

AGENT BASED MODELING FOR SIMULATION OF MICROBIAL COMMUNITY

A Paper
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Noah Daniel Tekeste

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Computer Science

December 2017

Fargo, North Dakota

North Dakota State University
Graduate School

Title

AGENT BASED MODELING FOR SIMULATION OF MICROBIAL
COMMUNITY

By

Noah Daniel Tekeste

The Supervisory Committee certifies that this *disquisition* complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Simone Ludwig

Chair

Dr. Anne Denton

Dr. Peter Bergholz

Approved:

December 12, 2017

Date

Dr. Simone Ludwig

Department Chair

ABSTRACT

Agent based modeling uses interacting agents and a governing rule to understand a complex phenomenon. It is an important mode of inquiry in the field of life sciences. For this paper a Haploid Evolutionary Constructor (HEC) tool was used for modeling and simulation of two sample models. The model samples were analyzed under the same and varying level of specific and non-specific substrates. In the first part of each experiment, the survival rate of the models was examined based on the model's inputs and outputs. However, since close association of microbes enhances the probability of Horizontal Gene Transfer (HGT) between organisms, HGT behavior was introduced within the populations in the second sets of experiments. This enabled them to adapt to dwindling resources in their environment and creation of new populations that are better suited for the ecosystem. Based on the obtained results, the behavior of the dominant population(s) is assessed.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my advisor Dr. Simone Ludwig, a very humble instructor and passionate researcher, who presented me with an incalculable opportunity to work with her. She has pushed me and steered me in the right direction, kept me in line and tirelessly advised me on the approaches and methodologies I should use starting from day one. This accomplishment would not have been possible without her.

This work wouldn't be possible without Dr. Peter Bergholz who introduced me to an array of new field and helped me challenge myself. His ideas and feedbacks have been the corner stone of this research and I couldn't thank him enough. Mr. Nick Dusek was among the key factors that helped me understand and envision the problem at hand and I am gratefully indebted to his hard work and valuable input in this project.

Special thanks to Alexandra Klimenko one of developers of Haploid Evolutionary Constructor software, who has diligently helped me understand configuration of HEC and her selfless attitude to give me her feedback regardless of her busy schedule.

I also want to express my profound gratitude to my parents Daniel Tekeste and Letekidan Araia who provided me with unfailing love and support since day one. I thank them from the bottom of my heart.

Finally, to the corner stone of my life, my refuge, and my hope. To the almighty God, for his grace and the energy as well as hope he instilled in me.

Noah Daniel Tekeste

DEDICATION

To my better half, my wife Winta, who encouraged me day in and day out throughout this
journey.

TABLE OF CONTENTS

ABSTRACT.....	iii
ACKNOWLEDGEMENTS.....	iv
DEDICATION.....	v
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
LIST OF ABBREVIATIONS.....	x
1. INTRODUCTION.....	1
2. AGENT BASED MODEL.....	4
3. RELATED WORK.....	6
3.1. Sugarscape.....	8
3.2. Schelling Model Of Ethnic Residential Dynamics.....	10
3.3. Haploid Evolutionary Constructor.....	13
4. METHODOLOGY AND EXPERIMENTS.....	21
4.1. Experiment 1.1.....	26
4.2. Experiment 1.2.....	30
4.3. Experiment 2.1.....	35
4.4. Experiment 2.2.....	39
5. CONCLUSION.....	42
REFERENCES.....	44
APPENDIX A. HEC SCRIPT FOR MODEL 1 (HGT NOT INCLUDED).....	46
APPENDIX B. HEC SCRIPT FOR MODEL 2 (HGT NOT INCLUDED).....	48

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1.	Model 1 population composition.	23
2.	Model 2 population composition.	24
3.	Prior and Posterior concentration of NS and SS of Experiment 1.1.	29

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1. An artificial society on sugarscape: societal evolution from a random initial distribution of agents.....	10
2. Initial configuration of one of schelling’s experiments. Model showing satisfied resident (a) and unsatisfied resident (b).....	11
3. Result of simulation of Schelling’s Model	12
4. Main HEC Interface.....	15
5. Initializing global variables.....	17
6. Adding new population.....	18
7. Initializing population specific variables	18
8. Initializing additional parameters	19
9. Initial model script and population trophic rings at the end of generation	19
10. Scheme of Model 1 substrates at the end of the given evolution.....	26
11. Trophic rings of Model 1 at the end of the given evolution	27
12. Population size dynamics of Model 1	28
13. Substrate (NS and SS) concentration dynamics of Model 1.....	28
14. Scheme of Model 1 substrates at the end of the given evolution (HGT included).....	30
15. Trophic rings of Model 1 at the end of the given evolution (HGT included).....	31
16. Population size dynamics of Model 1 (HGT included)	32
17. Substrate (NS and SS) concentration dynamics of Model 1 (HGT included)	32
18. Scheme of Model 2 substrates at the end of the given evolution.....	35
19. Trophic rings of Model 2 at the end of the given evolution	35
20. Population size dynamics of Model 2.....	36
21. Substrate (NS and SS) concentration dynamics of Model 2.....	37
22. Population size dynamics after removal of population 1 from the original Model 2	38

