growth curves.

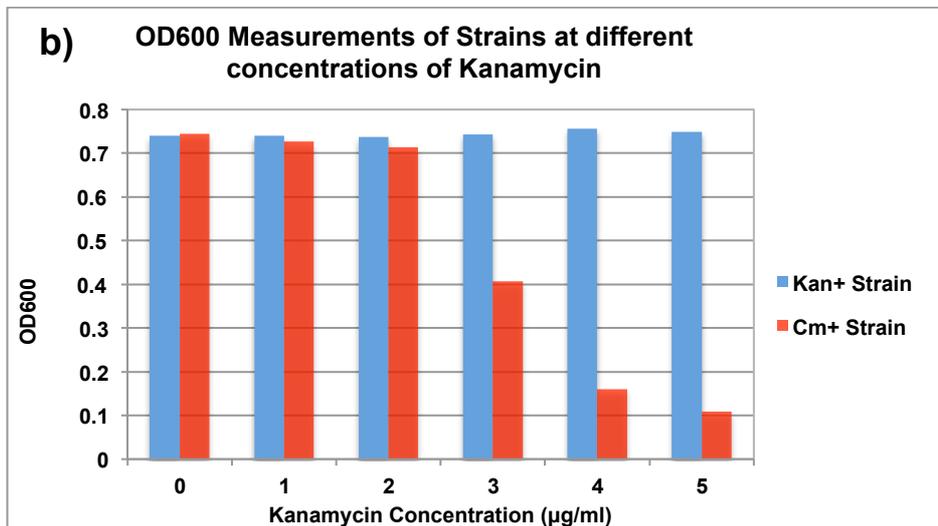
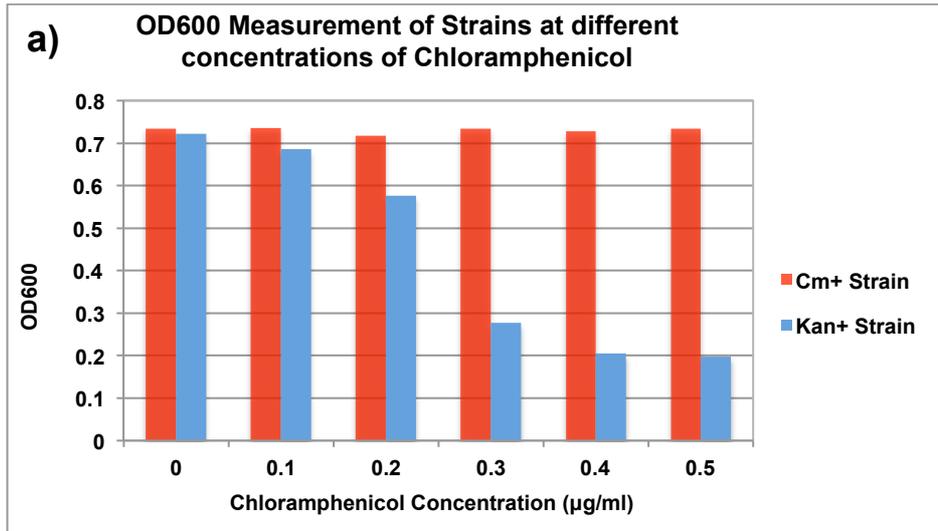


Figure 5-5: Preliminary growth assessment showing OD measurements at a time point of both strains in different concentrations of a) chloramphenicol and b) kanamycin.

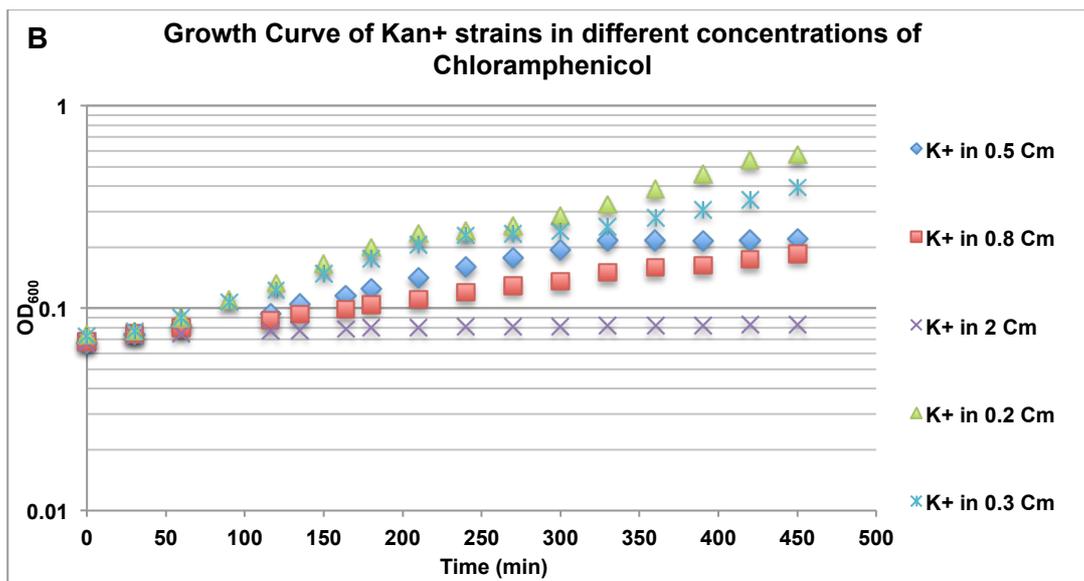
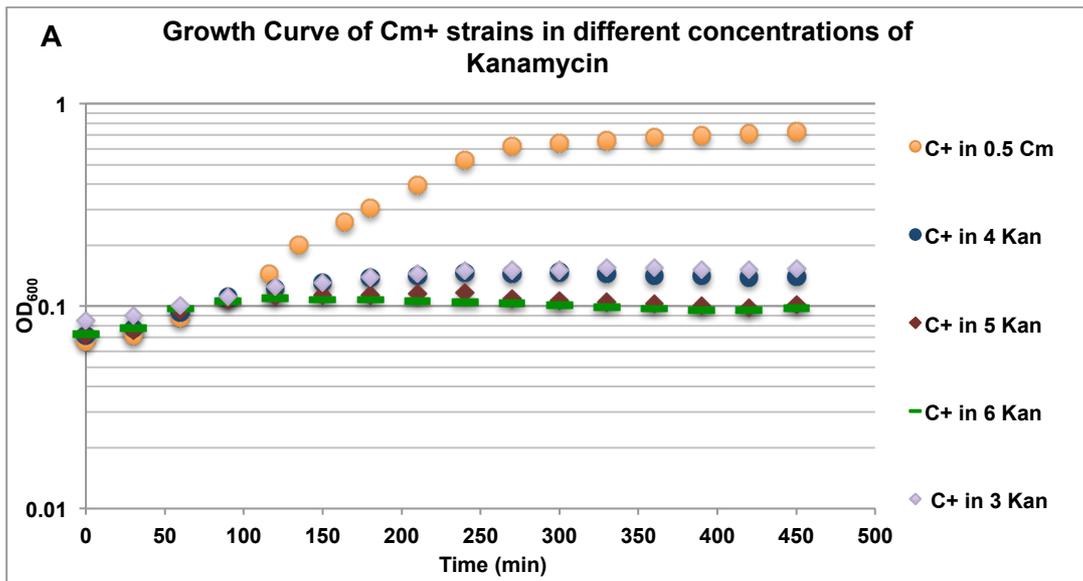


Figure 5-6: Detailed growth curve of both strains in different concentrations of antibiotics predetermined from the preliminary test. A) Cm+ strains in different concentrations of Kanamycin and B) Kan+ strains in different concentrations of chloramphenicol.

The detailed growth curves for antibiotic concentrations above the newly determined thresholds (0.2 μ g/ml chloramphenicol and 3 μ g/ml kanamycin) are shown in Figures 5-6a and b. Each of these curves represents the mean OD from triplicate experiments for each of the different antibiotic concentrations. In all cases, the K⁺ strain showed two phases of growth in media containing 0.2 and 0.3 μ g/ml chloramphenicol; between 3-4 hours it seems to commence stationary phase but then begins to grow exponentially again. This double growth could be attributed to the low concentration of chloramphenicol in the growth media, which wears out with time by getting mixed with by-products that accumulate as a result of growth. This could be problematic in the bioreactors when mixed cultures exist because though there is a continuous flow of antibiotic, given the mechanism of resistance of Cm⁺ strains in chloramphenicol and this tendency of Kan⁺ strains to exhibit double growth, the desired selective advantage may not be observed. This trend of double growth was neither observed in 0.5 μ g/ml chloramphenicol nor 3 μ g/ml kanamycin but the growth rate in 0.5 μ g/ml chloramphenicol was higher and thus was selected as the antibiotic and the concentration for the experiments. As before, the growth characteristics for both strains in the presence of a selective advantage were parameterized in terms of growth rate, maximum specific growth rate, half saturation constant and growth yield and the results are shown below.

Table 5-3: Summary of process parameters in the presence of antibiotics generated via Monod Kinetics and growth yield experiments for both strains.

Process Parameters (in the presence of 0.5µg/ml chloramphenicol)	Cm+			Kan+		
	Value	LCL	UCL	Value	LCL	UCL
μ_{max} (hr ⁻¹)	0.8571	0.817	0.9006	0.4156	0.3787	0.4604
K_s (g/l)	2.4587	2.1774	2.7400	1.0106	0.4468	1.5743
Yield	0.0403	0.0352	0.0454	0.0124	0.0082	0.0166

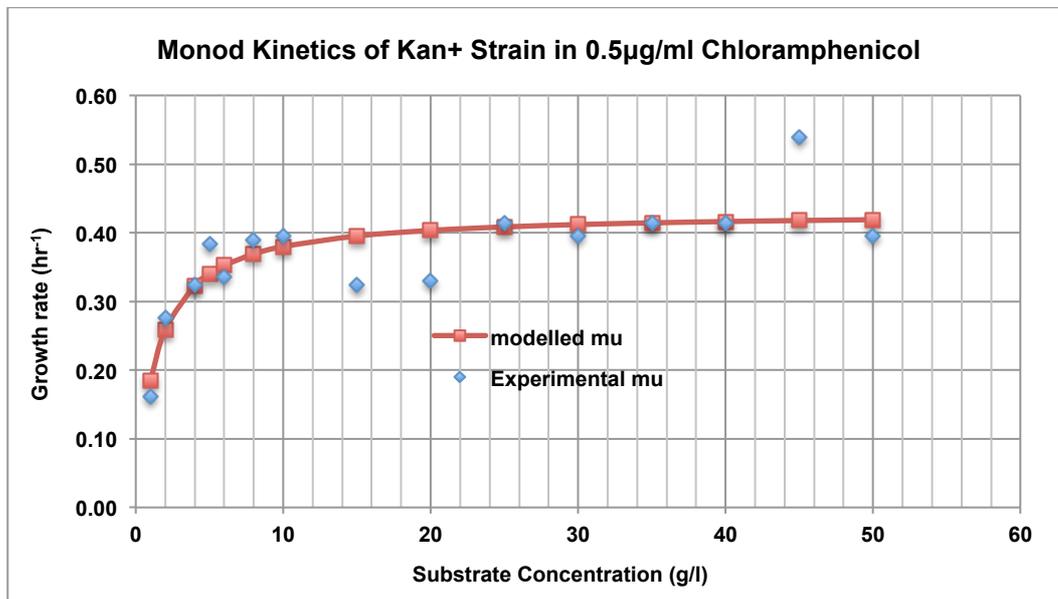


Figure 5-7: Monod Kinetics result of Kan+ strain in 0.5µg/ml Chloramphenicol

Using the same tryptone concentration of 10g/l as in the previous characterization experiment implies that the chemostat experiments involving selective advantage should be operating close to the maximum specific growth rate of Kan+ strain depending on the preset conditions in the chemostat.

6 Chemostat Experiments

In this chapter, three experimental setups for the continuous growth of the strains in the chemostat under the different conditions are presented. These are informed by developing a standard deterministic chemostat model to simulate the dynamics of the strains. The experimental results are then described and the dynamics are compared with those simulated by the deterministic model.

6.1 Neutral Dynamics with Immigration

In this section, an experiment where one would expect to see purely neutral dynamics is devised. Whilst there is evidence of stationary distribution of microbial taxa abundance fitting a neutral model in different environments (Sloan, 2006), there is no definitive evidence of neutral dynamics in microbial community assembly; though circumstantial evidence of neutral dynamics in studies with wastewater has been reported (Ofiteru *et al.*, 2010). Nee proposes that neutral dynamics may occur in microbial communities but they will be too slow to be observed in any experimental setup (Nee, 2005). Here two species having identical growth characteristics when not exposed to antibiotics have been created. Thus if the dynamics of this pair of organisms is indeed neutral and the speed is more akin to that observed by Ofiteru *et al.* (Ofiteru *et al.*, 2010) than suggested by the theoretical consideration of Nee (Nee; 2005) then one should be able to parameterize a neutral community assembly model.

This experiment is approached by first carrying out mathematical simulations using a deterministic model to inform the design characteristics and process parameters for the experimental setup. The actual experimental work is conducted under strict operating conditions and analyzed using high-resolution molecular tools. The experimental results are then reviewed in the light of the original simulations to check for any anomalies that might ultimately affect the ability to analyze the stochastic component of the dynamics. The deterministic model and experimental setup are discussed below.

6.1.1 Deterministic Model for Neutral Dynamics with Immigration Experiment

The equations of a chemostat describing bacterial growth served as the deterministic model for this work. For any system through which matter flows, a steady state mass balance can be described by,

$$\text{Inflow Rate} - \text{Outflow Rate} + \text{Generation of products} = \text{Rate of accumulation} \quad (6:1)$$

In a chemostat system, the death rate of cells is assumed to be negligible, as the system ensures the cells are maintained at the exponential phase. Also since there is only inflow of pure feed and no accumulation at steady state, the biomass balance at steady state reduces to,

$$\text{Generation of products} = \text{Outflow Rate} \quad (6:2)$$

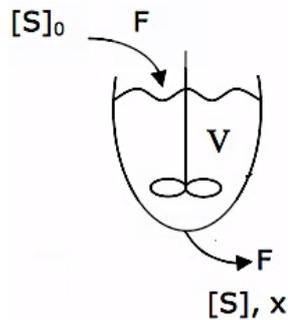


Figure 6-1: Schematic representation of mass balance equations around a bioreactor, where V = Volume of reactor, F = flow rate in and out of reactor as indicated by arrows, x = biomass and [S]₀ = original concentration of growth limiting substrate.

Figure 6-1 shows a classical schematic diagram of a completely mixed tank reactor, which, when operated at steady state becomes a chemostat. Then

$$r_x V = Fx \quad (6:3)$$

where V is the volume, r_x is cell growth kinetics and F, the flow rate. The cell growth rate, r_x , is the product of the specific growth rate of the cells, μ and the cell biomass itself, X ($r_x = \mu X$) while the flow rate to volume ratio (F/V) gives the dilution rate, D. Equation (6:3) above becomes,

$$\mu = D \quad (6:4)$$

This implies that at steady state, the dilution rate of the system is the specific growth rate of the cells. With any mixed tank reactor it is nearly impossible to achieve steady state (chemostat) conditions instantaneously and thus there is a period when the bacterial population densities and the substrates in the system adjust en route to steady state. It is important to know how long this adjustment will take for any experimental system so that samples can be taken once steady state has occurred. Thus a dynamic model is considered where,

$$\text{Rate of accumulation} = \text{Generation of products} - \text{Outflow Rate} \quad (6:5)$$

This becomes

$$V \frac{dx}{dt} = r_x V - Fx, \quad (6:6)$$

Or

$$\frac{dx}{dt} = \mu x - Dx \quad (6:7)$$

Where t is time. Substituting the expression of μ from the Monod kinetics in equation (5:1), equation (6:7) becomes

$$\frac{dx}{dt} = \frac{\mu_{max} S}{K_s + S} X - Dx \quad (6:8)$$

The substrate balance is given by

$$\begin{aligned} \text{Rate of accumulation} &= \text{Inflow Rate} - \text{Outflow Rate} + \\ &\text{Generation of products} \end{aligned} \quad (6:9)$$

Which for the mixed tank reactor becomes

$$V \frac{dS}{dt} = FS_0 - FS - \frac{\mu X}{Y} V \quad (6:10)$$

Or

$$\frac{dS}{dt} = D(S_0 - S) - \frac{\mu X}{Y} \quad (6:11)$$

Where S_0 is initial substrate concentration, S is the final substrate concentration, Y , cell growth yield. For the substrate, products are consumed rather than generated and the rate of consumption of the substrate to the cells produced is related by the growth yield (Y). Also, rate of inflow is same as that of outflow. Again substituting the expression of μ from Monod's Kinetics gives

$$\frac{ds}{dt} = D(S_0 - S) - \frac{\mu_{max} S}{K_s + S} \times \frac{X}{Y} \quad (6:12)$$

Where as before μ_{max} is the maximum specific growth rate and K_s is the half saturation constant. Equations (6:8) and (6:12) describe bacterial growth in the source (meta) community for the biomass and substrate balance respectively for each strain. A Matlab program was written that solved this system of ordinary differential equations (ODEs) using an ODE package (ODE45) that is designed for non-stiff ordinary differential equations. For the neutral case the parameters μ_{max} , K_s and Y in the equations were set to those determined experimentally, in chapter 5, in the absence of antibiotics, which are unique for each strain though they were very similar for both strains in this case.

The local community was modeled slightly differently to the source community because the influx of culture (cells and substrate) from the source community had to be taken into consideration. Also both populations exist in the local community (mixed culture) so there is an additional differential equation.

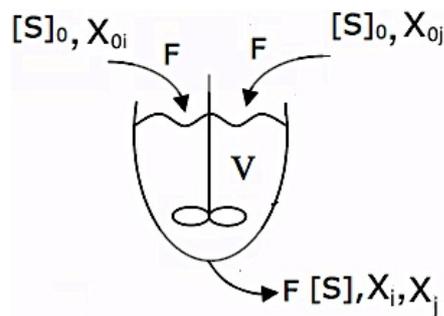


Figure 6-2: Schematic representation of mass balance equations around the local community. The subscripts i and j represent the Cm+ and Kan+ strains. In this case $X_{0i} = X_{0j}$ and will subsequently be written as X_0

The biomass balance is given by

$$\frac{dX_i}{dt} = D(X_0 - X_i) + \frac{\mu_{maxi} S}{K_{si} + S} X_i \quad (6:13)$$

where X_i is the abundance of the i th strain, X_0 is the abundance of the strains coming from the source community. The substrate balance is

$$\frac{dS}{dt} = D(S_0 - S) - \left(\frac{\mu_{maxi} S}{K_{si} + S} \times \frac{X_i}{Y_i} \right) - \left(\frac{\mu_{maxj} S}{K_{sj} + S} \times \frac{X_j}{Y_j} \right) \quad (6:14)$$

These three equations, two for biomass balance for both strains and the other for substrate balance, were encoded into a Matlab programme to simulate the dynamics in the local community. Simulations were conducted for both the large and small vessels using the parameters that were determined in Chapter 5 and the flow rates and volumes were adjusted in order to explore feasible experimental designs.

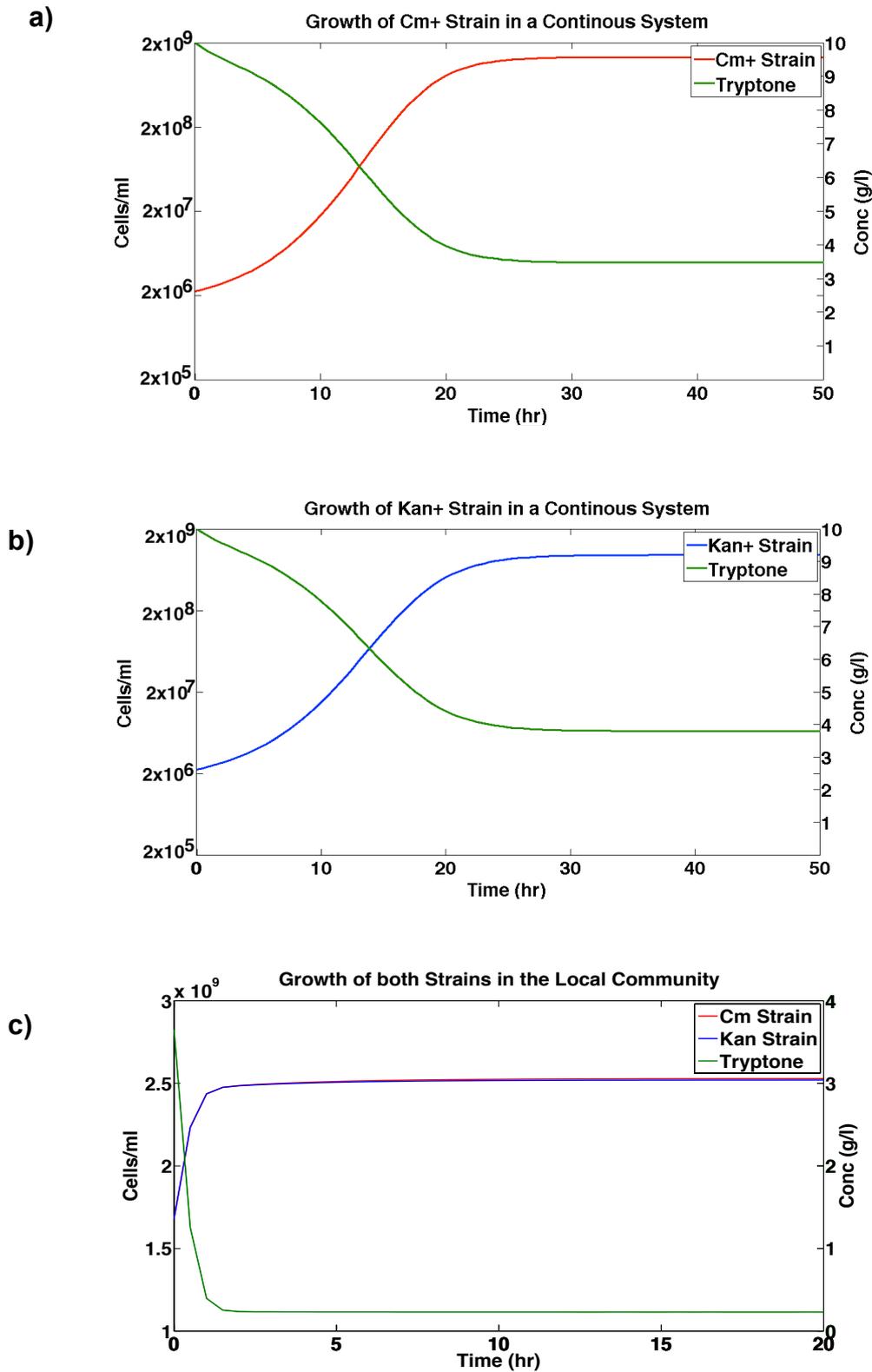


Figure 6-3: Simulation results for growth in both source communities using a dilution rate of 0.5hr^{-1} a) and b) and c) that for the local communities with a dilution rate of 0.33hr^{-1} when no advantage is applied.

Figure 6-3 gives the result of the simulations that were ultimately used to design the experiments. Figure 6-3a and Figure 6-3b show the growth of the Cm⁺ and Kan⁺ strains in the two large vessels that act as the source communities along with the depletion of substrate (tryptone) respectively. The consumption of tryptone is governed by the growth and yield relationships derived in chapter 5. This results in the growth of both strains at a rate of approximately 1.1hr⁻¹. The simulation predicts that the concentration of cells, and substrate reaches steady state conditions approximately 30 hours after the initial inoculation. The density of cells at steady state is approximately 1.7x10⁹ per ml for both strains. These results were achieved using a dilution rate of 0.5hr⁻¹. This dilution rate was chosen to ensure there was sufficient residual substrate in the effluent to support an actively dividing population when delivered to the small local community bioreactors. This meant that no additional fresh media was required, which simplified the design and allowed better control of the dynamics.

Figure 6-3c shows deterministic simulations of the abundance of the strains and the substrate in the small reactors that constitute the local communities. The residual media in the effluent from the source communities is predicted to have a substrate concentration of approximately 3.5g/l, which would support a growth rate of 0.5hr⁻¹ according to the calibrated Monod kinetics (Chapter 5), which is slower than for the source community. This means that for substrate utilization one would expect a mixed order (a combination of both zero and first order) kinetics, as K_s is close to S, and first order kinetics for cell growth. Steady state conditions in the local community occur quickly (< 3 hours). Of

course the growth rates in the completely mixed small reactors are expected to be even lower than those in the pure communities and the resulting growth rate from the simulation is approximately 0.37hr^{-1} . For this simulation, the growth rates of both strains are equal and the total cell concentration for both strains at steady state is approximately 2.5×10^9 .

The simulation results in Figure 6-3 are the final realization of multiple model runs where dilution rates were systematically adjusted to give a workable experimental design where steady state non-trivial behavior might be observed in a realistic timeframe. The two source community reactors were fed at a rate of 7ml/min. Culture was extracted at the same rate and entered the two manifolds, which was then pumped to the 10 smaller reactors and a flow rate of 0.55ml/min each. Thus each of the smaller reactors received migrants and substrate from the two different sources and so had a total rate of influent of 1.1 ml/min. The steady-state volumes of media in the source and local communities were 800ml and 200ml respectively.

6.1.2 Experimental Setup for Neutral Dynamics with Immigration Experiment

The experimental setup is intended to mimic the results achieved in the simulation. To achieve this, 2L bioreactors were used to house the source communities, 1L volumes were chosen for the manifolds that were used to distribute migrant populations to the local communities and 250ml volumes were used for the small reactors that house the local communities. Multistirrer platforms were used to achieve consistent stirring rates in all bioreactors, peristaltic pumps with eight rotors were used to accurately dispense volumes and the same tubing material, bore and lengths for connectivity were maintained throughout the experiment. Figure 6-4 shows the conceptual chemostat design for this experiment while Figure 6-5 shows photographs of the bioreactors used in this study.

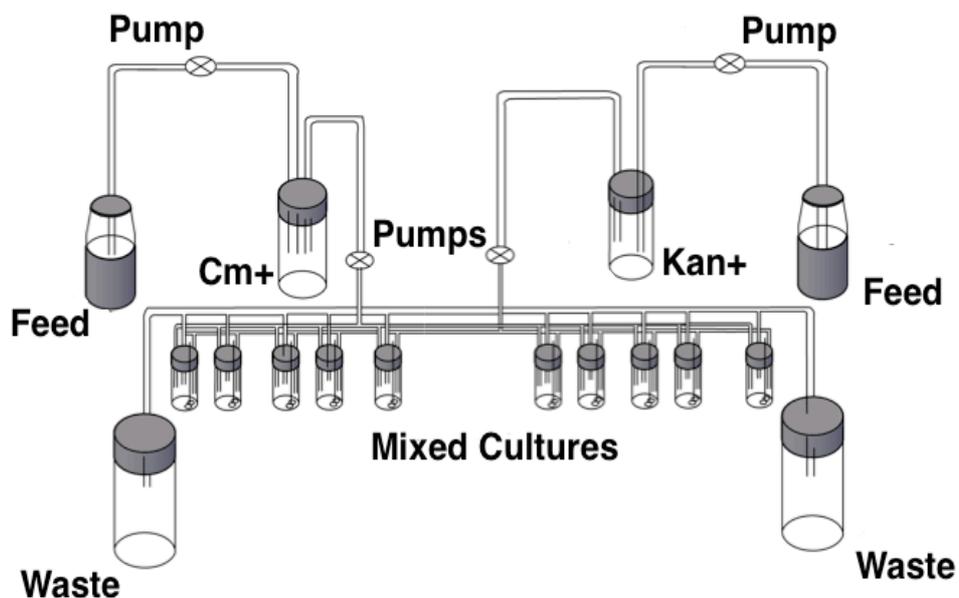


Figure 6-4: Conceptual chemostat design for neutral dynamics experiment showing arrangement of bioreactors, pumps and connecting tubes.

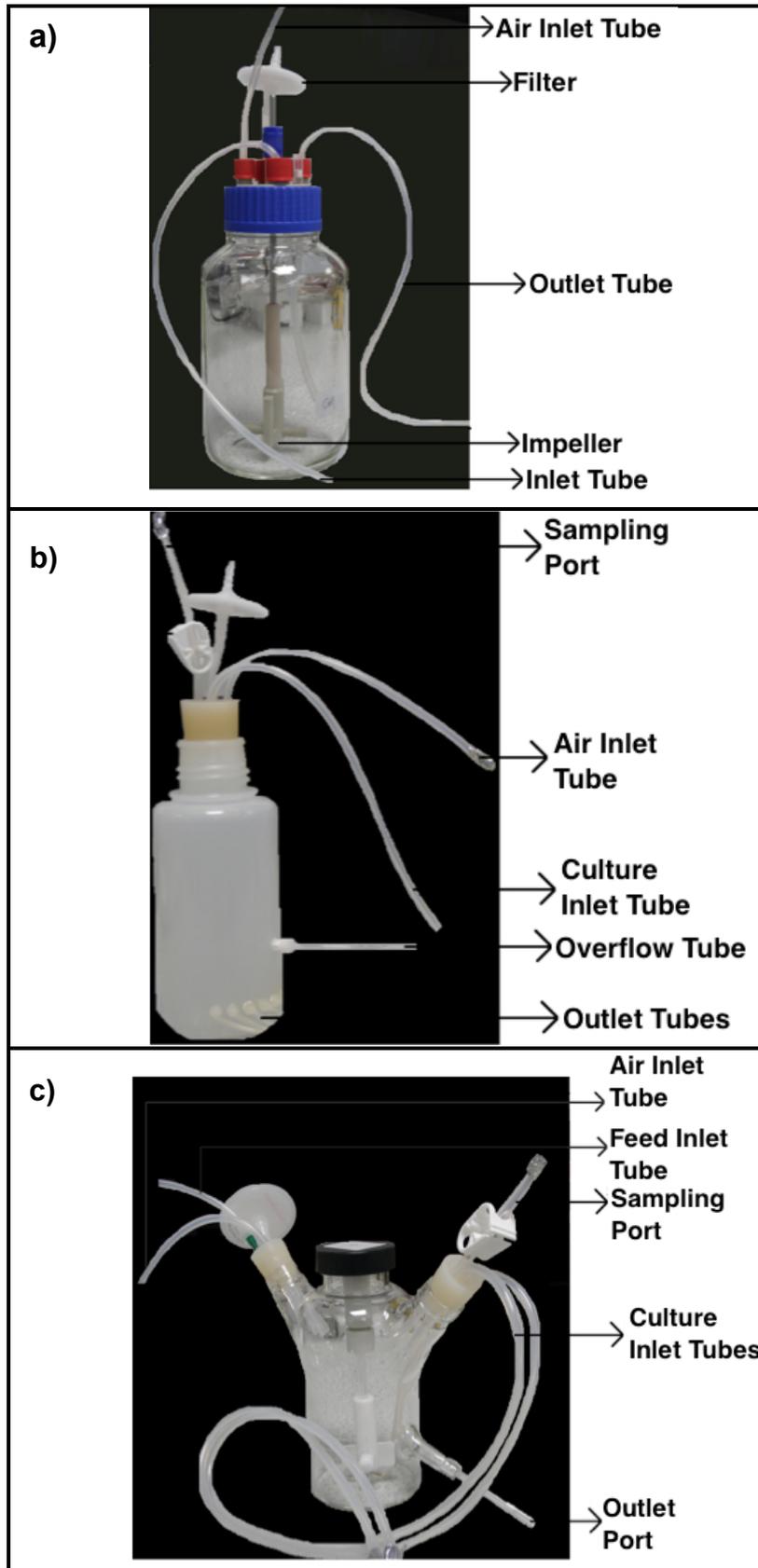


Figure 6-5: Figure showing individual pieces of bioreactors used in the continuous reaction experiments. a) The bioreactors used for the pure cultures of the strains in the source communities, b) The manifold used to distribute migrants and c) The reactors that hold the mixed culture of both strains that represent the local communities.

A total of fourteen bioreactors were used in this experiment - two source communities (Figure 6-4a) comprising separate pure cultures of Cm⁺ and Kan⁺ strains, 2 manifolds (Figure 6-4b) that accept migrants from the source and distribute them to local communities and then 10 small reactors (Figure 6-4c) housing local communities each of which are subject to identical conditions. The volume in the manifold was kept to a minimum using an overflow tube; this ensured that the dynamics was kept to a minimum and that manifold merely served to distribute culture from the source to the local communities. The glass jars for the small reactors were also modified to have overflows. This ensured that the flow rate in equaled the flow rate out and hence the fluid mass balance was in steady state. All bioreactors were seated on multistirrers with one central control knob to achieve consistent rotation speed and the working volumes used in the experiment were smaller than the reactor volumes, providing ample headspace for gaseous exchange; both factors affect aeration and thus could affect growth rates. To avoid contamination, silicone bungs were used as stoppers on the reactors and the exact tubing size was bored through them providing an airtight fit. Syringe filters of appropriate diameter corresponding to the working volume in each reactor were used to further prevent cross-culture contamination. Filtered air was supplied to each of the reactors using an air pump at a rate of 300ml/min.

The units were assembled into a temperature-regulated greenhouse, made of polycarbonate to help conserve heat, which served as the units' growth chamber. Temperature control was achieved by a temperature regulated heating device, whose temperature profile when set to 37°C (temperature for

experiments) gave a $\pm 2^{\circ}\text{C}$ accuracy. All equipment was calibrated before every experiment.



Figure 6-6: Picture showing assembled units of apparatus for the chemostat experiment.

For this experiment, the system has been set up according to the following conditions.

1. Pure cultures of the two strains were grown in each of the 2L bioreactors, one strain in each bioreactor.
2. Tryptone broth at a concentration of 10g/l was used as feed for the organisms.

3. Two double head peristaltic pumps were used to feed both source communities through tubes at a rate of 7ml/min each. A second tube left each of the reactors and went through the pump and into a manifold, one for each strain, also flowing at 7ml/min. Thus one double head peristaltic pump was used for each strain.
4. The culture, containing bacteria and tryptone, passed rapidly through each manifold and was divided between 10 tubes that were connected to the local community reactors through two 10-way multi-head peristaltic pumps flowing at 0.55ml/min each. Thus the two strains were not mixed until they entered the local communities. Extreme care was taken to ensure both strains arrived at the local community at the same time by using an eight rotor pump head, which dispenses accurate volumes, and also using tubes of same length, bore and material. Clearly 14ml/min, the total culture volume exiting the source community, is not equal to 11ml/min, which is the total volume of culture entering the local community ($0.55\text{ml/min} \times 2$ – total culture of both strain into 1 local community; $1.1\text{ml/min} \times 10 = 11\text{ml/min}$). The experiment was set up this way with an extra 3ml/min for logistic reasons – to ensure the system is always running and give the operator ample handling time.
5. In the small bioreactors, neither chloramphenicol nor kanamycin antibiotics were added. This means that neither of the strains had an advantage over the other.

6. The effluents from the local communities were connected via a 10-way multi-head peristaltic pump to a waste vessel, whose contents were safely disposed of daily.
7. The only variant in the system is the species and the growth rate has been showed to be almost identical deterministically. Thus the system is expected to behave neutrally.
8. All units were washed, sterilized and calibrated before the experiment.

PROCEDURE:

Overnight cultures of both strains were grown and using a 1/100 dilution, the source communities were inoculated accordingly and samples of the culture collected immediately afterwards. These pure cultures in the source communities were then grown as a batch for 30 minutes, and continuously thereafter. When the volume of culture in the manifolds reached a level just above the outlet ports to the local communities, the 10-way multi-head peristaltic pumps connected to the local communities were opened and effluents from the manifold bioreactors monitored to ensure that both strains arrived in the local communities at the same time. This occurred after 2 hours only then were samples collected. The experiment was conducted at a constant temperature of 37°C and lasted for 155 hours. Samples were collected every 2 hours over a 10hr period each day from each of the reactors for the duration of the experiment. Sampling occurred at the same time from each reactor and the samples were collected via a 3.2mm silicone tube attached to each reactor. Using a 10ml disposable syringe, 5ml of culture was

drawn from each reactor into corresponding 15ml sterilized labeled test tubes, which were placed on ice to help stop growth. The silicone tubes were subsequently sealed to prevent contamination. The 15ml test tubes were immediately vortexed and 2ml of culture from each distributed into two 2ml labeled sterilized eppendorf tubes (one for sample analysis and the other as a backup), which were spun in a centrifuge at 10,000 rpm for 10 minutes to prevent further growth; the supernatant was discarded safely. The resulting pellets were stored at -20°C for analysis. These steps were followed to provide an accurate account of the proportions of the strains at that time point.

The subsequent analysis involved pretreating the pellets at 95°C for 10 minutes to release the molecular contents of the cells including the DNA. A detailed qPCR analysis using reagents from Biorad was conducted on the burst cells and their contents to determine the proportions of both strains in all bioreactors. For each sample collected, qPCR was conducted with primers and probes that target the chloramphenicol resistance gene and those that target for the kanamycin resistance gene. These were done in triplicate and the average C_t value taken. The attractiveness of this process lies in the fact that the species used here have engineered so that there is just one antibiotic resistance gene in each individual. Thus there is a one-to-one correspondence between the antibiotic gene detected and the abundance of engineered bacterial strains. Figure 6-6 shows the proportions of both strains in the source community in terms of absolute and relative abundance while Figure 6-7 shows their proportions in the ten local communities in terms of relative abundance.

6.1.3 Experimental Result for Neutral Dynamics with Immigration

Figures 6-6a and b show results for the source communities in terms of their absolute abundance (shown on a semi-logarithmic scale) and relative abundance respectively. It can be clearly seen that the growth of the cells was maintained in the exponential phase and thus the abundances were approximately constant as expected of a continuous system. The total number of Cm⁺ and Kan⁺ cells in the source community are approximately 8.01×10^8 and 7.93×10^8 respectively. Importantly, the strains showed growth rates that were very similar; 1.3hr^{-1} for the Cm⁺ resistant strain and 1.2hr^{-1} for the Kan⁺ resistant strain. The system had achieved steady state in both source communities after approximately 4 hours. Sampling in the small reactors began at a time point (4 hours) when sufficient amount of culture could be collected.

Results for the ten replicate local communities are presented as time series of relative abundance of the strains in Figure 6-7. Whilst every effort was made to maintain the source community abundances at a constant level there were fluctuations. These probably arise from extrinsic factors such as temperatures changes (Hilfinger, 2011). In particular, fluctuations in temperature ($\pm 2^\circ\text{C}$) were observed. The fluctuation in the source community abundance will be reflected to an extent in the local communities (small chemostats) through the high immigration rate. However, all the small reactors demonstrate different time series of relative abundance and thus any fluctuation that is induced by the migrating population is superimposed by additional local fluctuations that differ from chemostat to chemostat. This variability between reactors in the

local community is shown in Figure 6-7c using sampling time 22 – 34 hours as an example. These local fluctuations are believed to be dominated by intrinsic stochasticity induced by the random birth death process. In the next chapter, this hypothesis is tested.

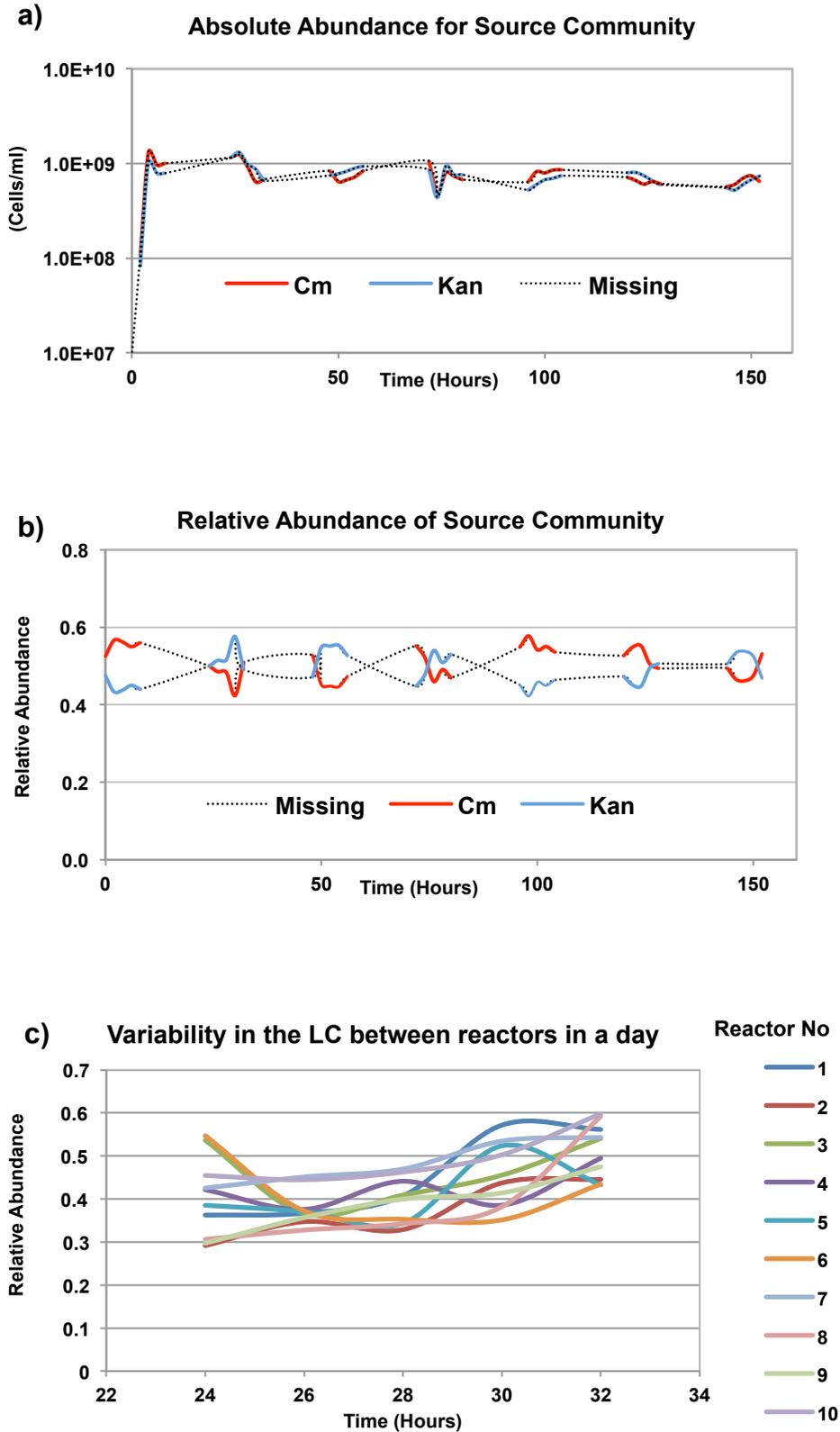


Figure 6-7: Diagram showing growth of both strain in the source community with a dilution rate of 0.5hr^{-1} in a) absolute abundance, b) relative abundance and c) diagram showing variability between reactors in the local community from sampling times 22 – 34 hours. Dilution rate in the local community was 0.33hr^{-1}