23.	Scheme of Model 2 substrates at the end of the given evolution (HGT included).....	39
24.	Trophic rings of Model 2 at the end of the given evolution (HGT included).....	39
25.	Population size dynamics of Model 2 (HGT included)	40
26.	Substrate (NS and SS) concentration dynamics of Model 2 (HGT included)	40

LIST OF ABBREVIATIONS

ABM.....	Agent Based Modeling
HGT.....	Horizontal Gene Transfer
NS.....	Non-specific Substrate
SS.....	Specific Substrate
HEC.....	Haploid Evolutionary Constructor
EBM.....	Equation Based Modeling
HPC.....	High-performance Computing

1. INTRODUCTION

In the past few decades, researches in the field of life sciences have increased in number due to the availability of rich data regarding many elements of the field. The availability of enormous data is primarily due to the advancement of data generation as well as the tremendous progress in the field of computing including hardware and software (An G., 2016). Nowadays, the availability of rich data has led computation to become one of the key components of many research areas. Hence, computing has become a vital element among an array of fields that depend on and use data for different purposes including biomedicine and healthcare. Generally, this computation is done using the help of models and other variables that characterize the system being studied and results in an accelerated discovery of new information and aids researchers to make better predictions in response to changing conditions. Basically, the principal purposes of a model are to enhance and deepen one's understanding of the systems being modeled and to help in decision making by providing a great deal of information about the modeled system. Formerly, the description of a system using mathematical concepts, also known as mathematical modeling, has been commonly applied to areas within social and life sciences to represent different systems. However, despite the popularity of mathematical modeling, agent-based modeling (ABM) has gained far more popularity in modeling systems especially in the field of life sciences.

Agent-Based Modeling is a powerful simulation modeling technique mostly used to model dynamic systems of interacting agents. This modeling paradigm has been widely used in modeling several applications such as flow, organizational, market, and diffusion simulations (Castiglione, 2006). The main characteristics of ABMs are that they are easy to construct, their ability to capture spatial heterogeneity, and provide an authentic representation of local characteristics that generate global dynamics.

A vital characteristic of ABM's is the possibility of asynchronous interactions between different agents as well as among agents and their environment. Agents in ABM follow a sequential schedule of interactions instead of performing actions simultaneously. This discrete-event setup allows the cohabitation of agents with different environmental experiences (Castiglione, 2006).

ABMs have become an increasingly important mode of inquiry for the life sciences. They are particularly valuable for complex systems that are not understood well enough to build an equation-based model (EBM). An EBM contains a set of equations as a model and the execution involves evaluation of these equations. However, ABM's model consists of a set of agents that encapsulate the behavior of the various individuals that make up the system and the execution is basically emulating these behaviors (Van Dyke Parunak, 1998). The behaviors and interaction of agents can also be formalized by equations, but more generally they are mostly specified through logical rules. This makes the modeling approach much more flexible. Furthermore, it is easily possible to consider individual variations in the behavioral rules (i.e. heterogeneity) and random influences or variations (i.e. stochasticity).

Although ABM construction is more attainable than other approaches involving mathematical equations, it has its own challenges. Like EBM, ABM allows the comparison of model output with observed system behavior. On top of this, ABM allows validation of individual level of comparing encoded behavior of agents with the actual behavior of the real agents. However, this requires additional data on the agents and more efforts in empirical research on the system being modeled. Additionally, dynamic systems that are modeled using ABM techniques are composed of adaptive agents which can also result in a synergistic effect. The adaptability and synergism are features that are challenging to encode when constructing an ABM. Furthermore,

for broad and complex models, the presence of a large number of rules can affect the state and behavior of agents, which are evaluated at each step. This extensive evaluation of probabilistic rules can lead to significant computation cost and reach computation limits quickly (An G., 2016). Regardless of the above fact, what makes ABM a very appealing methodology for the simulation of biological systems is, the use of rich data to study a model through analyzing the heterogeneity of interacting agents between them and interaction between agents and their environment.

This paper discusses agent-based modeling experiments to identify and analyze the existence of dominant agents as well as influential ones, also termed as keystone species, with the help of a simulation tool known as the Haploid Evolutionary Constructor (HEC). The models presented in the experiments are sample populations which behave like *E. Coli* in the way they reproduce, the size of population (colony), their mortality rates and their utilization of substrates. The analysis and discussion on this research also focuses how populations and their community are affected because of their substrate intake and utilization, the symbiosis that exists between them as well as presence or absence of several ecological factors like horizontal gene transfer and gene loss.