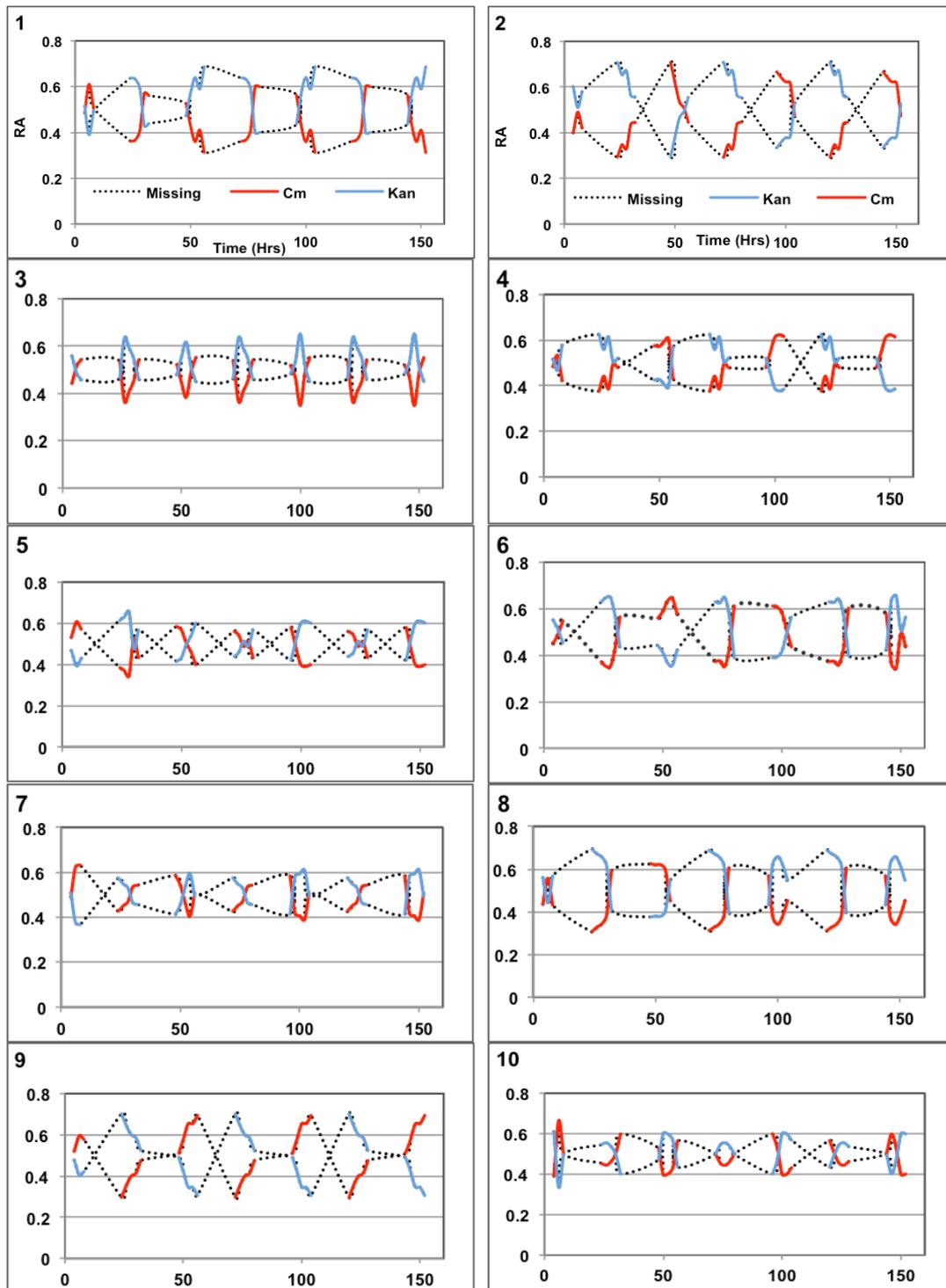


Figure 6-8: Growth of both strains in the 10 local communities when no selective advantage exists. Growth was achieved using a dilution rate of 0.33hr^{-1} . The legend shown for reactors 1 and 2 applies to all reactors – the red data points depict the Cm+ strain and the blue data points the Kan+ strain. The missing data points depict nighttime when sampling could not be conducted.

Quantitatively, the strains fluctuate between 0.7 and 0.3, with the highest fluctuations occurring in reactors 1, 2, 8 and 9. The pattern of fluctuation does not seem to be uniform across all reactors; reactors 1, 4, 6 and 8 show a degree of similarity that could be speculated as been driven by the immigrant population, reactors 2 and 9 share striking resemblance to each other, reactor 3 seems quite different from the others while smaller variances are observed in reactors 5, 7 and 10 and thus they all tend to be close to the mean value of 0.5.

It is clear that many of the generic features of neutral dynamics are exhibited in the experimental data: serial correlation, random variability between reactors (see Figure 6-6c) and the tendency to drift back to the source abundance. Thus superficially at least the chemostats appear to be delivering the neutral dynamics that they were designed for.

6.2 Non-neutral Dynamics in a Completely Insular Community

This experiment is motivated by certain agricultural and engineering applications, such as in bioaugmentation practices. In such cases, the aim is to introduce a new or enhanced function to the system. If the system is completely insular and so there is not a constant feed of immigrants to shape the community composition, then if a taxon has an advantage over others feeding off the same substrates, according to niche theories, then it should win out. Conversely, if it has a disadvantage then it should lose out. This is

commonly known as a competitive exclusion. Stochasticity may act to disrupt this competitive exclusion. To devise an experiment with the two strains of *E.coli* where such behavior is likely to be seen in realistic time frames, again the chemostat model is used. The conceptual chemostat design for this experimental set up is shown in Figure 6-9 below

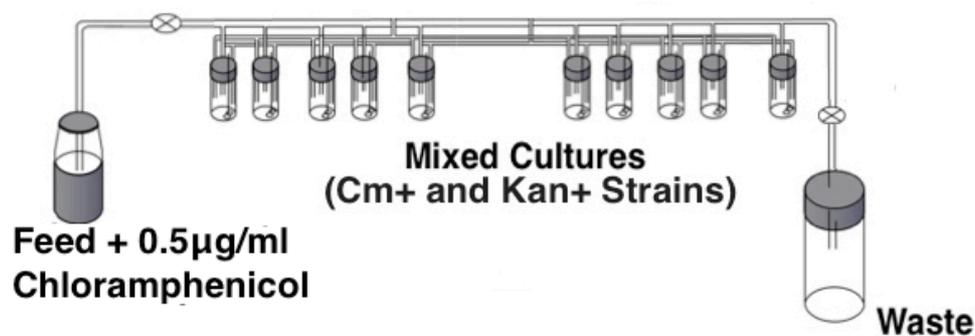


Figure 6-9: Schematic showing conceptual arrangement of chemostat for the Non-neutral dynamics in a completely insular community' experiment.

6.2.1 Deterministic Model for Non-neutral Dynamics in a Completely Insular Community

The model is that given in equations (6:9) and (6:10). Again, these balance equations were solved in Matlab and results of the simulation are shown in Figure 6-8. Note in this case immigrants are not fed into the local communities and hence equations (6:6) and (6:8) do not need to be solved.

In this case the parameters K_s and μ_{max} in the growth equations differ for the two strains. The parameters that were determined in chapter 5 when 0.5 μ g/ml chloramphenicol is added to the media were used and also a dilution rate of 0.17. The dilution rate was chosen as it provided the most appropriate dynamics given the capacity of pump. Higher dilution rates brought the system too close to washout.

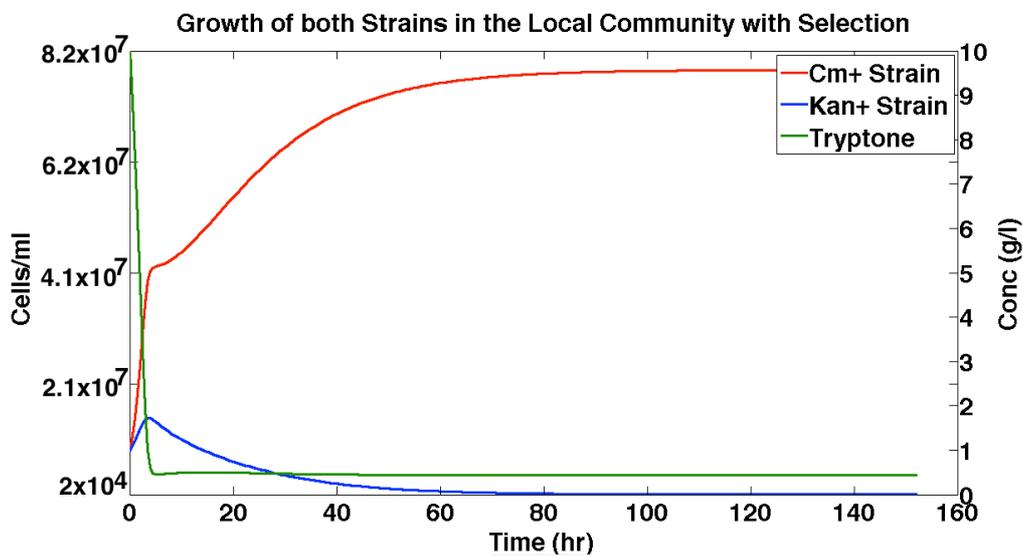


Figure 6-10: Simulation result of the local community when selective advantage is added to the system.

It can be clearly seen that the Kan+ strain is disadvantaged in the community as expected. Both strains seem to show the growth curve trend at the beginning of the experiment but due to the selective effects of the media, there reaches a point where the dilution rate exceed the growth rate for the Kan+ strain and it begins to wash out of the system. About 50 hours into the stimulation, a steady state condition is achieved for both strains in which the

total number of cells per ml of the Cm⁺ strain is approximately 8.2×10^7 while that of Kan⁺ is reduced to about 2.13×10^4 having starting off at the same concentration of 8.5×10^6 . The simulations suggests that 0.5µg/ml chloramphenicol with a dilution rate of 0.17 are appropriate conditions to impose in the experiments to demonstrate the effects of a selective advantage with no immigration.

6.2.2 Experimental Setup for Non-neutral Dynamics in a Completely Insular Community

For this experimental setup, one manifold and ten 250ml bioreactors for the local community were used. As migration effects are not considered, bioreactors for the source communities were omitted. A manifold with 10 exit tubes was used as the connection between the feed reservoir and the ten local communities. Again, tubes of the same length, material and bore were used along with peristaltic pumps for accurate dispensation of volumes. All reactors were stirred using a multistirrer with central control knob providing aeration and similar stirring speeds. In order to reproduce the simulation result, the effluent flow rate from the manifold reactor to each of the local communities having 200ml volume of culture was 0.55ml/min, giving a 0.17 dilution rate in each local community, the same as that of the simulation. For the selective advantage experiment, the system was set up as follows:

1. Tryptone broth was prepared at a concentration of 10g/l to which 0.5µg/ml chloramphenicol was added. This was the feed for the organisms.

2. A single head peristaltic pump supplied feed from the main reservoir to the manifold reactor at a rate of 6ml/min, which in turn supplied the local communities at a rate of 0.55ml/min via a 10-way multi-head peristaltic pump.
3. The effluents from the local communities were connected via a 10-way multi-head peristaltic pump to a waste vessel, whose contents were safely disposed of daily.
4. Both strains were introduced into each local community at the start of the experiment (time point zero). Extreme care was taken to ensure that the local communities were inoculated with the same volume of culture of both strains.
5. The strains were not inoculated any more during the course of the experiment neither was additional antibiotics added to the local communities at any point except the one conveyed via the feed.
6. All units were washed, sterilized and calibrated before the experiment.

PROCEDURE:

Cultures of both strains were grown overnight and using a 1/100 dilution, each local community was inoculated with both strains and samples of the culture collected immediately afterwards (sample collection at time zero). The mixed cultures in the local communities were grown batch wise for 30 minutes, and continuously afterwards with the introduction of the feed from the manifold reactor via the 10-way multi-head peristaltic pump. The experiment was conducted at a constant temperature of 37°C and lasted for 155 hours.

Samples were collected every 2 hours for 10 hours during the day from the all reactors. This was maintained for the duration of the experiment. Samples, as before, were taken by filling a 10ml disposable syringe and the samples were labeled and stored in the fridge. Again the abundance of the strains was quantified by qPCR.

6.2.3 Experimental Result for Non-neutral dynamics in a completely insular community

Figure 6-9 shows the relative abundance of both strains in the local communities. It can be clearly seen that with the addition of 0.5µg/ml chloramphenicol to the media, the Kan⁺ strain is disadvantaged, and as in the simulation has diminished to zero or a very low relative abundance after approximately 50 hours. Both strains began with approximately 1.38×10^7 cells/ml and at the end of the experiment, averaging across all the reactors, the Cm⁺ strain had about 6.44×10^8 cells/ml whilst the Kan⁺ strain had 7.75×10^5 cells/ml. The survival of a species in an environment depends on its ability to reproduce and transfer the genes responsible for its adaptability in that environment to its offspring. This is the basis of natural selection (Futuyma, 2009). In this experiment, the ability of the Cm⁺ strain to persist in an environment with chloramphenicol shows that the chloramphenicol resistance genes in the Cm⁺ strains are indeed in their genome.

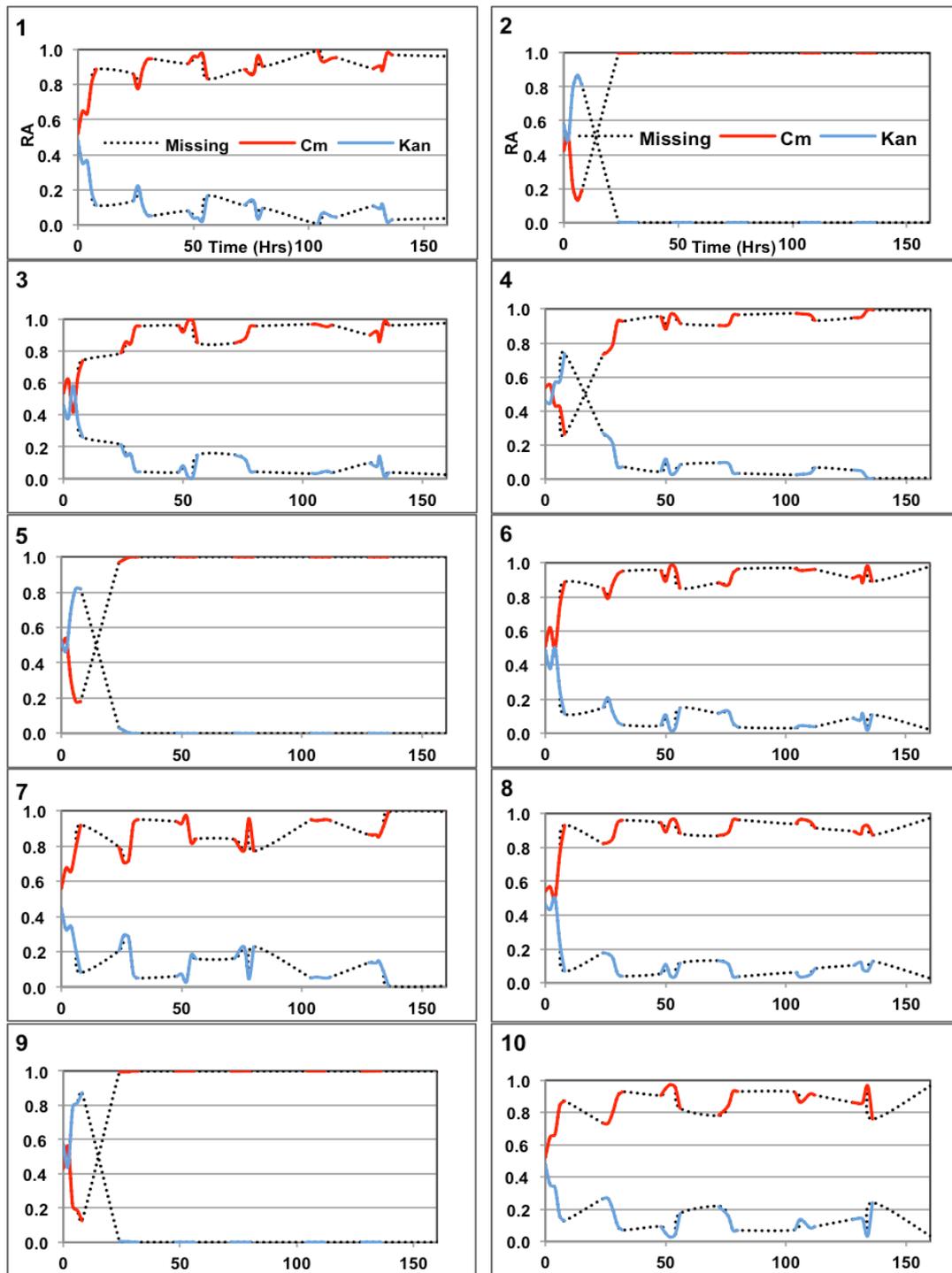


Figure 6-11: Growth of both strains in the 10 local communities when only selective advantage exists in the system. Growth was achieved using a dilution rate of 0.17hr^{-1} . The legend shown for reactors 1 and 2 applies to all reactors – the red data points depict the Cm+ strain and the blue data points the Kan+ strain. The missing data points depict nighttime when sampling could not be conducted.

The variability in relative abundance across all of the reactors vindicates the use of so many experimental replicates as recommended by Zhang (2004). Reactors 2, 5 and 9 show are similar to each other with the Kan⁺ strain almost washed out of the system and devoid of fluctuations observed in other reactors. The variability in the strains beyond 50 hours in the other reactors suggests that even when the Kan⁺ strain diminished to very low concentrations, the population can recover. This could not be simulated by the deterministic model. The fluctuation may reflect random demographic drift, but subjectively their amplitude looks larger than one might expect. It can thus be concluded that 0.5µg/ml chloramphenicol was an adequate concentration to apply for selection to be observed in the system. The experiment was done to show the effects of selective advantage in the community before adding migration as an ecological process onto it as seen in the next section.

6.3 Non-neutral dynamics with immigration

In this section, a selective advantage is conveyed on one of the strains when it immigrates into the local community, but the community is open to migrants of both strains. This is akin to engineering interventions where a species is introduced into a foreign environment to carry out a specialized function (Hosokawa, 2009). The environment may already have balanced population sizes of coexisting species with partitioned niches and thus the survival of the incoming species is crucial to the role for which it is being recruited to perform. Such scenarios arise in for example wastewater treatment plants and bioremediation strategies when inocula are introduced; the frequency and

duration of re-inoculation of species is often problematic posing serious challenges to bioremediation (Lebeau, 2011). Two dominant opposing forces responsible for community assembly are believed to be at work at such environments, that of stochasticity via re-inoculation and resource partitioning via selection, hence understanding the interplay between these forces could help improve the efficiency of such processes and maximize process cost.

To investigate this, I build on the experiment involving selection from the previous section by preserving the selective advantage in the local community and adding migration from the source community for both species. A schematic experimental set up is presented in Figure 6-12 below. Having given circumstantial evidence of demographic stochasticity and selective advantage using the modified strains under different chemostat conditions in the previous two experiments, I have shown that an environment can be successfully created where both deterministic and stochastic processes can be combined and investigated. In this way, an insight into the interplay between these forces can be gained. Again, using the process parameters generated both in the presence and absence of antibiotics from Chapter 5, the classic deterministic equations (6:6), (6:8), (6:9) & (6:10) of a chemostat for the conditions specified here, which would aid in the experimental design of this scenario, can be solved. Note that in this case, the equations for both the source community and the local communities are solved.

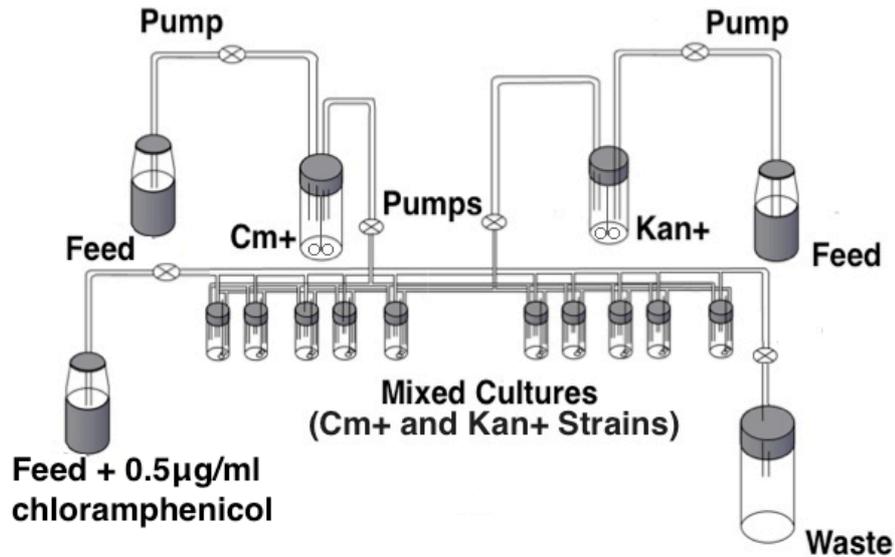


Figure 6-12: Diagram showing chemostat design for the 'Non-neutral dynamics with immigration' experiment.

6.3.1 Deterministic Model for Non-neutral dynamics with immigration

To incorporate the selective advantage only in the local community, an additional inflow of feed to the local community was used to which 0.5µg/ml chloramphenicol was added. For the purposes of the simulation then, different Monod kinetic parameters were used for growth in the source community from those used in the local communities. The results of the simulations are given in Figure 6-13.

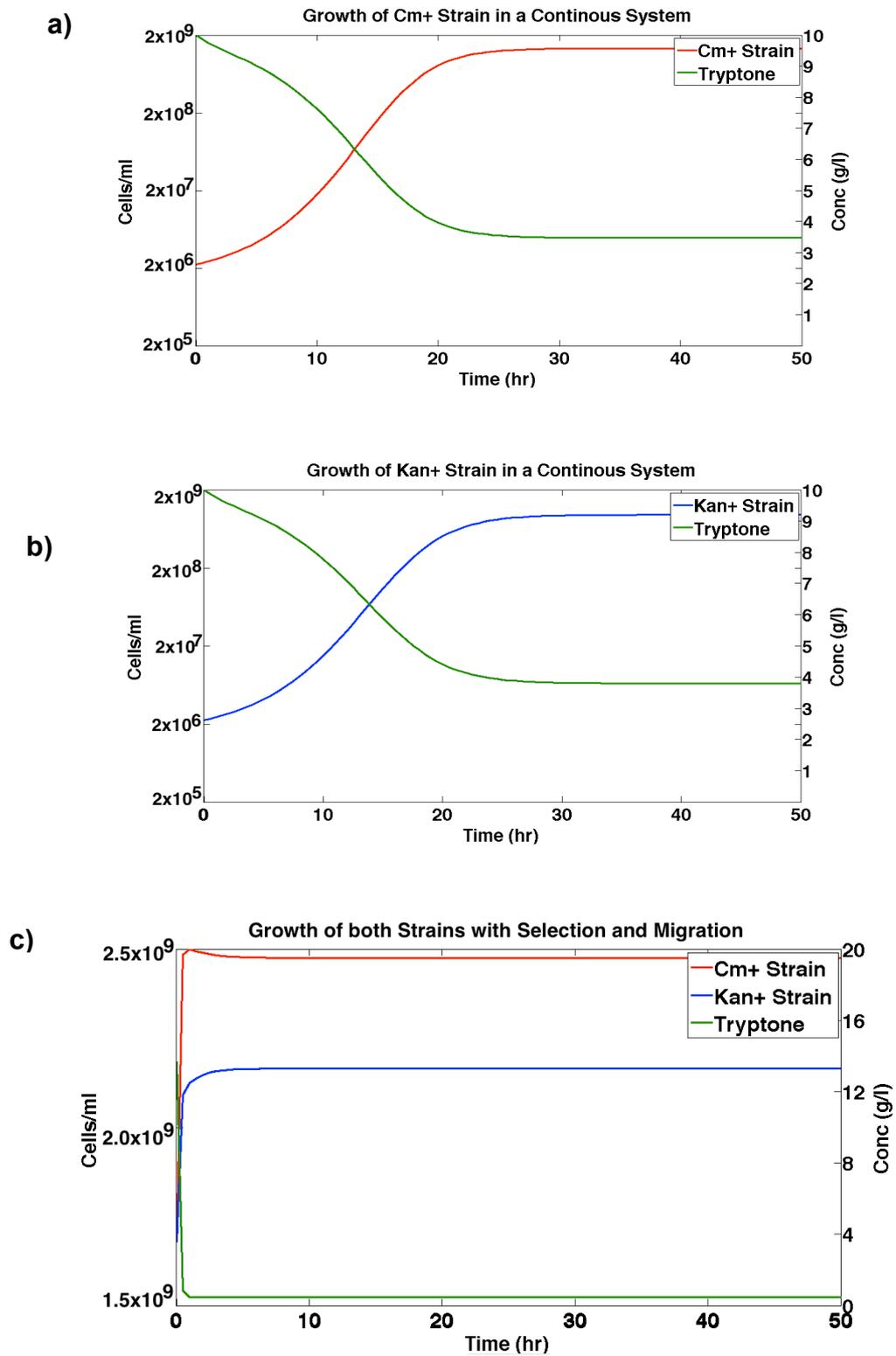


Figure 6-13: Simulation results for the growth of both strains as pure cultures in their respective source communities using a dilution rate of 0.5hr^{-1} (a and b) and c) in the local communities using a dilution rate of 0.42hr^{-1} when both selection and migration experiment.

Figure 6-13a and Figure 6-13b show simulation results for both strains in their respective source communities, achieved using a dilution rate of 0.5hr^{-1} . In both source communities, their growth rates are about 1.1hr^{-1} . Steady state conditions are achieved in both reactors at approximately 30 hours with approximately 1.7×10^9 cells per ml for both strains at the end of the simulation. Using this dilution rate, cells arriving in the local communities from the source communities would still be actively dividing.

Figure 6-13c represents the simulation results for one of the local communities achieved using 0.42hr^{-1} as the dilution rate. In the local community, steady state conditions also occur round 30 hours from the start of the simulation. Though selection exists here, the total substrate concentration, which is about 13.5g/l (approximately 3.5g/l residual media emanating from the source communities and 10g/l from the fresh media delivered only to the local communities) is higher than those in the source communities. The Cm⁺ strains grew at a rate of 0.8hr^{-1} , having 2.5×10^9 cells per ml at the end of the simulation, while the Kan⁺ strains grew at a rate of 0.44hr^{-1} and had 2.2×10^9 cells per ml at the end of the simulation, thus the percentage decrease in the population size of Kan⁺ strain compared to those of Cm⁺ strains is 12%. Although the same selective advantage was applied as in the previous experiment, it can be clearly seen that the population size of the Kan⁺ strain is maintained at orders of magnitude higher levels. This increase can be attributed to migration from the source communities and additional media, which is supplied to the local communities.

Also the growth rate of Kan⁺ strain from the simulation above (0.44hr^{-1}) and the maximum growth rate of the Kan⁺ strain in $0.5\mu\text{g/ml}$ chloramphenicol from the Monod Kinetics result shown in Table 5-3 (0.43hr^{-1}) are exact, depicting that the simulation results exist within the limits of the characterization of the species in $0.5\mu\text{g/ml}$ chloramphenicol. It also implies that using this dilution rate; the Kan⁺ species grew to their maximum in the environment. Replicating this simulation experimentally involves achieving the respective dilution rates in the source communities and local communities. The influent flow rate to each source communities is 6ml/min that to the local communities from each manifold reactor is 0.55ml/min and that from the feed inflow with the selective advantage is 0.33ml/min . The reactor volumes from previous experiments were retained.

6.3.2 Experimental Setup for Non-neutral dynamics with immigration

As both migration from the source communities to the local communities and selectivity only in the local communities are integral in this experiment, all the reactors were used - two 2L bioreactors for the source community, two manifolds and ten 250ml bioreactors for the local communities. Migration from the source communities to the local communities was incorporated, as in the 'neutral' experiment, through the manifold reactors while selectivity in the local communities was integrated by an additional reservoir of feed, to which $0.5\mu\text{g/ml}$ chloramphenicol was added, flowing continuously at a rate of 0.33ml/min through a 10-way multi-head peristaltic pump. The 250ml bioreactors were modified by fitting overflow tubes, which were connected to

the waste containers; the overflow maintained a fixed volume in the chemostats and ensured that inflow and outflow rates were the same. This was required because the initial waste pump was used for the additional feed delivery. Again, tubes of same length, material and bore were connected to the peristaltic pumps for accurate and uniform dispensation of volumes. All reactors were stirred using a multistirrer with central control knob providing aeration and similar stirring speeds. In order to reproduce the simulation result, the effluent flow rate from the manifold reactor to each of the local communities, having 200ml volume of culture, was 0.55ml/min and the feed inflow rate to each local community was 0.33ml/min, giving a total flow of 1.4ml/min into the local communities. Thus, a 0.42hr^{-1} dilution rate was achieved in each local community, as in the simulation. For the selective advantage experiment, the system was set up as follows:

1. Pure cultures of the strains were grown in each of the 2L bioreactors, one strain in each bioreactor.
2. Tryptone broth at a concentration of 10g/l was used as the feed for the strains in the source community.
3. Two double head peristaltic pumps were used to feed both source communities at a rate of 6ml/min each. The second tubing through the pump provided the connectivity from the source community to the manifold, also flowing at 6ml/min. One double head peristaltic pump was used for each strain.
4. The cultures from the manifold reactors were transferred to the local communities through two 10-way multi-head peristaltic pump to each of the 10 local communities at a rate of 0.55ml/min. Each manifold reactor

had a 10-way multi-head peristaltic pump, thus the mixing of both cultures was achieved in the local community. Extreme care was taken to ensure both strains arrived at the local community at the same time by using an eight rotor pump head, which dispenses accurate volumes, and also using tubings of same length, bore and material. Clearly 14ml/min, the total culture volume exiting the source community, is not equal to 11ml/min, which is the total volume of culture entering the local community ($0.55\text{ml/min} \times 2$ – total culture of both strain into 1 local community; $1.1\text{ml/min} \times 10 = 11\text{ml/min}$). The experiment was set up this way with an extra 3ml/min for logistic reasons – to ensure the system is always running and give the operator ample handling time.

5. In the small bioreactors, selective advantage in the form of $0.5\mu\text{g/ml}$ chloramphenicol was imposed. A continuous flow of the antibiotic into the small reactors was provided by adding $200\mu\text{l}$ of 10mg/ml chloramphenicol antibiotic stock to a 2L of tryptone broth at a concentration of 10g/l to give the desired $0.5\mu\text{g/ml}$ chloramphenicol concentration in the feed. This was replenished each time it ran down.
6. The 2L feed reservoir containing $0.5\mu\text{g/ml}$ chloramphenicol was supplied to the local communities via a 10-way multi-head peristaltic pump flowing at a rate of 0.33ml/min .
7. All effluents from the local communities were collected via overflow tube allowances made in the bioreactors. The overflow tubes were connected to a waste vessel, whose contents were safely disposed of daily.

8. Close attention was paid to ensure that the feed reservoirs never ran out, thus ensuring the system ran continuously.
9. All units were washed, sterilized and calibrated before the experiment.

PROCEDURE:

Overnight cultures of both strains were grown and using a 1/100 dilution, each 2L bioreactor with 800ml feed volume was inoculated with each strain and samples of the culture collected immediately to provide information on the number of cells at the start of the experiment. The pure cultures in the source communities were grown batch wise for 30 minutes and continuously thereafter while allowing culture to flow into the manifold. When the volume in the manifold got to a level just above the outlet ports to the local communities, the 10-way multi-head peristaltic pumps connected to the local communities were opened and effluents from the manifold bioreactors were monitored to ensure both strains arrived the local communities at the same time. This occurred after about 2 hours; samples were also collected. At the same time, the feed inlet streams of 10g/l tryptone broth containing 0.5 μ g/ml chloramphenicol connected via a 10-way multi-head peristaltic pump supplying only the local community were opened to bring about the selectivity desired in each local community. The experiment was conducted at a constant temperature of 37°C and lasted 10 hours during the day at an interval of 2 hours throughout the duration of the experiment. Sampling was done as before via a 10ml disposable syringe and samples were labeled and stored safely. In addition, qPCR analysis was conducted and the proportions of each strain in each reactor quantified.

6.3.3 Experimental Result for Non-neutral dynamics with immigration

Figure 6-11a shows results for the source communities; plotted on a semi-logarithmic scale. It can be clearly seen that both strains showed similar growth, which was expected because there was no antibiotic in the source community vessels. The growth rate in the source community was 1.5hr^{-1} for Cm^+ and 1.6hr^{-1} for Kan^+ strain and the total number of cells per ml at the end of the experiment was 5.5×10^8 and 5.7×10^8 respectively. As a result of the dilution rate used, the growth of the strains in both reactor types was maintained in the exponential phase, showing they were still actively dividing as predicted by the simulation before migrating to the local communities. Their relative abundance is shown in Figure 6-11b.

Figure 6-12 shows the relative abundances of the strains in the ten local communities. Clearly, the abundances of the two strains are not equal, as selectivity exists. Importantly, the abundance of the Kan^+ strain is maintained at a level well above that reported for the experiment with no immigration (see Section 6.2). The average number of cells per ml at the end of the experiment for the Cm^+ strain is 1.32×10^9 while that of Kan^+ strain is 5.76×10^8 , implying a 56.4% decrease in the population size of the Kan^+ strain compared to that of the Cm^+ strain.

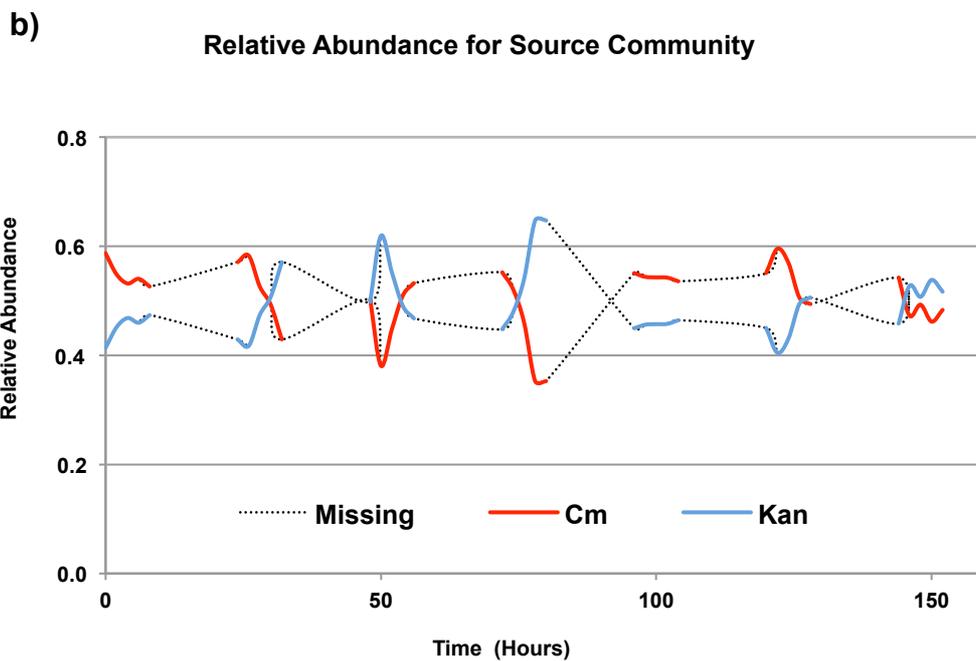
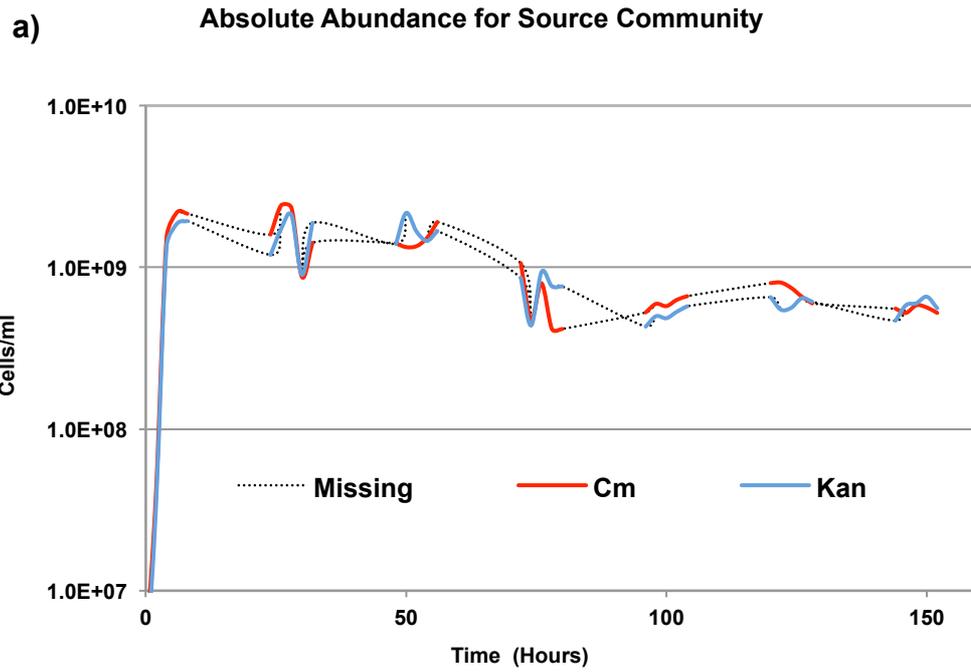


Figure 6-14: Diagram showing growth of both strain as pure cultures in the source community in terms of a) absolute abundance and b) relative abundance with a dilution rate of 0.5hr^{-1} .

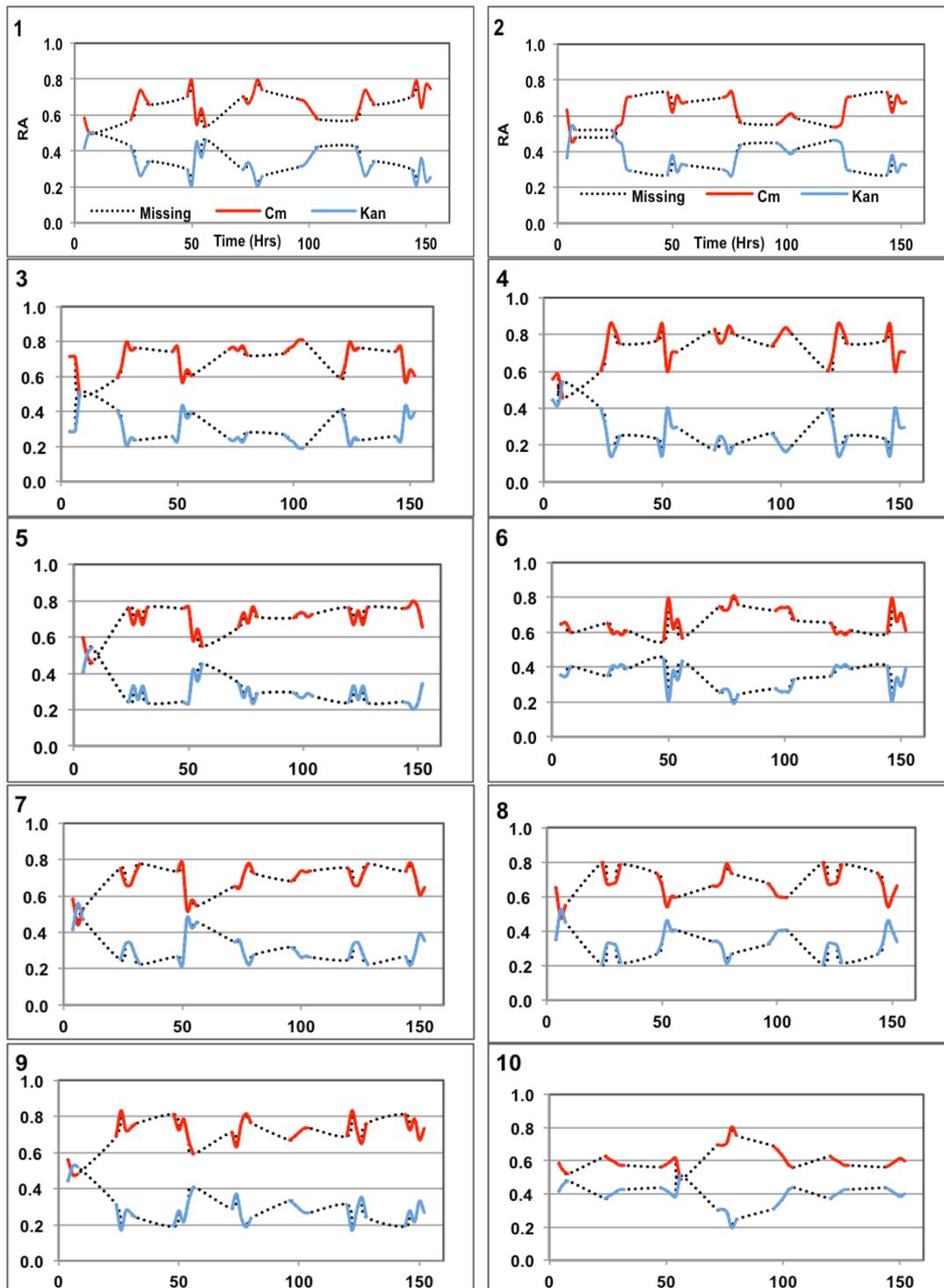


Figure 6-15: Growth of both strains in the local communities when both migration and selection exists with dilution rate 0.42hr^{-1} . The legend shown for reactors 1 and 2 applies to all reactors – the red data points depict the Cm+ strain and the blue data points the Kan+ strain. The missing data points depict nighttime when sampling could not be conducted.

Critically whilst there is a degree of correlation between the abundances in the local and source communities, each of the local communities do exhibit unique dynamics. Furthermore there appears to be serial correlation in the deviation from the source community abundances, which indicates some sort of systematic deviation rather than purely random measurement error around a trend.