2. AGENT BASED MODEL

A model is a more simpler way of representing a system which is designed mainly to convey a better understanding of the system and its operation. Of the several ways of representing a system, the most popular methods include mathematical, statistical, and computational methods. While mathematical and statistical methods use mathematical formulas and statistical processes, respectively, to study the behavior of a system, a computational method involves developing computer programs to build models.

The rule based approach of ABMs, together with the computational modeling techniques, enables an easier way of studying different agents and simulation of complex applications in the field of biology. At the heart of the ABM structure lies individual agents with a set of characteristics and a set of rules to guide their interaction with one another as well as with their environments. The existence of different agents with different sets of characteristics and the dynamics of a system makes the modeling problem challenging, but an exciting topic to research. Dynamic systems are systems that are composed of adaptive agents that can respond to stimulus from other agents or the environment.

When designing an ABM, the rules that govern the interaction of the agents are encoded in software as simple logical rules. Associated with these rules are random number generation mechanisms that create randomness in the system to mimic the real-world model. Starting with the rules, agents reconstruct populations are generated and the simulation of the real-world interaction is constructed. This helps to understand and study both the interactions and relationships that exists between agents. When designing an agent based model for biological studies the following key characteristics should be examined:

1. Customizable structure: an agent is defined by a set of rules and in ABM a new agent can be introduced to the model or even an existing one can be modified without altering the whole model.
2. Emergent properties: Interaction between agents often leads to a synergism, a higher-level relationship between group of agents, with much more sophisticated behavior and effect which is greater than the individual agents. (Politopoulos, 2007).
3. Abstraction: The concept of abstraction can be implemented when constructing an ABM by omitting unnecessary details. Furthermore, an ABM can be designed even without complete knowledge of the system resulting in a simple but verifiable model (An, 2013).
4. Volatility: Biological systems exhibit random fluctuations in their behavior. Hence agent based models can be designed to include this stochastic behavior through the random number generating function. Behavior of agents are determined starting from the population level with the following different kinds of trajectories to determine the agent behaviors based on the probability of the behavior of the individual agents. These behaviors are later translated as that of the agent's rules (An, 2013).

The simplicity in the implementation and construction of ABM makes it widely used and more favored modeling method compared to other modeling systems that often involve differential equations. Nowadays, researchers build a more suitable, convenient, and easy to understand model for their research problem and can further analyze and investigate their hypotheses using experiments involving ABM. Before the agent-based model can be developed, one must choose the appropriate software or modeling toolkit to be used, which mainly depends on the type of problem at hand.

3. RELATED WORK

Microbial communities often consist of different types of microorganisms that interact with one another. There are different kinds of interactions between microorganisms, which help form an ecological network with various kinds of relationships between them. Basically, these relationships are categorized into three namely Predation, Competition, and Symbiosis. Symbiosis, which is the interaction between two organisms living in close physical association can further be classified to Mutualism, Commensalism and Parasitism.

- Predation: an exploitation in where an organism, the predator, kills and consumes another pray organism.
- Competition: a relationship where organisms affect one another negatively to fight for the supply of resources used by both.
- Mutualism: a relationship both the interacting microorganisms benefit from the existence of one another.
- Commensalism: also known as one sided relationship where a microorganism can benefit from the existence of the other without neither harming or helping it.
- Parasitism: when an individual organism attempts to profit at the expense of the other with or without killing the host organism.

With recent findings showing the presence of certain assembly rules and the fact that interaction differs among species that either form or avoid partnership with other species, the study of these intricate relationship styles is vital in producing a better model for a microbial ecosystem (Shashkova, 2016).

Nowadays, Agent Based Modeling has become an important tool for researches involving interacting agents. ABM can help in testing, evaluating and visualizing processes through studying

the interaction between the autonomous agents. Through studying the patterns and relationships between these agents and their environment, it can also predict the outcome of an overall change in the population or the effect of disturbances in the environment (Garcia, 2005).

Although a better understanding of basic cellular processes was manifested by the recent advancement in the field of genomics and related researches, the question of how the individual cellular activity contributes to the overall population dynamics was vital in the field of both ecological and clinical studies. The agent based modeling approach is a crucial method to understand the effect of an individual cell among both other autonomous agents and its environment.

Through pre-defined rules and attributes that relate to individual components in a microbial ecosystem, an ABM can examine and predict a developing trend within the population by focusing on the laws that govern the interaction of the individuals only. This methodology allows the local differences among different populations within the environment to be considered in the model (Murphy, 2008). Sugarscape model and Schelling model of ethnic residential dynamics are one of the classical and major driving forces to popularizing the use of ABM in different fields. They show how simple and interpretable rules for agents could simulate behavior that was interesting in areas like demography, anthropology, sociology, and economics. They carry out functions such as consumption, reproduction, tribal growth, survival, and migration patterns that are commonly seen in microbial communities as well.

3.1. Sugarscape

Sugarscape is an agent based model which was created to simulate the distribution of wealth in a society. This model was developed by Epstein and Axtell in 1996 (Epstein & Axtell, 1996). The first model represents an artificial society in a two dimensional 50 x 50 grid. Grid cells contain a sugar level which is the amount of sugar a cell has and the sugar capacity which is the maximum value of sugar a cell can hold. At any point, a sugar level and a sugar capacity are given, sugarless points with low capacity are called deserts and some sugarless points have high capacity. The fundamental elements of this model are: agent, rules, landscape and sugar (resource). Agents start working in random on their primary situation, assets, and all their internal area. Whereas some subgroups of the internal states remain unchanged within the agent's life, the existence of other subgroups depend on time. Additionally, some states are spatial and different for some and common for others. The spatial time independent states are the primary assets, maximum life time, vision and metabolism rate. The overall independent states include: time needed for increasing vision, poverty limit (reaches to zero), spatial time dependent states such as agent situation in landscape, real asset in sugar units. The agent executes rules simultaneously in searching for sugar. The movements of the population is an emergent result of simple spatial activities by agents.

As in most ABMs, the four important characteristics in this model are:

1. Perception: the ability of an agent to sense agents and resources in its environment. Hence in this model agents can perceive the amount of sugar the current cell has.
2. Performance: the capacity of an agent to perform different tasks to maintain its survival. This set of tasks can range from moving to interacting and communicating with other agents in its environment. Agents in this model move and consume sugar for their survival.

3. Memory: ability of an agent to record its previous action and states to attain an optimal outcome in the upcoming generations.
4. Policy: this are set of guidelines that enable the agent to come up with the best possible strategy given a certain criterion in its current state. The memory of its previous states is helpful in such cases where a new strategy is required. In the sugarscape model, the agent may look for cells with the highest sugar levels.

The rules of the sugarscape ABM are executed on a 50x50 grid, which consists of 2500 cells as individual agents. There may be other agents than sugar cells. The distribution of sugar is shown in Figure 1 in which the amount of sugar cells can be predefined according to the growth rate and it can rise and be searched by another agent for existence of sugar or production. As shown in Figure 1, the amount of sugar is high in the first quadrant (northeast) and third quadrant (southwest) parts of the grid. The first quadrant shows random initial distribution of agents in the available grid. However, in the third quadrant the flow of agents towards the areas with highest amount of sugar level is noticed. In Part 4 (southeastern quadrant), the agents concentrate their activities on the sugar peaks that lead to the formation of two colonies. In cases where resources can fall short often, agents with high metabolism or low vision can easily perish (Abdou, Hamill, & Gilbert, 2012).

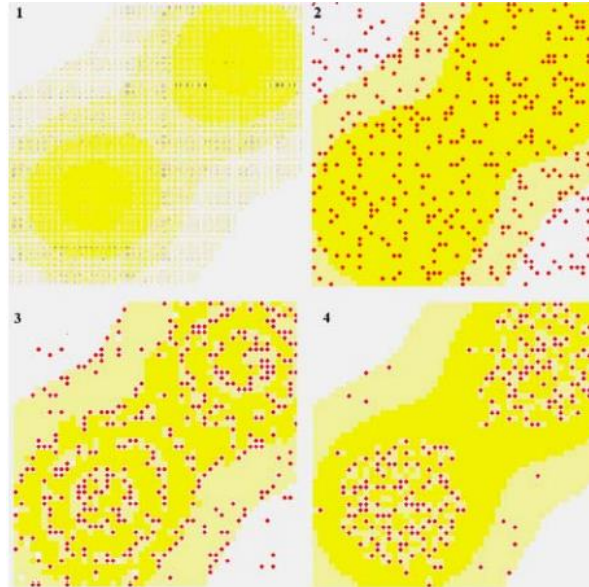


Figure 1. An artificial society on sugarscape: societal evolution from a random initial distribution of agents (Rahman, 2007).

3.2. Schelling Model of Ethnic Residential Dynamics

This is one of the oldest classic examples implemented using an agent based model to analyze the residential segregation. Racial segregation has always been a problem in the United States despite the many efforts extended to desegregate schools, churches, and neighborhoods. In 1971, Thomas Schelling, an American economist created an ABM to help explain why segregation was difficult to combat (Hatna, 2012).