6.4 Discussion

This chapter described the process of developing a set of experiments that ultimately will be used to determine the relative importance of deterministic and stochastic community assembly processes in an engineered system. The high degree of control that was engineered in to the two strains of bacteria provides the unique opportunity of rationally designing the experiments based on predictions rather than just vague conceptual ecological theory. Thus, based on the tunable growth characteristics that were characterized in Chapter 5, one could parameterize a deterministic chemostat model and simulate a range of different scenarios, honing in on a set that could feasibly be reproduced in laboratory experiments. Of course the deterministic experiments are not capable of capturing demographic stochasticity, which throughout this thesis has been suggested as a possible important factor. There will always be noise in the measurements from a system like the one presented here. The question becomes, is that noise purely measurement noise or is it a systematic and serially correlated noise associated with a stochastic population process? If the stochasticity in the time series of abundance can be characterized by a weighted random walk then this

would suggest that the randomness is more than just measurement noise and is fundamental to the dynamics. This will be addressed in the next chapter. At this stage then, mere observations on the behaviour of the experiments are made. Firstly, if the mean behaviour across all reactors is considered, then it is in general agreement with the deterministic model predictions. So, for the neutral dynamics with immigration the mean relative abundance of both taxa is 0.5. For the non-neutral dynamics in a community that is completely isolated, one strain becomes mono dominant after about 50 hours. For the non-neutral dynamics with immigration then the strain with the advantage reaches a higher relative abundance but does not win out completely. What is clear is that none of the reactors behave identically. There are local fluctuations about the abundance of the strains in the source community, but any fluctuations seem to be pulled back to the source community abundances. So the reactors do exhibit many of the hallmarks of stochastic dynamics. A subjective observation is that the stochastic dynamics do seem to be rapid for such a large (about 10^{11}) population. This was previously noted in a real engineered system (Ofiteru *et al.*, 2010).

7 Stochastic Modeling

In this chapter, the behaviour of the stochastic dynamics is explored prior to outlining the steps taken to calibrate a stochastic model using experimental data from the chemostat experiments. The calibration procedure is assessed using simulated data, the procedure is then applied to the chemostat data and the biological implications of the results are discussed.

7.1 Stochastic Differential Equations (SDE)

The stochastic differential equations used to describe the ecological processes observed in this study were first used by Ofiteru *et al.*, who successfully calibrated the model using the dynamics of two functional groups of bacteria in the Palo Alto Regional Water Quality Control Plant. Ofiteru *et al.* (Ofiteru *et al.*; 2010) demonstrated how the model cast in discrete time as a Markov process described by transition probabilities can be recast as a set of continuous stochastic differential equation (SDE) for large populations. In the Markov Chain process, the transition probabilities are written in terms of discrete individual replacement events in the population; time is not explicitly referenced. In order to move to the SDE representation, the time taken for a discrete event to occur needs to be related to real time; measured in seconds, minutes or hours. In Ofiteru *et al.* (2010), the SDE was applied to data from a large sewage works and as a consequence both the total population and the time between discrete replacements could only be speculated on. Thus time was scaled in a way that ultimately delivered parameters that did not require

this knowledge to interpret them. With the careful experiments conducted in this thesis, the number of replacements occurring within a given time period is known so an estimate of the average time between replacement events is known and also the total population size. This gives a richer interpretation of the parameters that are delivered by calibrating the SDE model. Thus discrete event time is related to real time in a marginally more intuitive way.

The model makes the following assumptions. In a community saturated with N_T individuals, an individual must die or leave the system for the assemblage to change. The dead individual is replaced by an immigrant from a source community with probability m , or by reproduction by a member of the local community with probability $(1 - m)$. Now let a be the average time between individual replacement events in the population and define a scaled time $\tau = t/a$. Then in a time period $\Delta\tau = 1$, one can expect a single replacement in the community as a whole. For a particular species within the community, say the i^{th} species, comprising N individuals, the probabilities of an increase by one, no change and a decrease by one individual were explained in the literature review but are repeated here for completeness:

$$\Pr(N + 1/N) = \left(\frac{N_T - N}{N_T}\right) \left[mp + (1 + \alpha)(1 - m) \left(\frac{N}{N_T - 1}\right) \right] = b_N \quad (7:1)$$

$$\Pr(N/N) = \frac{N}{N_T} \left[mp + (1-m) \left(\frac{N-1}{N_T-1} \right) \right] + \left(\frac{N_T-N}{N_T} \right) \left[m(1-p) + (1-m) \left(\frac{N_T-N-1}{N_T-1} \right) \right] \quad (7:2)$$

$$\Pr(N-1/N) = \frac{N}{N_T} \left[m(1-p) + (1-\alpha)(1-m) \left(\frac{N_T-N}{N_T-1} \right) \right] = d_N \quad (7:3)$$

where p is the relative abundance of the i th species in the source community and α is an advantage parameter.

In general, if N_T is large enough to consider $X = \frac{N}{N_T}$ to be a continuous variable, then the Markov Chain process can be described by the SDE,

$$dX = M d\tau + \sqrt{V} dW_\tau \quad (7:4)$$

where W_τ is a standard Wiener process and for small $\Delta\tau$,

$$E(\Delta X | X(\tau) = x) = M \Delta\tau, \quad (7:5)$$

and

$$E((\Delta X)^2 | X(\tau) = x) = V \Delta\tau \quad (7:6)$$

For the discrete time model then, the smallest change that can occur is the replacement of one individual, which is expected to occur in a time $\Delta\tau = 1$.

So during this time period, the possible changes in X are, $\Delta X = \pm \frac{1}{N_T}, 0$.

Consequently, for $x = \frac{N}{N_T}$, we have,

$$\begin{aligned} E(\Delta X | X(\tau) = x) &= \frac{1}{N_T} (b_n - d_n) + 0 \\ &= \frac{1}{N_T} (b_n - d_n) \Delta \tau \\ &= \frac{1}{N_T} (m(p - x) + 2\alpha(1 - m)x(1 - x)) \Delta \tau \quad (7:7) \end{aligned}$$

where b_n is equation (7:1) and d_n is equation (7:3) and

$$E((\Delta X)^2 | X(\tau) = x) = \frac{1}{N_T^2} (2(x(1 - x)) + (m(1 - 2x)(p - x)) \Delta \tau \quad (7:8)$$

which gives us M and V . In Sloan et al. (Sloan, 2006), it is argued that the second term in V is negligible and it is reasonable to ignore it. Thus equation (7:4) becomes

$$\begin{aligned} dX &= \frac{1}{N_T} (m(p - X) + 2\alpha(1 - m)X(1 - X)) d\tau \\ &\quad + \sqrt{\frac{1}{N_T^2} (2X(1 - X))} dW_\tau. \quad (7:9) \end{aligned}$$

Now we can convert to real time using $d\tau = dt/\alpha$ so that,

$$dX = \frac{1}{aN_T} (m(p - X) + 2\alpha(1 - m)X(1 - X))dt \quad (7:10)$$

$$+ \frac{1}{\sqrt{aN_T^2}} \sqrt{(2X(1 - X))}dW_t.$$

7.2 Model Calibration

For the purposes of calibration, one can conveniently map equation (7:10) onto a simple linear model. Note that, in contrast to the paper by Ofiteru *et al.* (2010), the p values that represent the relative abundance in the source community are actually measured here, so

$$dX = m_1 Y_1 + m_2 Y_2 + \varepsilon \quad (7:11)$$

where, $Y_1 = p - X$, $Y_2 = X(1 - X)$, $m_1 = \frac{m dt}{N_T a}$, $m_2 = \frac{2\alpha(1-m)dt}{N_T a}$ and $\varepsilon = \frac{1}{\sqrt{aN_T^2}} \sqrt{(2X(1 - X))}dW_t$. While dW_t is normally distributed, $N(0,1)$, ε is not.

However, equation (7:11) gives us a straightforward method of calibrating the unknown parameters. For the purposes of this study, we use a weighted least square regression to calibrate the unknown parameters since the variances of the experimental data set are not constant. It reflects the behavior of the random error of the model and differs from ordinary least square regression in that it uses weights to estimate the regression coefficients (Montgomery et al; 2012). Performing a weighted least squares regression analysis, using

observations of the dependent dX and independent variable in which the weights are $2(X(1 - X))^{-1}$, gives estimates of the parameters m_1 and m_2 . The weighted errors should be normally distributed and hence the standard residual error produced by the least squares analysis should be $\sqrt{\frac{1}{aN_T^2}}$. Thus, the entire original model parameters in equation (7:10) can be retrieved from a weighted least squares analysis.

7.3 Simulating generic behaviour of the system and validating the calibration method

In the literature review the generic stochastic process for neutral community dynamics, as introduced by Hubbell (2001) and subsequently extended by Sloan et al. (2006), was introduced. Here a slightly modified version is presented that would allow for a more intuitive interpretation of any parameters that are ultimately calibrated using experimental data. However, neither in the literature review nor in the mathematical development above have I demonstrated what stochastic dynamics might look like in microbial population under the variety of experimental setups that were undertaken. In assessing the experimental design, a deterministic model was deployed, but that clearly doesn't incorporate the stochastic demography that, throughout this thesis, have been suggested might be critical in shaping community structure. So prior to conducting the calibration method outlined above, the results of simulations are presented.

To perform the simulation, the Euler-Maruyama method is used to give numerical realisations of the solution of the continuous differential equation (7:10). Thus if we let the sequence of random variables,

$$Y_1, Y_2, Y_3, \dots, Y_N \quad (7:12)$$

be approximations of the solution to equation (7:10) on the time interval $[0, T]$ at equally spaced discrete time points,

$$0 < \tau_1 < \tau_2 < \tau_3 < \dots < \tau_N = T \quad (7:13)$$

where, $\tau_{i+1} - \tau_i = \Delta T$ is a constant. Then we recursively deploy,

$$Y_{i+1} = Y_i + \frac{1}{aN_T} (m(p - Y_i) + 2\alpha(1 - m)Y_i(1 - Y_i))dt + \frac{1}{\sqrt{N_T^2 a}} \sqrt{Y_i(1 - Y_i)} \Delta W_{\tau_{i+1}} \quad (7:14)$$

where,

$$\Delta W_{\tau_{i+1}} = W_{\tau_{i+1}} - W_{\tau_i},$$

and thus the random variables $\Delta W_{\tau_{i+1}}$ are independent normal random variables with expected value 0 and variance ΔT . This algorithm was encoded in MATLAB.

Using this approach, one can simulate some of the generic dynamics associated with the stochastic differential equation. This has been done for a population size, $N_T = 1000$, which would be considered very small indeed for microbes in a chemostat. However, using a small population for the purposes of demonstrations helps to ensure that stochastic dynamics occurring over relatively short timescales can be seen. The replacement rate, $a = 0.00833$, corresponds to a replacement occurring approximately every 4 seconds which yields a fairly rapid turnover such that the approximate population doubling time is just over 1 hour. In the experimental set up, large pure cultures of the two strains of bacteria were maintained and they enter the small chemostats in equal proportion where their mixing then occurs, thus in the simulations shown here immigration is occurring at $p_i = 0.5$.

7.3.1 Neutral model with Migration

For the first demonstration simulation, realizations of the dynamics of two species that are behaving in accordance with Hubbell's original model are shown. Hubbell's theory (2001), and the earlier theory of Island Biogeography by McArthur and Wilson, are founded on the idea that in an open insular community which is linked to the outside world the rate of immigration dictates how similar the insular community is to that of the outside world source community. High immigration will yield local communities that strongly resemble the outside and low immigration will give local communities whose structure might be quite different (McArthur & Wilson, 1956). In Hubbell's

model, which is the basis of the SDE employed here, the dynamics are neutral and there is no speciation occurring in the local community, so the dissimilarity in community structure between local and source community is manifest as more variability in the abundance of species in the local community.

Figures 7-1 – 7-3 give the realization of the dynamics simulated by the model with different values of m , the probability that a death event is replaced by an immigrant. In Figure 7-1, $m = 0.1$, it is 0.4 in Figure 7-2 and 0.8 in Figure 7-3 respectively. The number of individuals (N_T) and mean replacement time (a) were the same in all simulations. In addition, at the start of the simulation, the relative abundance of the two strains was assigned to be 0.5. With the probability of the source abundance of the i^{th} species (P_i) set to 0.5, it can be seen that a mean of 0.5 resulted in all cases as expected. This reflects the fact that in equation (7:10) the immigration term premultiplies $p_i - x$ and so when x is greater than p_i , the first term is negative and effectively pulls the abundance back towards p_i and conversely when x is less than p_i , the first term is positive and draws the abundance up towards p_i . So the first term always attempts to bring the local population of species i to the abundance it has in the source community. Superimposed on this are the random walk, governed by the Wiener Process in the second term of equation (7:10). The relative importance of the deterministic pull towards p_i and the random-walk process are determined by the magnitude of m . Thus, when m is increased in Figure 7-2 and in Figure 7-3, the magnitude of the random fluctuations about p_i decrease. Indeed if histograms of the abundances that are realized in a

long simulation (3000 hours) are plotted, an approximation to the stationary abundance distribution is obtained. (See Figure 7-1(b) - Figure 7-3(b)). It is readily apparent that the mean of the distribution is always 0.5 but the variance decreases with increasing immigration probability. In this case the distribution looks normal, however, it was shown in (Sloan *et al.*, 2006) that the abundance is in fact beta distributed. This simulation demonstrates the increased uncertainty in the abundance of local populations that comes with being increasingly isolated.

i) Case 1: $m = 0.1$

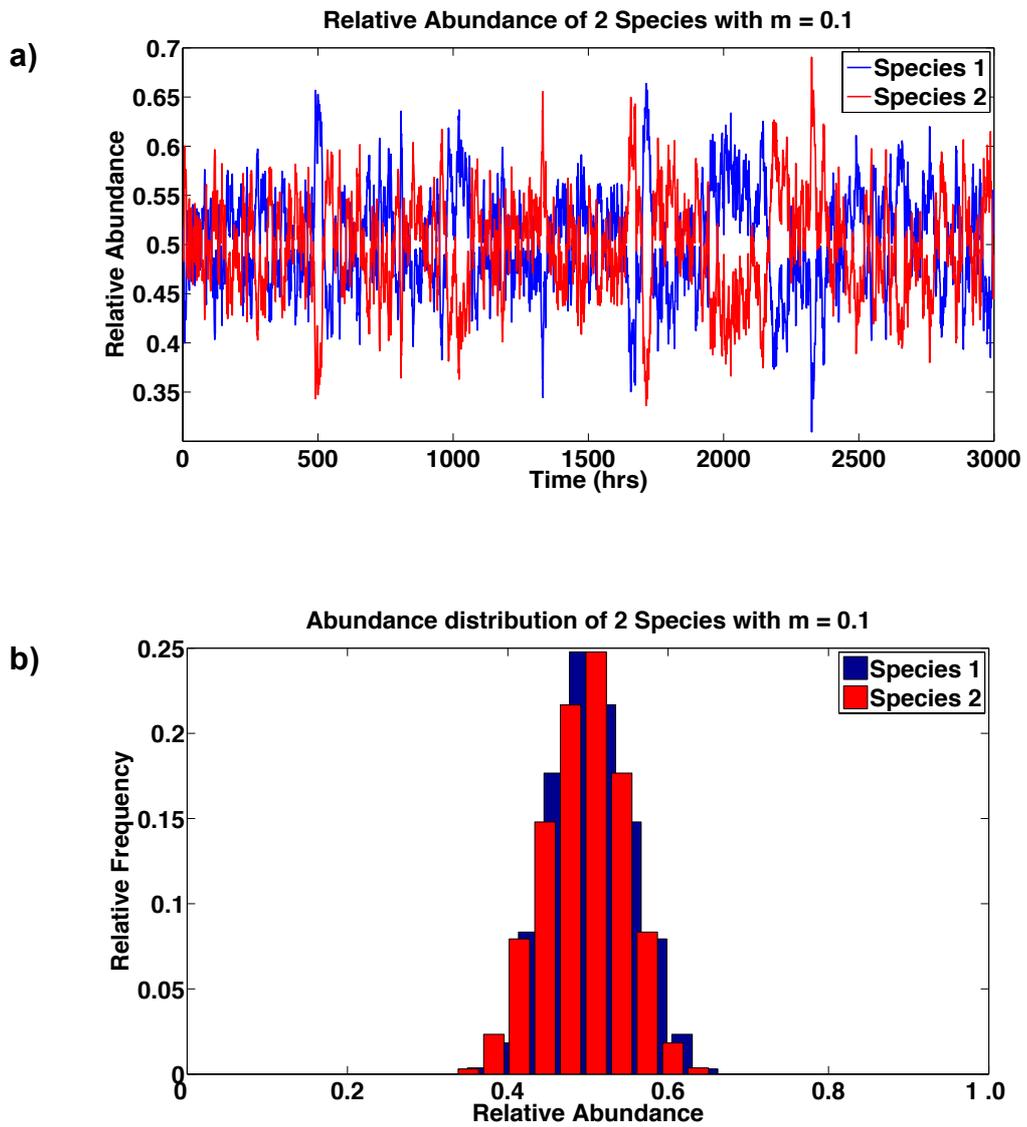


Figure 7-1: Simulation results from the stochastic model with 0.1 as migration rate. Graph a) shows the relative abundance of both species as the simulation runs while b) shows the relative occurrence of each species at specified relative abundances.

ii) Case 2: $m = 0.4$

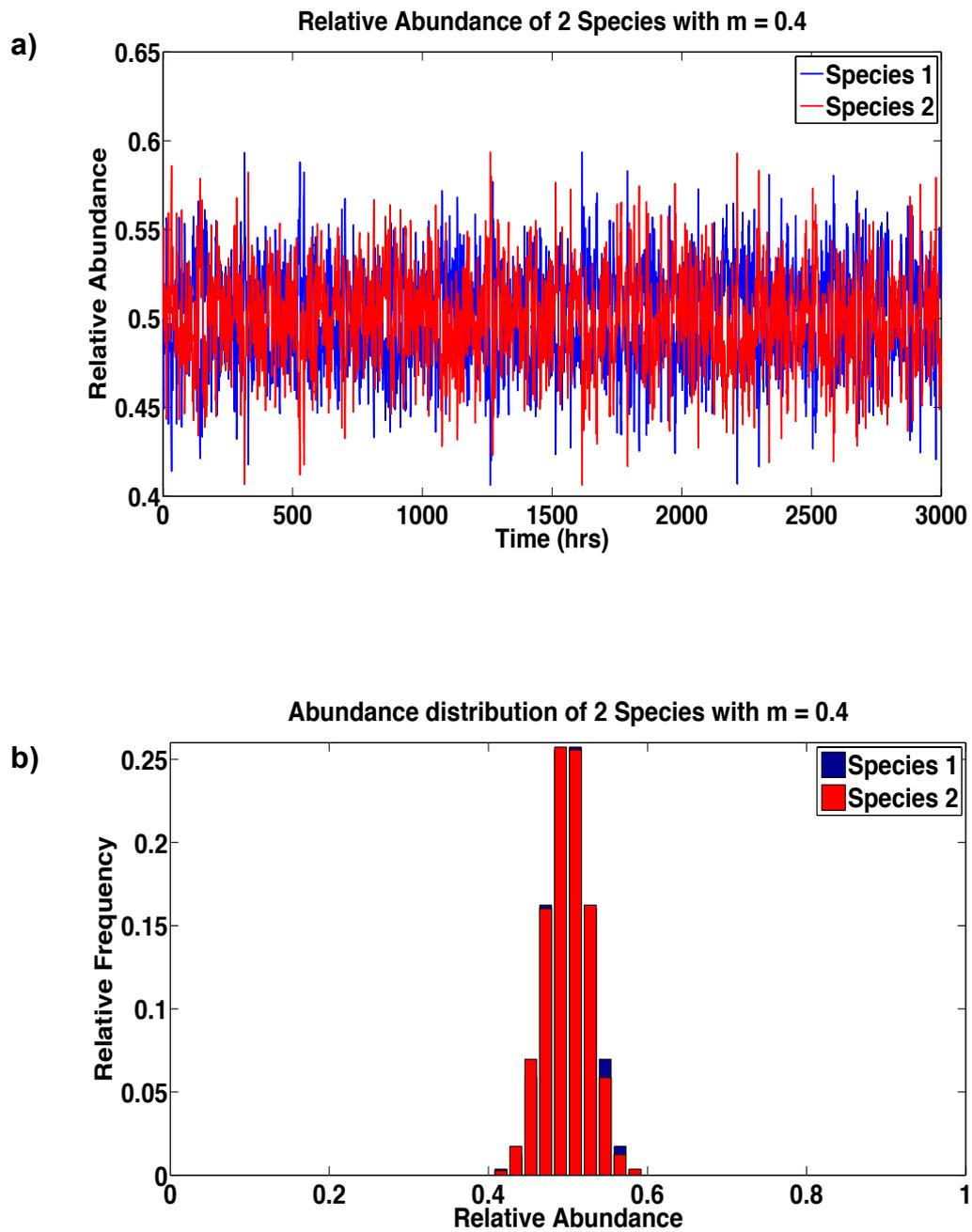
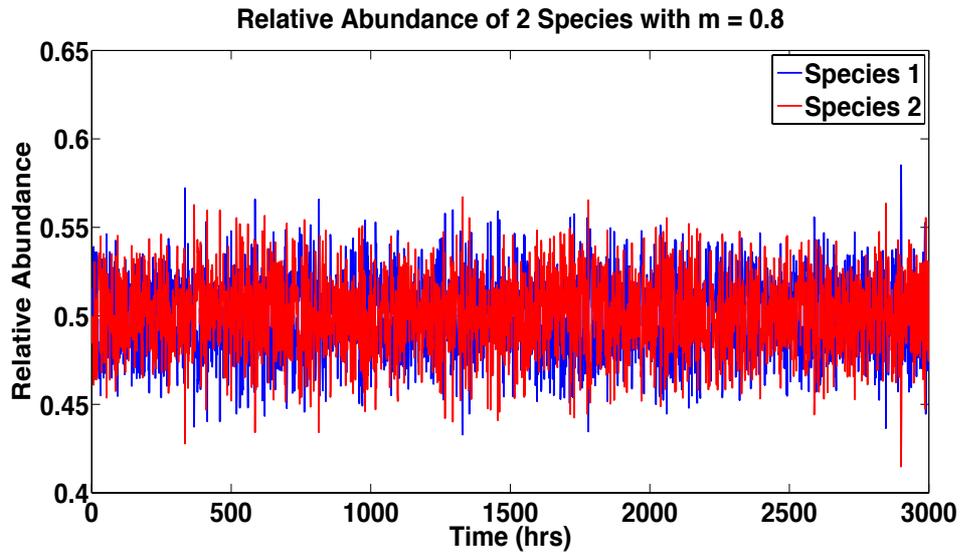