In the Schelling model, a city is represented by a square grid of cells in which each cell represents a residence. A cell can be any of the three colors white, black or grey according to the color of the agent living there. A white cell refers to a white agent living there, a black cell for a black agent and grey if it is empty. Simulation starts by randomly placing agents on the grid and it is assumed that the agents have a threshold of tolerance of the other neighboring ethnic groups. The threshold defines whether the agent is happy with its neighbors, that is if cells to the north, south, east, west, north east, north west, south east, south west of the agent have a proportion more

than the given threshold or whether it decides to move when the proportion falls below the given threshold. For example, a 50% threshold of tolerance would mean that the agent is happy to stay in its location if at least four of its eight neighbors are of the same color as itself. On the other hand, the agent is not content and will try to move to another neighborhood if this proportion falls below 50%.

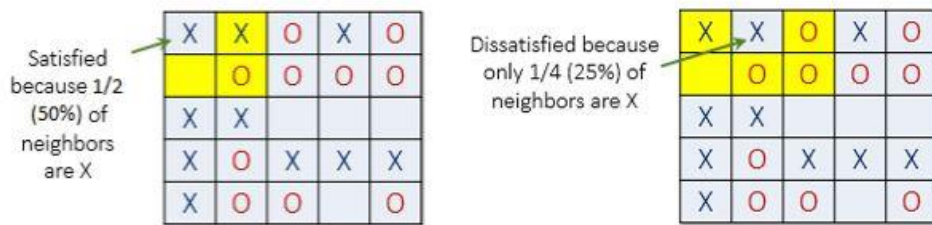


Figure 2. Initial configuration of one of Schelling's experiments. Model showing satisfied resident (a) and unsatisfied resident (b) (Hatna, 2012).

The result given below indicates a simulation with 2000 agents with the upper-left panel showing the starting random allocation of black and white agents on the grid. The other three panels are the results of the simulation with different thresholds of tolerance. In the simulation with a 37.5% threshold of tolerance, at least three of an agent's eight neighbors must be of the same color for the agent to stay put and not move to another place (Abdou, Hamill, & Gilbert, 2012). With a 50% threshold of tolerance at least four neighbors, and with 75% at least six out of eight neighbors should be there for an agent to be happy with its location.

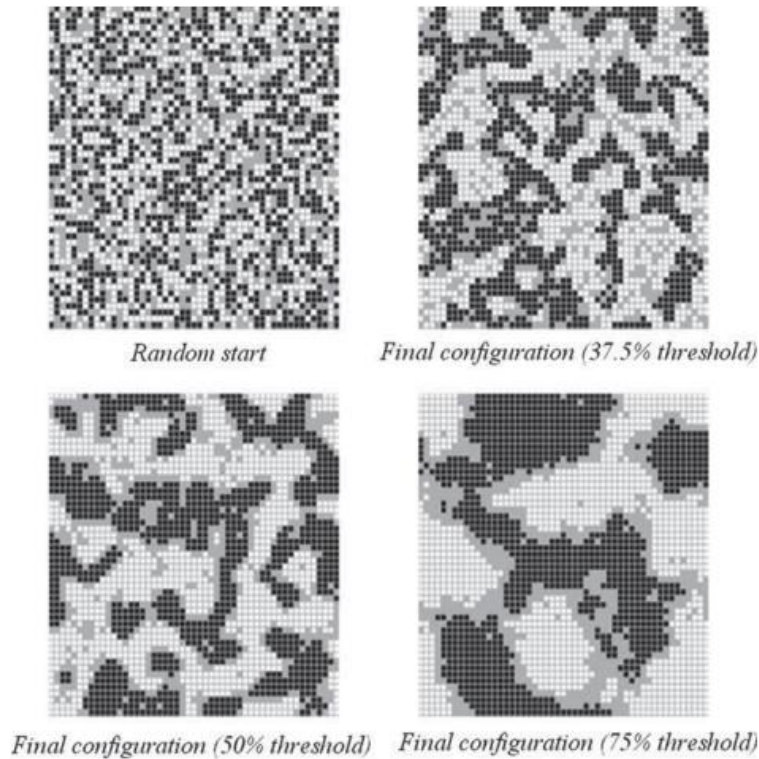


Figure 3. Result of simulation of Schelling's Model (Abdou, Hamill, & Gilbert, 2012).

The Schelling's model demonstrated that agents would disaggregate themselves from other agents when the other agents surrounding them are of different race. The model that illustrates the emergent behaviors that are mainly caused due to continuous aggregation and disaggregation of agents.