Figure 7-2: Simulation results from the stochastic model with 0.4 as migration rate. Graph a) shows the relative abundance of both species as the simulation runs while b) shows the relative occurrence of each species at specified relative abundances.

iii) Case 3: $m = 0.8$

a)



b)

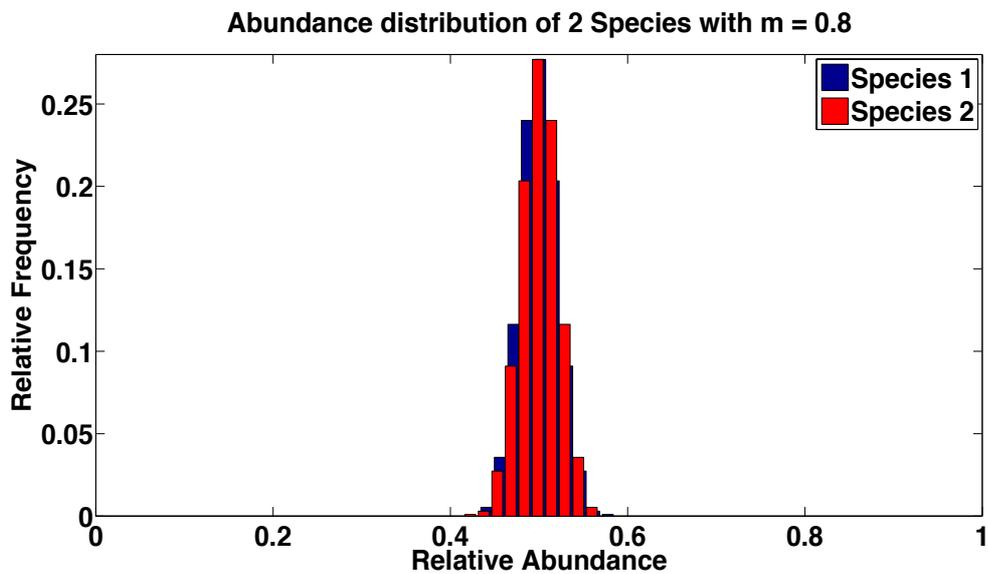


Figure 7-3: Simulation results from the stochastic model with 0.8 as migration rate. Graph a) shows the relative abundance of both species as the simulation runs while b) shows the relative occurrence of each species at specified relative abundances.

7.3.2 Non-Neutral model with immigration

The model presented here differs from Hubbell's original in its ability to retain stochastic dynamics whilst representing bacteria with different specific growth rates. When an advantage is conveyed on one species then automatically this results in a disadvantage for the other species. In a closed-deterministic system, the bacterial species with the advantage would win out and become mono dominant but in an open system, then immigration can maintain the population of the disadvantaged species at values greater than 1. In Figure 7-4, one of the species has been given a significantly higher probability of birth than the other ($\alpha = 0.75$), the two species start with the same relative abundance and have the same relative abundance in the source community. In the simulation the species with the advantage rapidly dominates the system but does not completely take over. The variance in abundance of the taxa is significantly lower than for the completely neutral scenario with the same immigration rate ($m = 0.1$) that was presented in Figure 7-1. If one considers the second term in Equation (7:10) then this decrease in variability is inevitable when the abundance X is close to zero or one.

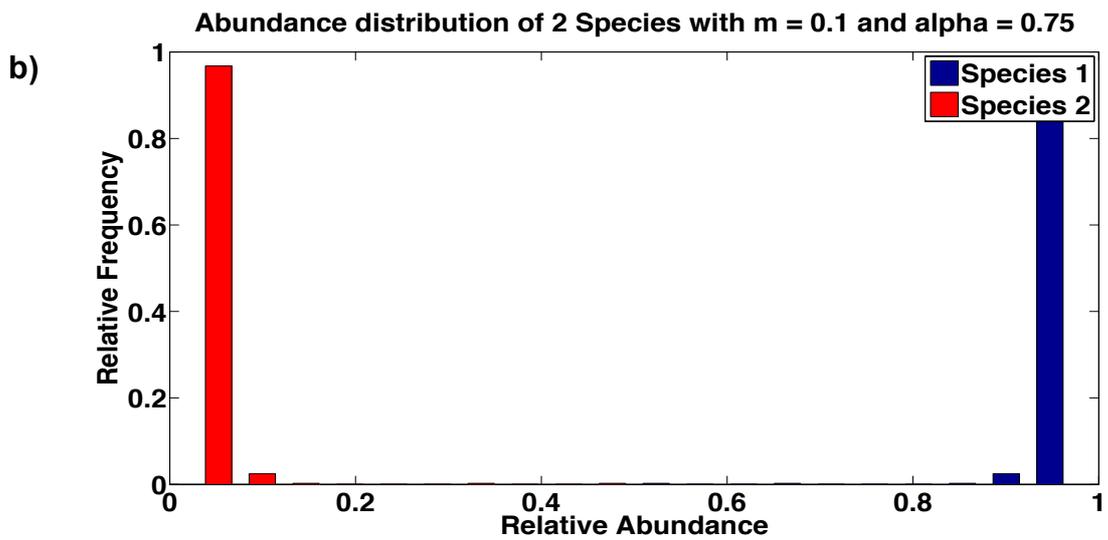
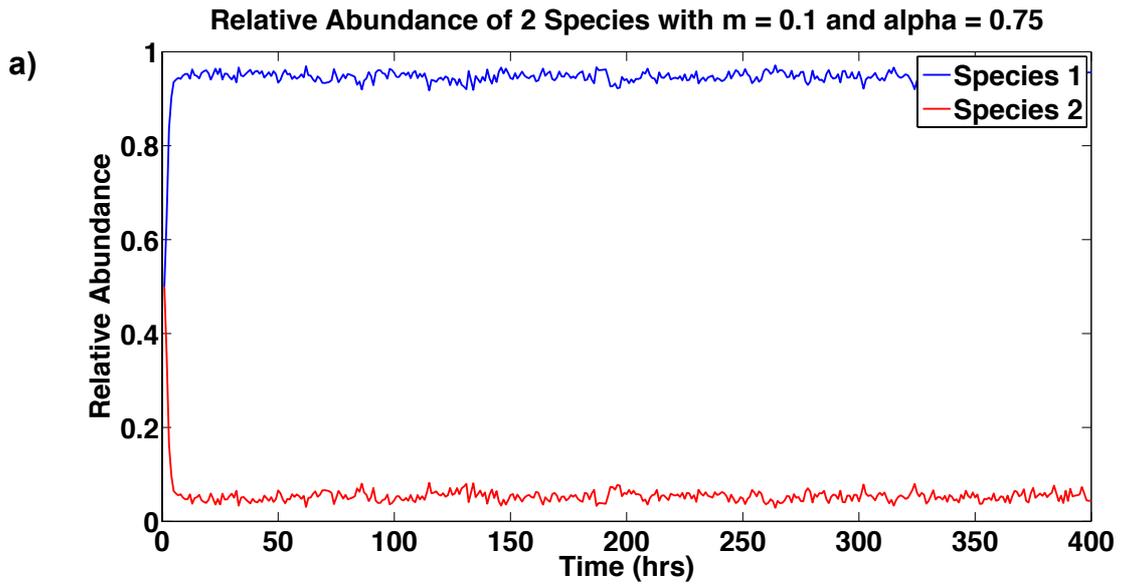


Figure 7-4: Simulation results from the stochastic model with 0.75 selective advantage and no migration into the community. Graph a) shows the relative abundance of both species as the simulation runs while b) shows the relative occurrence of species at specified relative abundances.

7.3.3 Completely insular neutral community ($m=0$)

One interesting scenario to consider is when the local community becomes completely cut off from the outside world and the species behave neutrally; such a scenario has been suggested to prevail in certain communities where ecological advantage between the species is weak and migration non-existent (Nemergut, 2013). Figure 7-5 shows two different realizations of the dynamics simulated by the SDE when $m = 0$. In one simulation, the red species become dominant and in the other the blue one becomes dominant. This occurs purely by chance. The histograms of residuals (Figure 7-5b and c) are for the abundances of both species attained when one realization of the dynamics tend toward one species being mono dominant. If an average over many realisations is taken, then the stationary distribution for both species would have two bars with relative frequency 0.5; one at 0 and the other at 1, indicating that there is a 50% chance of the species either becoming mono dominant or going locally extinct. This behaviour contrasts dramatically with the deterministic model where the abundance of both species would tend to remain constant at their initial value ad-infinitum. A comparison between the graphs of drift (Figure 7-5) and selective advantage (Figure 7-4) show that drift can have a similar catastrophic effect on a community composition when selective advantage is weak or non-existent between species that are ecologically similar.

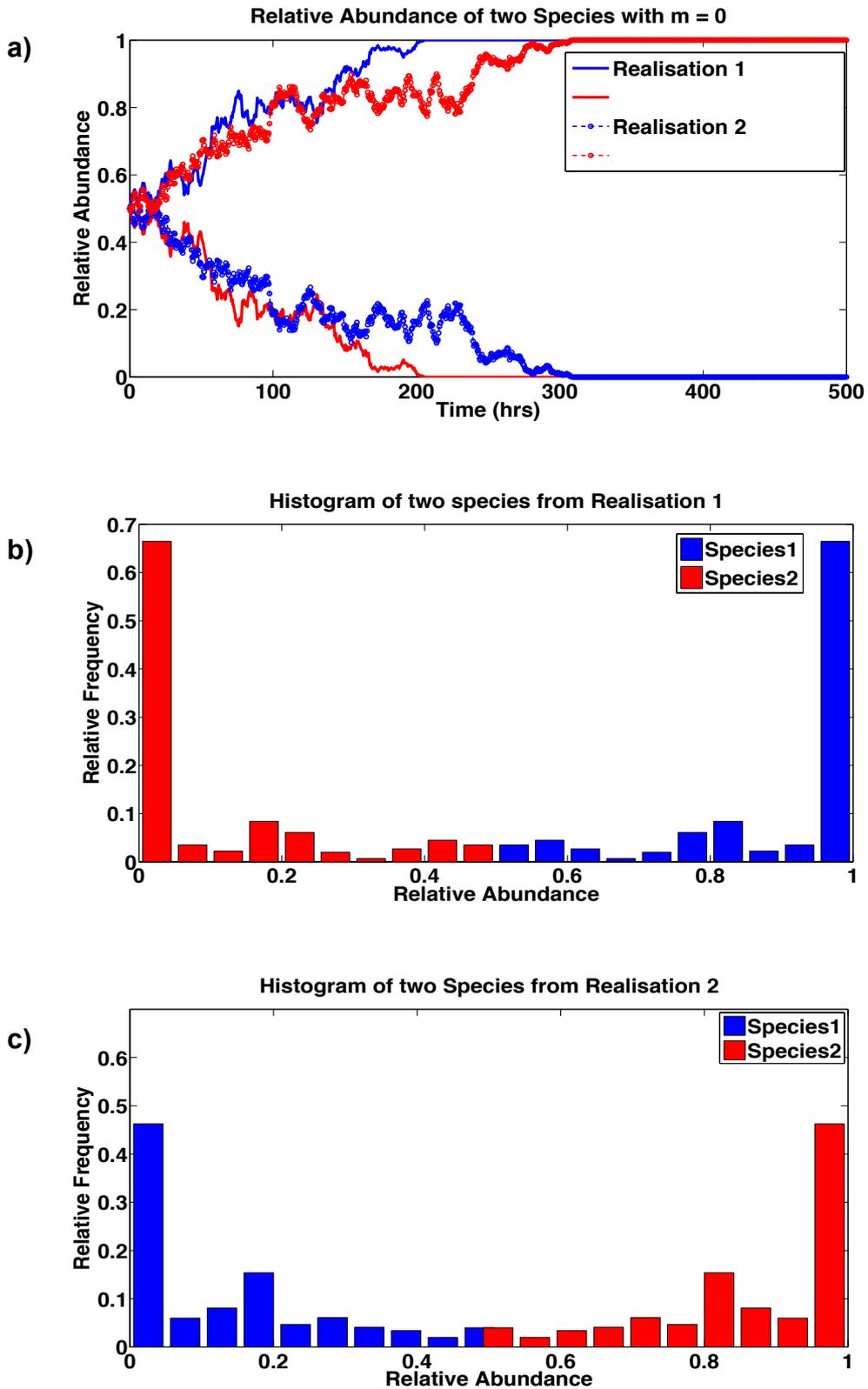


Figure 7-5: Results from the stochastic model when drift alone occurs (no migration rate and selective advantage). Graph a represents two realizations, showing that each species can be the dominant one and this is solely based on chance, while graphs b & c show the relative occurrence of each species at specified relative abundances for both realizations.

7.3.4 Testing the Calibration Technique: Retrieving Process Parameters

Ultimately the goal is to be able to calibrate the model parameters and quantify the goodness of fit using data from the experimental chemostats. To do this, it is first necessary to test whether the calibration method works. The simulations shown in section 7.3 above were conducted with the specific aim of demonstrating generic properties of the model simulations under different combinations of parameters. They don't map directly onto the scenarios in the experimental setup; although elements of these generic behaviors will be apparent in the experimental results. To proceed with testing the calibration procedure, a second set of simulations were conducted again for a week long period, but in this case the parameters were selected to at least notionally map onto the experimental setup. Of course, at this stage the exact values of the parameters are not known rather randomly selected parameters that will yield qualitatively similar time series of abundance as are observed in the experiments are used. The aim here is to ascertain whether or not the correct parameter values can be retrieved when the calibration method is applied to the simulated data. Table 7-1 shows the parameters used to conduct the simulations for the three scenarios. Note that where parameters m_1 or m_2 are not listed then they have been set to zero. The table also gives the parameter value that is calibrated, its upper and lower confidence limits at the 95% level and a p-value for the parameters. Here N_T is arbitrarily chosen to be 10^4 . Note that the values of a are of the order 10^{-5} , which may seem small, but the units are days^{-1} thus the values corresponds to a replacement event occurring approximately every second. The calibration algorithm clearly does a good job of retrieving the parameters from the simulated data. The confidence limits

are narrow and the p-values indicate that the estimates are highly significant. For this stochastic model, quoting for example, R-Squared as a measure of goodness of fit would be meaningless and could indeed potentially mislead. Capturing the variance is an integral part of the stochastic model and the R-squared could be very low but provided the residuals capture those imposed during the simulation then a good fit is attained. Now since the value of a is inversely proportional to the standard error in the weighted residuals, if through the calibration procedure, the value of a used in the simulation can be retrieved then this indicates that the variance that was originally imposed has been captured. Thus it is reassuring that very similar values of a can be retrieved by the calibration.

Table 7-1: Table showing simulation results from the calibration models

	Stochastic Models	Process Variable	Assigned Value	Output Value	LCL	UCL	pValue
1	Neutral dynamics with immigration	m_1	0.4167	0.4232	0.3941	0.4525	1.3784e-25
		a	1200	1118.4			
2	Non-neutral dynamics in an insular community	m_2	0.0588	0.0603	0.0530	0.0676	6.3907e-52
		a	1700	1748.4			
3	Non-neutral dynamics with immigration	m_1	0.5	0.5051	0.4795	0.5306	3.2464e-28
		m_2	0.17	0.1681	0.1592	0.1769	6.6555e-26
		a	1200	1194.1			

It appears, therefore, that the weighted least square analysis is an appropriate calibration method.

7.4 Calibration of Chemostat Experiments

Having shown that the calibration method can accurately retrieve the parameters used in the simulations, I proceed with confidence to calibrate the three chemostat experiments described in Chapter 6. This is done by inputting the time series data obtained experimentally into the corresponding calibration models. After obtaining the parameters, they are used to undertake a simulation to see if the realization of the stochastic process, at least qualitatively, resembles each corresponding set of experimental data. The implications of the parameters obtained are then discussed.

7.4.1 Neutral dynamics with immigration

The first experiment that was conducted in the chemostat involved no antibiotics and so the strains had the same specific growth rate. The concentration of bacteria in the large source community reactors was monitored and so there is no need to calibrate the abundance in the source community. The immigrants of both strains entered the local community in roughly equal abundances. The calibration results are given in Table 7.2.

Table 7-2: Resulting process parameters from the ‘Neutral dynamics with immigration’ calibration model.

Stochastic Models	Process Variable	Output Value	LCL	UCL	pValue
Neutral dynamics with immigration	m_1	0.4956	0.3831	0.6080	6.2701e-16
	Standard Error	0.0886			

It is clear from the table that the parameter value m_1 obtained from the calibration is very highly significant with a p value = $6.2701e-16$. The standard error is for the weighed residuals. The model for the errors is an important component of the stochastic model; indeed it characterizes the stochastic component of the dynamics. Thus, again merely quoting, for example, an R-squared value tells us little about the goodness of fit of the model; there will inevitably be a high degree of variability and hence a low R-squared value. Similarly the actual residuals in the model are not normally distributed, it is only the weighed residuals that are expected to be normally distributed. So instead, as demonstrated in Figure 7-6, we check that the distribution of residuals obtained from the real experimental data looks very similar to the distribution that is obtained from simulating a long time series of the dynamics (4000 hours).

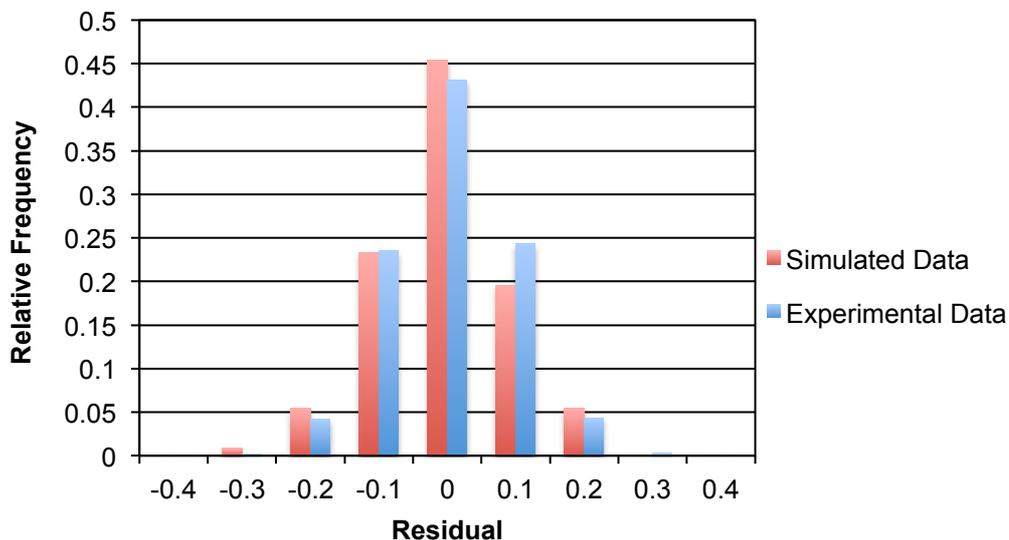


Figure 7-6: Histogram of residuals for the neutral model with immigration from both the experimental chemostats (blue) and a long simulated time series (red).

From the histogram shown in Figure 7-6, it can be seen that the distribution of the residuals is similar for the experimental data and for a long time series of abundances simulated using the fitted parameters. The similarity suggests that the stochastic process captures the variability in the real data. This is further demonstrated by presenting the simulated data for a period of 200 hours in Figure 7-7; subjectively the dynamics look similar to those presented for the neutral dynamics with immigration chemostat experiment in Chapter 6. The biological relevance of the parameters calibrated and their correspondence with the biological and physical conditions in experimental chemostat will be discussed later, once the other calibrations have been performed.

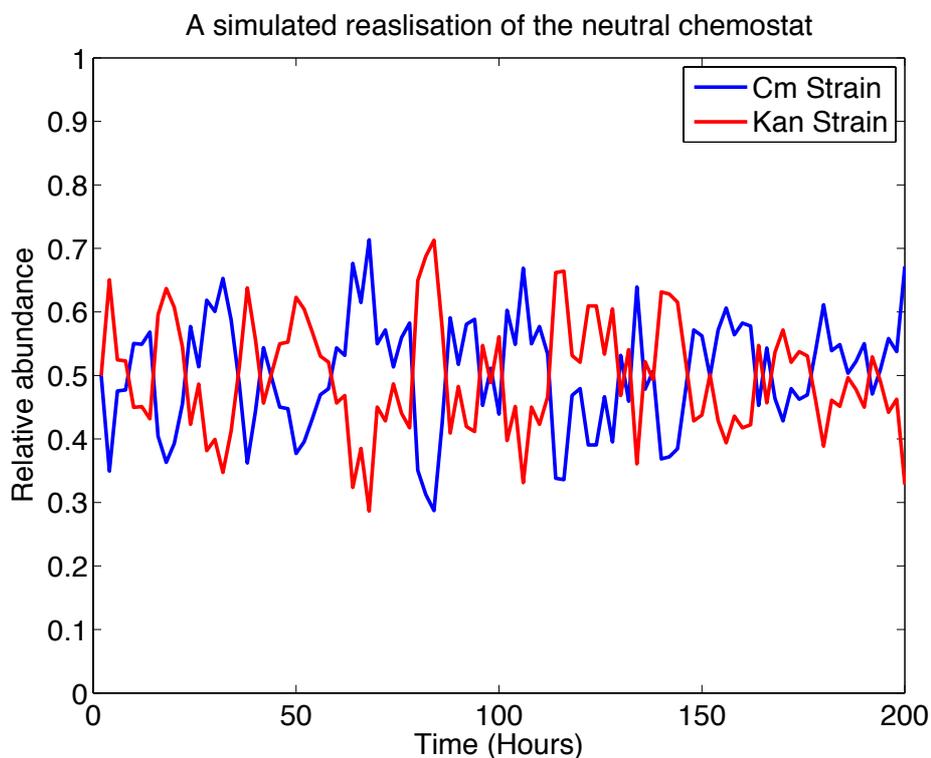


Figure 7-7: Graphical result of simulation of the dynamics of both species in the local community modeled using the stochastic process with calibrated parameters.

7.4.2 Non-neutral dynamics in a completely insular community

In the second chemostat experiment, immigrants were not allowed to enter the local community. Instead, each of the ten small reactors was inoculated with nutrient along with a low concentration of 0.5µg/ml chloramphenicol and equal concentrations of the two *E.coli* strains. As expected and clearly seen, one species wins out, although there were fluctuations in the abundance on route to mono dominance that suggested stochastic demography might play a role. The calibrated model parameters are given in Table 7-3. Here the parameter that is related to the selective advantage is only significant at the 90% level. Because the stationary distributions are either for mono-dominance or complete extinction depending on the strain, in-the-long-run the variance will diminish to zero. Thus to demonstrate that the calibrated parameters deliver dynamics that are consistent with the observed dynamics, an ensemble of stochastic simulations are run and recovered and their mean and the mean +/- 2 x Standard deviations plotted. The observed time series of abundances for the experiments are then plotted in Figure 7-8, which shows that they fall within two standard deviations of the predicted mean.

Table 7-3: Resulting process parameters from the ‘Non-neutral dynamics in an insular community’ calibration model.

Stochastic Models	Process Variable	Output Value	LCL	UCL	pValue
Non-neutral dynamics in an insular community	m_2	0.0902	-0.0105	0.1907	0.0788961
	Standard Error	0.1176			

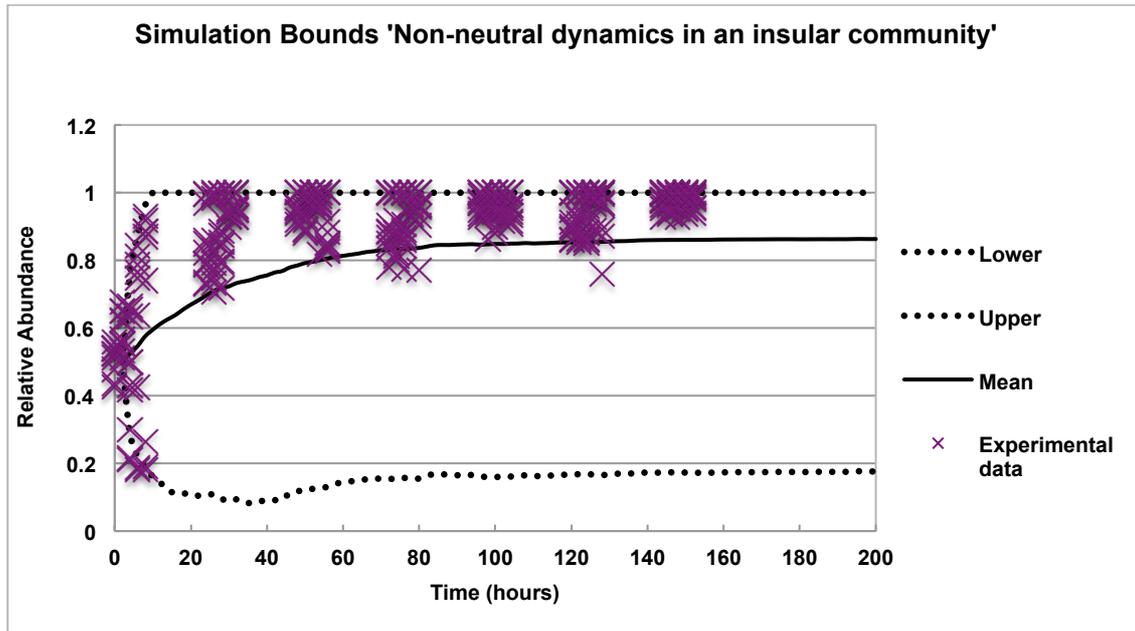


Figure 7-8: Graph showing that the experimental data from the 'Non-neutral dynamics in a completely insular community' experiment fall within the upper and lower limits set by the standard deviations of the calibrated mean.

In Figure 7-8 we note that for the simulations, two times the standard deviation does yield wide confidence limits, which highlights the fact that for this scenario, capturing the stochastic demography is vital in characterizing the dynamics of the system. A single stochastic realization using the simulated parameters is shown in Figure 7-9. After approximately 50 hours, the disadvantaged strain – Kan+ has gone extinct. By chance, even given its selective advantage, the chloramphenicol resistant strain can also sometimes be the one that goes extinct. The experimental data do fall within two standard deviations of the mean.

Stochastic Modeling of Non-neutral dynamics in an insular community

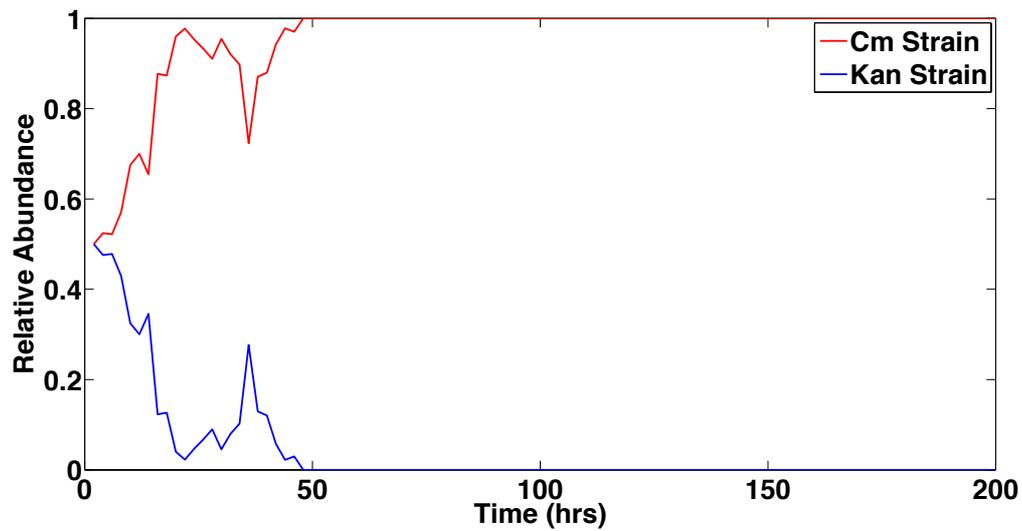


Figure 7-9: Graphs showing the stochastic modeled dynamics of both species in the local community for the ‘Non-neutral dynamics in a completely insular community’ experiment.

7.4.3 Non-neutral dynamics with immigration

The last chemostat experiment involved immigration of both strains from their respective source communities, where the populations were maintained at very similar abundances, to ten replicate local communities (chemostats) to which nutrient with $0.5\mu\text{g/ml}$ chloramphenicol was added. The deterministic modeling suggested that the selective advantage was insufficient for a single taxon to completely dominate; instead based on that modeling, the chloramphenicol resistant strain was expected to have a relative abundance of about 0.55. In the experiments, the mean relative abundance of the Cm+ strain was approximately 0.67. Here I calibrate the SDE on the data from all ten time series of the relative abundance of the chloramphenicol resistant

strain. There are two calibrated model parameters for this experiment - m_1 and m_2 for the drift and selective advantage terms respectively. The calibrated parameters are shown in Table 7-4 below. The p values of both m_1 and m_2 indicate that the calibrated parameters are again highly significant. The variance is expected to be high due to the stochastic and selective properties of the system so reporting the R-squared value or showing the histogram of residuals would lend little significance and goodness of fit to the data, thus as in section 7.3.2, multiple stochastic simulations are run and the results show that the experiment data lie within the upper and lower limits of the calibrated mean which is the mean $\pm 2 \times$ Standard deviation. This data is shown in Figure 7-10. Note that the mean of the simulated data stabilizes at 0.67, which is the same as for the experimental results.

Table 7-4: Resulting process parameters from the ‘Non-neutral dynamics with immigration’ calibration model.

Stochastic Models	Process Variable	Output Value	LCL	UCL	pValue
Non-neutral dynamics with immigration	m_1	0.3454	0.2795	0.4114	4.0617e-21
	m_2	0.3038	0.2320	0.3757	4.8783e-15
	Sta. Error	0.0869			

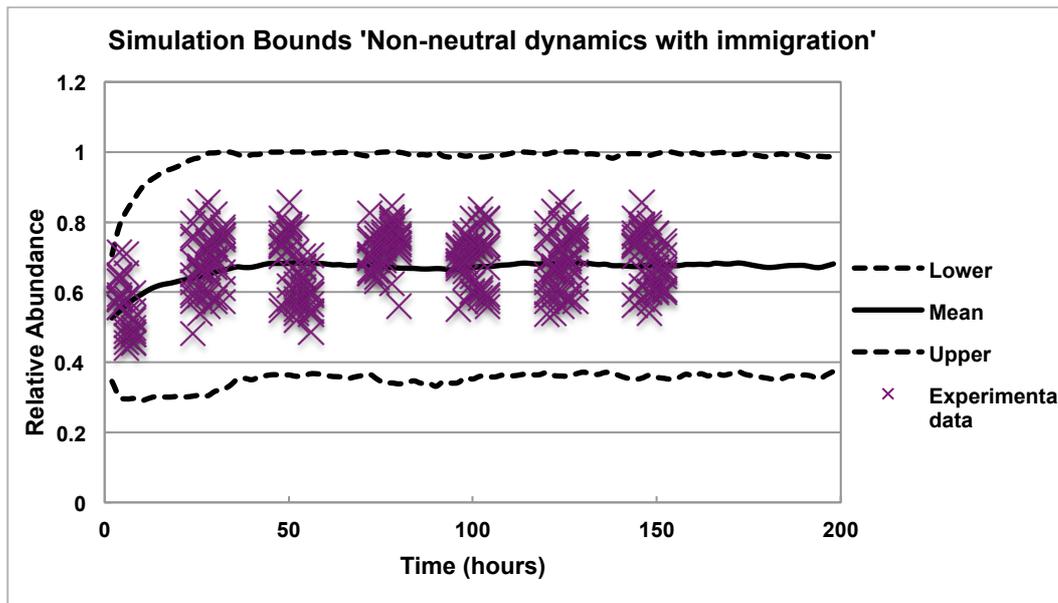


Figure 7-10: Graph showing that the experimental data from the ‘Non-neutral dynamics with immigration’ experiment fall within the upper and lower limits set by the standard deviations of the calibrated mean.

From Figure 7-10, it is clear that the experimental data lie within the upper and lower limits set by two times the standard deviations of the mean. Again the confidence limits are wide, suggesting that stochastic demography is important. A single stochastic realization using the calibrated parameters is shown in Figure 7-11. Also the summaries of all calibration parameters along with the standard error for each chemostat experiment are listed in Table 7-5 below. As before, the biological implications of the calibrated parameters are discussed in the next chapter.

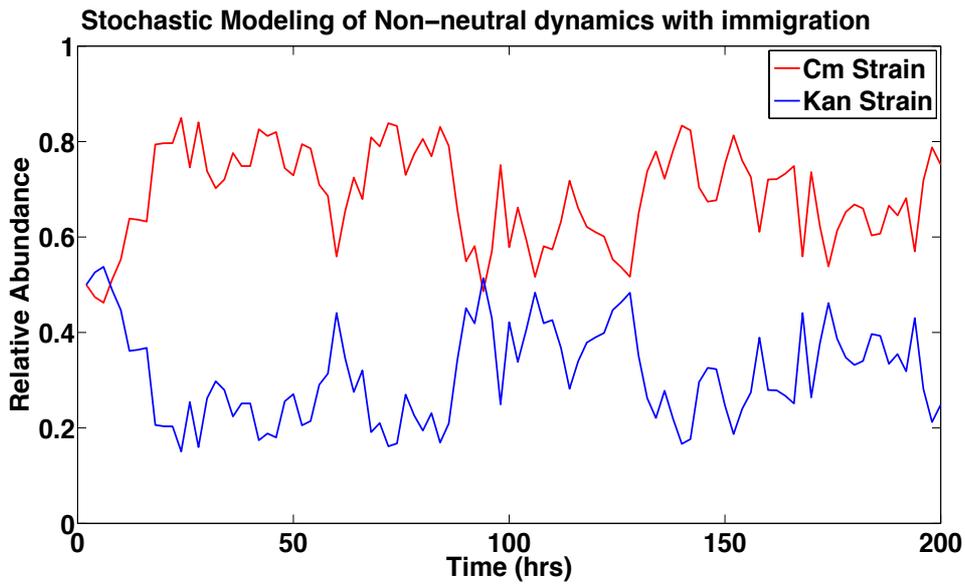


Figure 7-11: Graph showing the stochastic modeled dynamics of both species in the local community for the ‘Non-neutral dynamics with migration’ experiment.

Table 7-5: Summary of fitted model parameters for each chemostat experiment using corresponding calibration models.

	Stochastic Models	Process Variable	Output Value	LCL	UCL	pValue
1.	Neutral dynamics with immigration	m_1	0.4956	0.3831	0.6080	6.2701e-16
		Sta. Error	0.0886			
2.	Non-neutral in insular community	m_2	0.0902	-0.0105	0.1907	0.0788961
		Sta. Error	0.1176			
3.	Non-neutral dynamics with immigration	m_1	0.3454	0.2795	0.4114	4.0617e-21
		m_2	0.3038	0.2320	0.3757	4.8783e-15
		Sta. Error	0.0869			

8 Discussion and Conclusion

The main aim of this thesis was to determine whether or not demographic stochasticity could be observed in highly controlled experiments with very well defined microbial communities. The motivation was the growing body of literature that suggested that much of the variance in community composition samples taken from a suite of similar environments could be explained by neutral community assembly models (Woodcock *et al.*, 2007; Sloan *et al.*, 2006). With the exception of one previous study, the models have been applied to explain taxa abundance distributions in single samples from multiple locations. Thus the first fitting routines, developed by, for example Volkov and Etienne fitted a theoretical approximation to taxa abundance distribution to data from single sites (Volkov *et al.* 2003; Etienne *et al.*, 2005). Whilst this gave circumstantial evidence of neutral assembly, it has been pointed out that many parameter combinations could give very similar quality of fits (Ofiteru *et al.* 2010; Harte, 2003). Thus, researchers subsequently sought to use abundance data from multiple sites. Sloan *et al.* (Sloan, 2006) devised a method that was specifically tailored for microbes that did not rely on any assumptions about the mechanisms in the source community. Subsequently, Harris (Harris *et al.*, 2015) used the result that the multispecies steady-state distribution of abundances is a Dirchlet distribution to devise a Bayesian method of calibrating the neutral model, which is now state-of-the-art in fitting such models. However even with these fitting routines some of the parameters are compound. Thus, for example, in all methods for fitting

stationary data the parameters N_T and m are never explicitly estimated, instead, their product $N_T m$ is fitted. This can imply that the models fit but for the wrong reasons; it could be that a range of different processes could yield the same stationary distribution. Thus, a recognition has emerged that if one is to truly determine whether stochastic demography (and neutral dynamics as a subset of this) is critical then one needs to actually observe the dynamics and show that they do actually adhere to model stochastic dynamics (Etienne *et al.*, 2005; Ofiteru *et al.*, 2010). That was the purpose of this thesis. In addressing this, I hope to also address some of the other controversies, most notably the observation that neutral dynamics would be slow (Nee, 2005); potentially too slow to be observed in any very large microbial populations.

The careful design of organisms and experiments that have been described in this thesis gave strains of *E.coli* that were truly neutral in the absence of antibiotic and could be tuned to exhibit different growth kinetics. Thus if neutral dynamics does occur then it should be seen in the abundances of the two strains housed in a chemostat. The fact that 10 identical chemostats were used provides the statistical power to calibrate the stochastic differential equation that is based on a simple birth-death-immigration process. In all three experiments, the SDE was successfully calibrated.

Thus from the successful calibration of the stochastic model and, in particular, the requirement to weight the residuals in the least squares analysis, it superficially looks like the experimental strains of *E.coli* are exhibiting the effects of demographic-stochasticity in their population dynamics. So for

example, in the neutral dynamics experiments, it would appear that this is the first time neutral dynamics are actually observed. However, when interpreting the calibrated parameters in the light of the careful measurements of the experimental system then a dichotomy arises.

Before comparing the calibrated parameters with those directly measured in the specific experiments, it is worth reviewing some of the parameters that have emerged in other efforts to calibrate, in particular, the neutral version of the model. The goal was to look at the dynamics, whereas in most previous studies (Ofiteru *et al.*, 2010; Sloan, 2006) the stationary distribution of taxa abundances predicted by the model has been calibrated against the observed abundance distributions in one or more similar samples. This yields a compound parameter N_{Tm} , for the local community. The values for N_{Tm} , especially in engineered microbial communities, but also in some natural analogues with rapid population turnover, seem to be reasonably consistent. So for ammonia oxidizing species in wastewater treatment works $N_{Tm} = 55$ and for heterotrophs $N_{Tm} = 198$ (Ofiteru *et al.*, 2010); and in experimental aquatic microcosms, $N_{Tm} = 10.3$ (Horner-Devine, 2003). In addition, above an uncertain threshold in population size, N_{Tm} seems to be scale invariant. In the one previous attempt to calibrate this model on time series data (Ofiteru *et al.*, 2010), the calibrated N_{Tm} was consistent with that calibrated using stationary distributions. The implication of having N_{Tm} values of the order 10 – 1000 is that for very large microbial populations, the probability, m , that a replacement event occurs through immigration, must be very small indeed. This would suggest that microbial populations are very insular and that once a

large community has established it would be difficult to perturb its structure by immigration.

The advantage of actually having explicitly measured the density of organisms in the chemostats and having controlled the volume of the fluid in the chemostats and the flux to and from them now comes into play. From Chapter 5, a very accurate knowledge of the growth characteristics is known and from the experimental design, one can calculate the growth rate and yield in the chemostats under the operating conditions that are imposed. Thus there are several sources of information that can inform the experimental parameter values that are completely independent of any assumed model or calibration technique.

So starting with immigration, throughout the thesis, immigration has been defined to be the probability that when there is a replacement in the community it is by a migrant rather than a local birth. When the chemostat is saturated with organisms and has reached steady state then the number of organisms leaving the system in the effluent per unit time, A , is the number of replacements per unit time. Similarly the number of organisms entering the system per unit time, B , is the immigration rate. Thus the probability that a replacement event occurs through immigration is approximately B/A . The density of organisms in the source chemostats and in the local chemostats were obtained and thus the values of N_T and the p_i are known, which are the population size and relative abundance of the i^{th} strain in the source populations respectively. Note that p_i changes with time due to fluctuations in

the source community. The mean time between replacements in each of the communities can also be calculated. Since my interest is in testing whether all ten of the chemostats are behaving stochastically then I have averaged these experimental variables across the chemostats and presented them in Table 8-1.

Table 8-1: The key variables measured in the local communities in the three different experiments.

Experiments	m	a	Experimental N_T	aN_T
Neutral dynamics with immigration	0.71	1.68E-11	1.81E+11	3.04
Non-neutral dynamics in an insular community	-	9.17E-11	6.61E+10	6.06
Non-neutral dynamics with immigration	0.46	7.94E-12	3.80E+11	3.02

For convenience I have repeated the parameter estimates obtained from fitting the SDEs to the data in Table 8-2.

Table 8-2: Summary of fitted model parameters for each chemostat experiment using corresponding calibration models.

	Stochastic Models	Process Variable	Output Value	LCL	UCL	pValue
1.	Neutral dynamics with immigration	m1	0.4956	0.3831	0.6080	6.2701e-16
		Sta. Error	0.0886			
2.	Non-neutral in insular community	m2	0.0902	-0.0105	0.1907	0.0788961
		Sta. Error	0.1176			
3.	Non-neutral dynamics with immigration	m1	0.3454	0.2795	0.4114	4.0617e-21
		m2	0.3038	0.2320	0.3757	4.8783e-15
		Sta. Error	0.0869			

Remembering that $m_1 = \frac{m dt}{N_T a}$ and $m_2 = \frac{2\alpha(1-m)dt}{N_T a}$ then if the experimental values are substituted into the expression, then $m_1 = 0.47$ for the neutral case and $m_1 = 0.31$ for the non-neutral dynamics with immigration. Thus there is an excellent correspondence with the calibrated parameters. For m_2 , the knowledge of the exact values of α from the experiments is not known but the growth rate is. If it is assumed that the calibrated parameters are correct and then estimate what the value of α would have to be for that to be the case, for the completely insular community, $\alpha = 0.13$ and for the community with immigrants $\alpha = 0.41$. Given that in the closed community the whole population is exposed to antibiotic for the duration of the experiment, whereas in the open community a large proportion of the births must occur from new immigrants that are only newly exposed to the antibiotic then these values seem entirely reasonable.