ABMs are good for exploring how micro-scale processes give rise to macro-scale phenomena because most of these phenomena are not easy to be represented mathematically or statistically. The primary reason is because agents can influence one another (mutualism), which can rather simply be modeled using agent-based modeling. For example, building a model to study how agents influence each other's purchasing choices. Another advantage of ABM is it enables one to experiment with the models and analyze different scenarios. Using a computer model, one can set many experiments using a range of parameters. For example, it is better to use a model of

an airplane to test flying under various conditions than to use a real one, in which the cost of experimentation is high. As a result of this great advantage of experimentation and analysis, ABM has been used in many areas of research including social study, economics, and biology.

Nowadays, it is not difficult to find tools that facilitate developing an ABM. Besides Swarm, an open source agent-based simulation package, Repast, Netlogo, and Mason are among commonly used tools for agent-based model implementation.

3.3. Haploid Evolutionary Constructor

A Prokaryote is a microscopic unicellular organism that consists two domains named Bacteria and Archaea. Naturally, Prokaryotes exist as part of microbial communities, which has a complicated structure whose metabolic pool often form a complete cycle. These communities are represented by a complicated trophic chain with a degree of closeness between them. They have an extensive symbiotic relationship in the ecosystem, which results in the creation of trophic cycle also known as trophic ring. Trophic rings are created when metabolic products of one species or strain are used or can in certain conditions be used for food by others. In Prokaryotic communities, the close bacterial association enhances the probability of horizontal gene transfer between different bacteria, which enables them to obtain new functions. Horizontal gene transfer is a mutational process in which organisms share genetic materials among one another. It is one of the processes that enables organisms to acquire new capabilities within a generation and enables them to adapt to their ecosystem (Aoki, 1999). A bigger percentage of prokaryotes is unculturable due to lack of critical information on the organisms' biology. For instance, failing to replicate essential aspects of their environment can be a reason for an organism to become non-culturable (Stewart, 2012). Conversely, it is the culturable prokaryotes, that represent the minor fraction of the overall diversity of organisms, that show high metabolic and reproduction activities. The evolution of such

highly-integrated prokaryotic communities has its qualitative specific factors and cannot be reduced to the evolution of the distinct population composing them. Therefore, even though the reproduction rate of prokaryotes is remarkably high, an experimental study of their evolution is difficult as it requires the study of the whole prokaryotic community.

Modeling of evolution is one of the primary challenges of 21st century biology, mathematics, and computer science (Lashin, 2014). Biological systems are much more complex, interlinked and often many or most of the quantitative values needed for an experiment are missing (Wooley, 2005). One of the main methods to study evolutionary process is through mathematical modeling. This type of modeling is convenient to describe different trophic interactions, spatial distribution of organisms, genetic structure of population and other environmental factors.

Traditional approaches to evolution and population process modeling include methods of population dynamics and methods of population genetics. The population genetics methods are generally based on methods of the probability theory and mathematical statistics. Although, traditional approaches allow the study of the evolution of population genetic structure they fail to convey the means for modeling of the population dynamics process in detail (Lashin, 2014).

The Haploid Evolutionary Constructor (HEC) is software package, which was developed using the java programming language. It simulates the operation of a population network of haploid organism, like microbes, which are trophically linked with substrate-product relationships under the environmental effect. With varying degree of description detail, HEC provides a means of simultaneously describing the prokaryotic community at different levels of its biological organization: genetic, metabolic, population, and ecological flexibility. Population variation in both size and diversity may arise due to selection and mutation during the model simulation process. The important element of this methodology in general, and HEC in particular, is its

provision of modeling evolutionary and population processes that require a significant rearrangement of structure of a model during the simulation process.

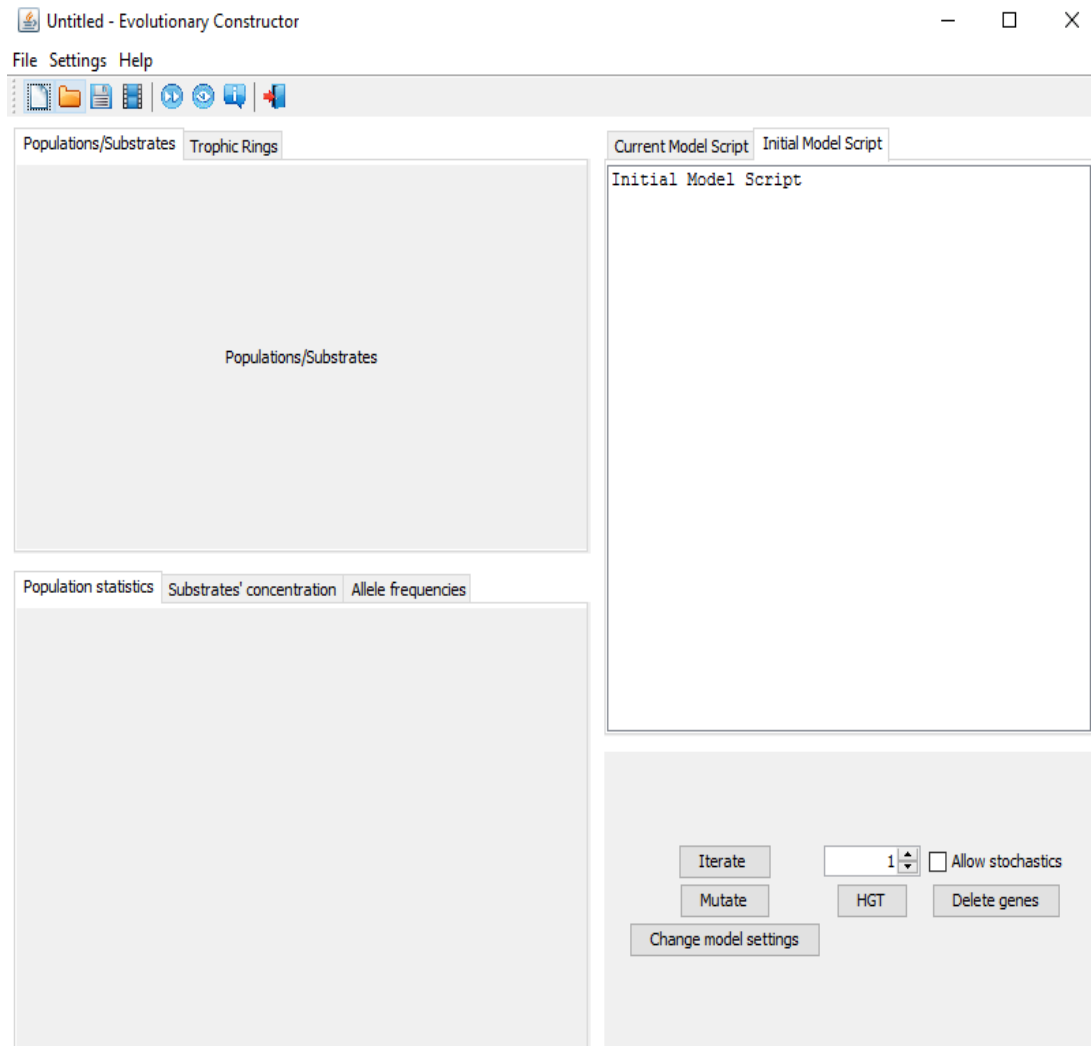


Figure 4. Main HEC interface. Snapshot taken from Haploid Evolutionary Constructor Software developed by Sergey A. Lashin (Lashin, 2014).

HEC allows two ways of creating a model:

1. Writing a script using the HEC format.
2. Creating a script using the new model wizard.

The HEC script is a look alike of C or Matlab style script with each script containing two essential parts. The first part, which is the model declaration section, starts with a declare statement and ends with an end declare statement. This section contains the declaration of environmental volume and flow, the number and concentration of specific and non-specific substrates, the reproduction and mortality rate among other parameters as well. A substrate is an input to a biochemical reaction. Enzymes act on substrates and yield products, which themselves often become substrates for further reactions. Substrates in an ecosystem are considered non-specific if many or all members of a community have the enzyme(s) necessary to metabolize it. As a result, organisms may compete for the non-specific substrate if it is in limited supply. Conversely, a substrate may be considered specific if only one or a few members of the community have the enzyme(s) necessary to metabolize it. In this case, competition for the substrate only occurs between those who share the corresponding enzyme(s), and as such, is specific to a subset of the community.

The second part of HEC script, which is the model calculation section, starts immediately after the end of the declaration statement and initializes the number of generations and types of operations, such as mutation and HGT, performed in the population. A thorough explanation of different sections of the script is given in the Methodology and Experiments section of this paper.

A model script can also be created using HEC model wizard which is a step by step guide to creating a script by setting the required parameters.