So at this stage everything still points to the model being an accurate representation of the dynamics in the community and the model parameters and the calibrated parameters appear to be consistent. That is until one considers the standard error in the weighted residuals that was estimated in the calibration routine. In the model, the standard error should be $\sqrt{\frac{1}{aN_T^2}}$. For the neutral case substituting in the measured experimental variables gives 1.3×10^{-6} , while for the insular communities where one taxon has an advantage this gives 1.5×10^{-6} and for the communities with immigration and selective advantage the procedure gives 9×10^{-7} . Clearly these are several orders of magnitude smaller than the standard errors in the model fit to the data, which are 8.9×10^{-2} for the neutral case, 1.2×10^{-1} for the insular community and 8.7×10^{-2} for the community where migration and selection exists. This is problematic, because if the standard error were indeed this small then the random fluctuations in the data would be almost imperceptible. Thus a dichotomy arises. I observe what looks like neutral dynamics and I can calibrate an SDE that represents neutral dynamics but the random component in the experimental observed dynamics are much larger than those that would be simulated by the model using the measured population size N_t and population turn over a . This lends weight to Nee's argument that it would be difficult to observe neutral dynamics in a large population (Nee, 2005). It also suggests that the model used is some how wrong, which is extremely frustrating since to all intents and purposes the dynamics have the characteristics of neutral or stochastic dynamics. Suppose, however, that there is some spatial correlation in the births and deaths in the chemostat. This is not an unreasonable assumption since despite all efforts these

systems are never perfectly mixed, so when an individual leaves the system, then it departs at the same time as its neighbours. In the model it is assumed that births and deaths occur completely at random anywhere in the community. In reality we might suppose that when there is a birth, the offspring appear immediately adjacent to the parent cell, whereas in the model there is no spatial representation. What is more, some aggregation of cells in loose flocs can be observed, and in most engineered systems there will be a degree of flocculation and biofilm formation whether intentional or not. Therefore, rather than considering individual cells let us group cell together and assume that they are in some way clumped.

What size of clumps would yield the stochastic dynamics that was observed in the chemostats? If the cells in the chemostat are grouped into clumps then they will exit the chemostat in those clumps, so if N_{cT} is the total number of clumps and a_c is the time for a clump to leave the system, then $N_{cT} a_c = N_T a$, meaning that the turnover of the population is just the same but that it occurs in clumps rather than individuals. N_{cT} will be smaller than N_T and a_c will be longer than a . Now if we suppose that the standard error in the model fit equates to $\sqrt{\frac{1}{a_c N_{cT}^2}}$, then we get the values of N_{cT} shown in Table 8-3. Note that the compound parameters m_1 and m_2 remain unchanged when replacing N_T with N_{cT} and a with a_c .

Table 8-3: The number of bacterial clumps that would yield the observed dynamics and an estimate of the number of bacteria per clump.

Experiment	N_{c_T}	N_T/N_{c_T}
Neutral with immigration	2.75E+06	6.58E+04
Non-neutral dynamics in an insular community	8.88E+05	7.44E+04
Non-neutral dynamics with immigration	4.08E+06	7.91E+04

The values of N_{c_T} that would yield the stochastic dynamics observed in the experiments are indeed much smaller than the values of the total numbers of individuals in the population, N_T . However, it is interesting to consider the number of individuals per clump that would be required, which can be estimated by N_T/N_{c_T} (Table 8-3). The number of individuals per clump would be very similar across all three experiments at approximately 10^4 . This may seem like a large number of individuals, but the density of bacteria in the chemostats is of the order 10^9 per ml. Thus, if the clumps comprise of the order 10^4 bacteria then a clump would occupy approximately 10 nano-litres of fluid in the chemostat; barely visible with the human eye. In reality what looks like flocculation in the communities in the chemostats can be seen.

Thus the modeled stochastic dynamics would match the experimental results if the population size were N_{c_T} rather than the number of individuals N_T . A similar situation occurs when considering allele frequencies in population

genetics. Indeed, the concept of effective population sizes sits at the heart of population genetics; thus times to fixation or extinction or allele frequencies are always expressed as a function of effective population size rather than the census population size; effective populations are always smaller than census populations. A number of different formulae have been derived, over the years, to estimate effective population sizes. Indeed, the origin of the concept of effective population is almost lost in the collective memory. It was originally introduced by Sewall Wright as a fix. He observed that theoretical predictions of the expected frequency distribution of neutral mutated alleles made using the census heterozygote populations did not match the observed allele frequencies (Wright, 1938). The effective population size is the number that would give the correct distribution. Subsequently the reasons for requiring an effective rather than census population size have been attributed variously to the number of breeding pairs, the spatial structure, and to age-structured dynamics. This thesis shows that for the ecological situation of bacteria in a chemostat, stochastic dynamics do occur and that they can be identified as being important even when there is selection. However, the theoretical model used here, which is based on random births, deaths and immigration, only matches the experimental findings if an effective population size is adopted. It can be speculated that this is required because the bacterial populations are spatially structured (clumped). When the size of the clumps needed to yield the required appropriate effective population sizes are estimated, they seem entirely reasonable and consistent across all the experiments. Others have suggested that effective population sizes for bacteria are small (Fraser *et al.*, 2009) and have attributed this to a range of possible mechanisms, such as

predation. Here I actually observe neutral dynamics in very simple experiments where the most of these mechanisms have been eliminated and yet an effective population size is still required. I suspect that clumping is ubiquitous and its presence promotes the importance of stochastic dynamics in all bacterial populations, rendering drift an important component of even large community dynamics.

9 Further Research

Researchers (Vellend, 2010; Nemergut, 2013) have strongly suggested that ecological patterns of species-area, species diversity, population density and abundances, observed or inferred in fauna, flora and microbial kingdoms all stem from four major ecological processes, namely selection, drift, dispersal and speciation. They advocate that ecological trends result from one or more of the aforementioned processes or a combination of all four processes. So far in this thesis, due to logistics and time constraints, the chemostats were run for only two weeks, thus there was not enough time for speciation to be studied. There is evidence that speciation can occur quickly in bacteria, especially in stressful environments, and thus in large bacterial populations it may play a role that is as important as dispersal. The population dynamics reported are based on three ecological processes – drift, selection and dispersal (migration).

Therefore a robust, detailed and controlled experiment aimed at observing speciation, or at least the accumulation of mutations in the genome, would further inform the community assembly process and lead to new models. Speciation would increase diversity locally at a rate that depends on the environmental conditions (McPeck, 2007). A modest extension to this study would be to measure the accumulation of antibiotic resistance in the two strains at locations away from the artificially inserts – *pepA* region of the *E.coli* genome. By calibrating a model that incorporates speciation the relative importance of immigration and evolution could be determined. If the rate of

evolution is high in comparison to immigration then it might act as a stabilizing force in the face of environmental fluctuations. It might help design for the long-term stability of engineered biological systems.

A key area of future research is to explore the generality of the finding that, to get theory and experiments to match, one needs to assume an 'effective community size' that is orders of magnitude smaller than the real number of individual bacteria in the model. This is profoundly important because it means that stochastic dynamics can be influential even in very big populations. It has been speculated, on the basis of circumstantial evidence, that this is due to clumping of bacteria in space. This should be tested in a wide variety of environments because, if it is true, then demographic stochasticity will need to be accounted for in engineering design.

10 Bibliography

1. Addams, L. et al (2009), "Charting our water future – Economic frameworks to inform decision making".
2. Archunan, G. (2004), "Microbiology", New Dehli: Sarup & Sons.
3. Beck, S. et al (2015), "The role of stochasticity differs in the assembly of soil and root associated fungal communities", *Soil Biology & Biochemistry*, Vol. 80.
4. Bekins, B. A. et al (1998), "A comparison of zero order, first order and monod biotransformation models", *Groundwater*, Vol. 36, No. 2.
5. Bell, G. (2001), "Neutral Macroecology", *Science*, Vol. 293.
6. Bronzino, J. (2000), "The Biomedical Engineering Handbook" 2nd Edition, Volume 2, Florida: CRC Press LLC.
7. Caruso, T. (2011), "Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale", *ISME* 5, 1406-1413.
8. Casali, N. (2003). *E.coli Plasmid Vectors. Methods in Molecular Biology*. Vol 235, 27-48.
9. Chave, J., (2004), "Neutral theory and community ecology". *Ecol. Lett.* 7, 241- 253.
10. Clark, D. P & Pazdernik, N. J. (2013), "Molecular Biology", 2nd Edition, Massachusetts: Elsevier
11. Clark, J. S. et al (2007), "Resolving the biodiversity paradox", *Ecology Letters*, Vol. 10, pp. 647-659.
12. Costantino, N. and Court, D. L. (2003), "Enhanced Levels of Red-mediated recombinants in mismatch repair mutants". *Proc. Natl. Acad. Sci. USA*, Vol. 100, pp. 15748-15753.
13. Curtis, Thomas P., Ian M. Head, and David W. Graham. "Peer reviewed: theoretical ecology for engineering biology." *Environmental science & technology* 37.3 (2003): 64A-70A.
14. Davies, P. S. (2005), "The Biological Basis of Wastewater Treatment", Strathkelvin Instruments Ltd.
15. de Crecy, E., Metzgar, D. et al (2007), "Development of a novel Continuous Culture Device for Experimental Evolution of Bacterial Populations", *Appl Microbial Biotechnology*, Vol. 77, pp. 489-496.
16. Dornelas, M. et al. (2006), "Coral reef diversity refutes the neutral theory of biodiversity". *Nature* 440, 80-82.
17. Elad, Y. & Baker, R. (1985), "The Role of Competition for Iron and Carbon in Suppression of Chlamyospore Germination of *Fusarium* spp. By *Pseudomonas* spp.". *Phytopathology* 75:1053-1059.
18. Etienne, R. S. & Olf, H. (2004), "How dispersal limitations shape species - body size distributions in local communities", *American Naturalist* 163: 69-83.
19. Ferrenberg, S. et al (2013), "Changes in assembly processes in soil bacterial communities following a wildfire disturbance", *The ISME Journal* 7: 1102-1111.

20. Fierer, N. & Lennon, J.T. (2011), "The generation and maintenance of diversity in microbial communities" *American Journal of Botany* 98(3): 1-10.
21. Fraser, Christophe, et al., (2009), "The bacterial species challenge: making sense of genetic and ecological diversity." *science* 323.5915, pp. 741-746.
22. Futuyma, D. (2009), "Evolution", 2nd Edition, New York: Sinauer Associates.
23. Gause, G.F. (1934), "The struggle for existence", IX. Williams and Wilkins. Baltimore, Maryland.
24. Gewin, V. (2006), "Beyond neutrality-ecology finds its niche" *PLoS Biol* 4(8): e278.
25. Grady Leslie, C. P. Jr. et al, (2011), "Biological Wastewater Treatment", 3rd edition, Florida: CRC Press.
26. Graham, D. W. & Smith, V. H. (2004), "Designed ecosystem services: application of ecological principles in wastewater treatment engineering", *Front Ecol Environ*, Vol. 2(4), pp.199-206.
27. Han, L. (2002), "Physiology of *Escherichia coli* in Batch and Fed-batch Cultures with Special Emphasis on Amino Acid and Glucose Metabolism", Royal Institute of Technology. Stockholm, Sweden.
28. Harpole, W. S. & Tilman, D. (2006), "Non-neutral patterns of species abundance in grassland communities", *Ecological letters*, Vol. 9, pp.15-23.
29. Harpole, W. (2010), "Neutral Theory of Species Diversity", *Nature Education Knowledge* 1(8): 31.
30. Harris, K. et al, (2015), "Linking statistical and ecological theory: Hubbell's unified neutral theory of biodiversity as a hierarchical Dirichlet process"
31. Harte, J., (2003), "Ecology: Tail of death and resurrection." *Nature* 424.6952, pp. 1006-1007.
32. Heijnen, JJ. & Roels, JA. (1981), "A macroscopic model describing yield and maintenance relationship in aerobic fermentation processes", *Biotechnol. Bioeng*, Vol. 23, pp. 739-763.
33. Hendrickson, E. R. et al (2002), "Molecular analysis of *Dehalococcoides* 16S ribosomal DNA from chloroethene-contaminated sites throughout North America and Europe", *Applied and Environmental Microbiology*, Vol. 68, No. 2, pp. 485-495.
34. Hilfinger, A. & Paulsson, J. (2011), "Separating intrinsic from extrinsic fluctuations in dynamic biological systems", *Proceedings of the National Academy of Science*, Vol.109, No. 29, pp. 12167–12172.
35. Holmes, I., Harris, K., and Quince, C. (2012), "Dirichlet Multinomial Mixtures: Generative Models for Microbial Metagenomics", *PLOS One*, Vol. 7, Issue 2.
36. Holt, R. D. (2006), "Emergent Neutrality", *Trend in Ecology and Evolution*, Vol. 21, No. 10.
37. Horner-Devine, M. C., Leibold, M. A., Smith, V. H. & Bohannan, B. J. M. (2003), "Bacterial diversity patterns along a gradient of primary productivity". *Ecol. Lett.* Vol. 6, pp. 613–622.

38. Horner-Devine, M. C., Carney, K. M. & Bohannan, B. J. M. (2004), "An ecological perspective on bacterial biodiversity", *Proc. R. Soc. London*, Vol. 271, pp. 113-122.
39. Hosokawa, R. et al (2009), "Autochthonous bioaugmentation and its possible application to oil spills", *World J Microbiology Biotechnology*, Vol. 10.
40. Hubbell, S. P. (2001), "The Unified Neutral Theory of Biodiversity and Biogeography", Princeton University Press.
41. Jorgensen, S. E. & Bendoricchio, G. (2001), "Fundamentals of Ecological Modeling", 3rd Edition, Oxford: Elsevier.
42. Jorgensen, S. E., Chon, T-S., Recknagel, F. (2009), "Handbook of Ecological Modeling and Informatics", Massachusetts: Witpress.
43. Kimura, M., and Ohta, T. (1971) *Theoretical Aspects of Population Genetics*. Princeton, NJ, USA: Princeton University Press.
44. Lebeau, T. (2011), "Bioaugmentation for in situ soil remediation: how to ensure the success of such a process". In: Singh, Ajay, Parmar, Nagina, Ramesh, Kuhad, C. (Eds.), "Bioaugmentation, Biostimulation and Biocontrol", Chapter 7. *Soil Biology*, Vol. 28. Springer-Verlag, Berlin Heidelberg, pp. 129 - 186.
45. Leibold, M. A. & McPeck, M. A. (2006), "Coexistence of the Niche and Neutral Perspectives in Community Ecology", *the Ecological Society of America*, 87(6), pp. 1399-1410.
46. Lenski, R. E. (1986), "Coexistence of two Competitors on One Resource and One Inhibitor: A Chemostat Model Based on Bacteria and Antibiotics", *J. theor. Biol.* Vol. 122, pp. 83-93.
47. Logan, J. et al (2009), "Real Time-PCR: Current technology and applications" Norfolk UK: Caister Academic Press.
48. Martin, A.J. (1927), "The activated sludge process" London, UK: Macdonald and Evans.
49. McPeck, M. A (2007), "The Macroevolutionary Consequences Of Ecological Differences Among Species", *Palaeontology* Vol. 50, pp. 111- 129
50. Miller, J. H. (1995), "Discovering Molecular Genetics: A case study with problems and scenarios".
51. Monod, (1950), "La technique de culture continue theorie et applications", *Annls. Inst. Pasteur, Paris*, 79:390-401.
52. Montgomery, D. C., Peck, E. A. & Vining, G. G. (2012), "Introduction to Linear Regression Analysis" 5th Edition, New Jersey: John Wiley & Sons Inc.
53. Mouquet, N. & Loreau, M. (2003), "Community patterns in source-sink metacommunities", *Am Nat*, Vol. 162, pp. 544-557.
54. Nee, S. (2005), "The Neutral theory of Biodiversity, do they numbers add up?" *British Ecological Society, Functional Ecology*, Vol. 19, pp.173-176.
55. Nemergut, D. R. et al, (2013), "Patterns and Processes of Microbial Community Assembly", *Microbiology and Molecular Reviews*, Vol. 77, No. 3, pp. 342-356.
56. Novick, A. & Szilard, L. (1950). Experiments with the chemostat in spontaneous mutations of bacteria. *Proc. Nat. Acad. Sci.* 36:340-345.

57. Ofiteru, I. D. et al (2010), "Combined niche and neutral effects in a microbial wastewater treatment community", PNAS, Vol. 107, No. 37, pp. 15345-15350.
58. Palmeri, L. Barausse, A. & Jorgensen, S. E. (2014), "Ecological Processes Handbook" 1st Edition, Florida: Francis & Taylor Group, CRC Press
59. Prosser, J. I, Bohannan, B. J. M, et al (2007), "The Role of Ecological Theory in Microbial Ecology", Nature Publishing Group, Vol. 5, pp. 384 – 392.
60. Roettig, A. & Steinbuechel, A. (2013), "Acyltransferases in Bacteria", Microbiology and Molecular Biology Reviews, Vol. 77, pp. 277-321.
61. Rosenzweig, M.L. (1995), "Species Diversity in Space and Time", Cambridge University Press, New York, NY.
62. Rosindell, J. et al (2012), " The Case for Ecological Neutral Theory", Trends in Ecology and Evolution, Vol. 27, No. 4.
63. Saiklay, P. E. & Oerther, D. B. (2004), "Bacterial Composition in Activated Sludge: Theoretical Analysis of Varying Solids Retention Time on Diversity", Springer Science, Vol. 48, pp. 274-284.
64. Sato, K. et al (2003), "On the relation between fluctuation and response in biological systems", Proceedings of the National Academy of Science, Vol. 100, No. 24, pp. 14086-14090
65. Scholar, E. C. & Pratt, W. B. (2000), "The Antimicrobial Drugs", (2nd ed.) New York: Oxford University Press.
66. Schumann, W. (2006), "Dynamics of the Bacterial Chromosome: Structure and Function. Weinheim: John Wiley & Sons, pp.100-103.
67. Simon, S. (2003), "Sustainability and Water Management", Glasgow: Bath press.
68. Sinclair, A.R.E., Fryxell, J.M., Caughley, G. (2006), "Wildlife ecology, Conservation and Management", (2nd ed.) Massachusetts: Wiley-Blackwell.
69. Sloan, W. T et al (2006), "Quantifying the roles of immigration and chance in shaping prokaryote community structure", Environmental Microbiology, 8(4), 732-740.
70. Sloan, W. T et al (2007), "Modeling taxa abundance distributions in microbial communities using environmental sequence data", Springer Science, Vol. 53, pp. 443-455.
71. Smith, H. L and Paul Waltman, (1995), "The theory of the chemostat: Dynamics of microbial competition", (1st ed.) Cambridge University Press,
72. Stroo, H. F. et al, (2013), "Bioaugmentation for Groundwater Remediation", New York: Springer Science & Business Media.
73. Tchobanoglous, G et al, (2003), "Wastewater Engineering Treatment and Reuse", (4th ed.) New York: McGraw Hill Education.
74. Tilman, D. (2004), "Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion and community assembly", PNAS, Vol. 101, No. 30
75. Volkov, Igor, et al., (2003), "Neutral theory and relative species abundance in ecology." Nature 424.6952, pp. 1035-1037.

76. Wagner, M. & Loy, A. (2002), "Bacterial community composition and function in sewage treatment systems". *Curr. Opin. Biotechnol.* Vol. 13, pp. 218–227.
77. Wagner, W. et al, (2002), "Microbial community composition and function in wastewater treatment plants" *Antonie van Leeuwenhoek* 81:665-680.
78. Williams, F. M. (1971), "Dynamics of microbial population", in BC Patten (ed.), *Systems analysis and simulation in Ecology*. New York: Academic Press, pp. 197-265.
79. Wright, G. D. & Thompson, P. R. (1999), 'Aminoglycoside Phosphotransferases: Proteins, Structure and Mechanism', *Frontiers in Bioscience*, Vol. 4
80. Wright S., (1938). "Size of population and breeding structure in relation to evolution". *Science* **87** (2263), pp. 430–431.
81. Woodcock, S. et al, (2007), "Neutral Assembly of Bacterial Communities", *FEMS Microbiological Ecology* Vol. 62, pp.171-180.
82. www.drugbank.ca/drugs/DB01172
83. www.microbelibrary.org/component/resource/laboratory-test/3202-indole-test-protocol
84. www.unwater.org/downloads/waterscarcity.pdf
85. Xiao, J. & VanBriesen, JM. (2005), "Expanded Thermodynamic Model for Microbial True Yield Prediction", *Biotechnology & Bioengineering* Vol.93, No.1, pp. 110-120.
86. Yu, D. et al, (2000), "An Efficient Recombination System for Chromosome Engineering in *Escherichia coli*", *PNAS*, Vol. 97, No. 11, pp. 5978-5983.
87. Zhang, W & Fuller, G. N. (2004), "Genomic and Molecular Neuro-oncology", Massachusetts: Jones and Bartlett Publishers, Inc.