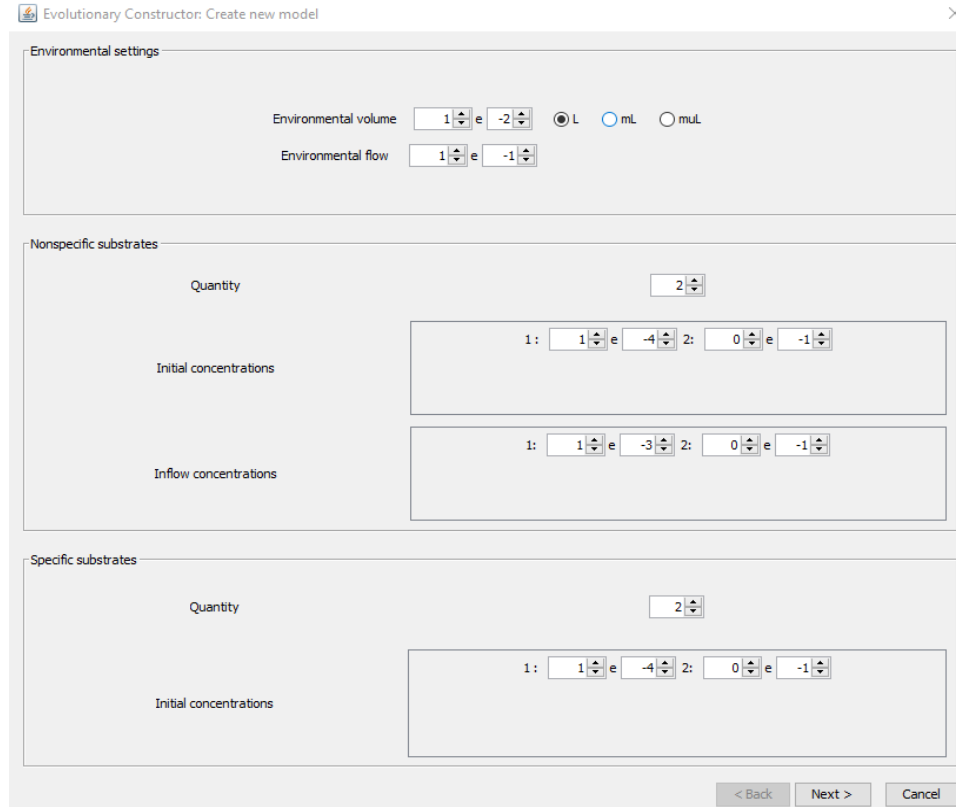


Figure 5. Initializing global variables. Snapshot taken from Haploid Evolutionary Constructor Software developed by Sergey A. Lashin (Lashin, 2014).

The following snapshots (Figure 5 to Figure 9) describe the step by step model creation using the HEC software wizard. Figure 5 initializes global variables such as environmental setting, number and concentration of specific and non-specific substrates. The step shown in Figure 6 enables addition of population(s) to a model. After adding the desired population(s), what follows is the initialization of population specific variables like size of population, trophic strategy, death coefficient, consumption rate of Non-specific Substrates (NS) and Specific Substrates (SS) and other variables. The snapshot for this initialization is given in Figure 7.

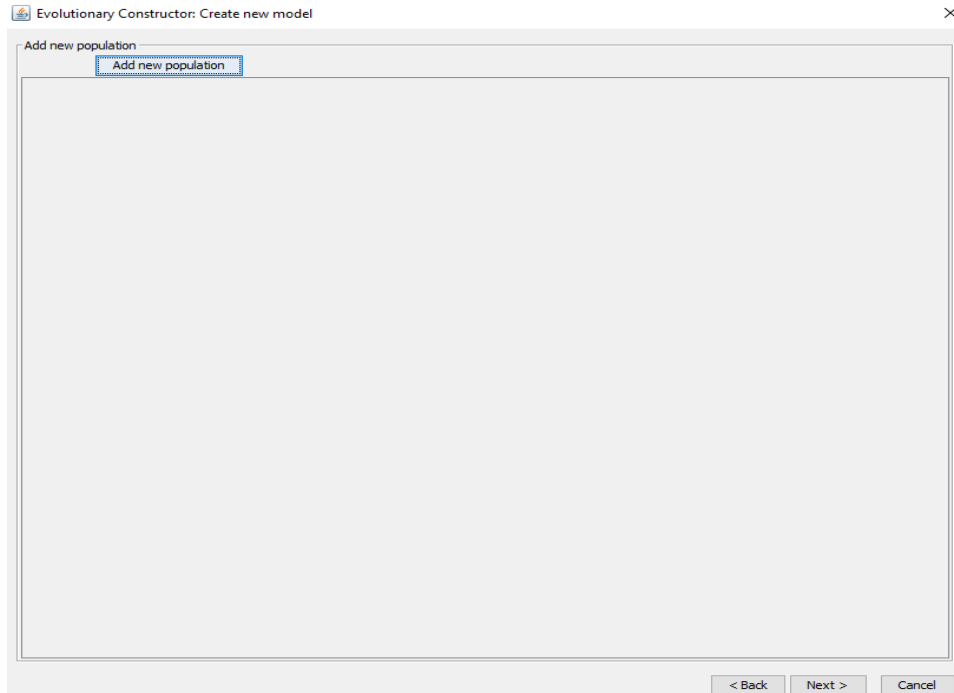


Figure 6. Adding new population. Snapshot taken from Haploid Evolutionary Constructor Software developed by Sergey A. Lashin (Lashin, 2014).

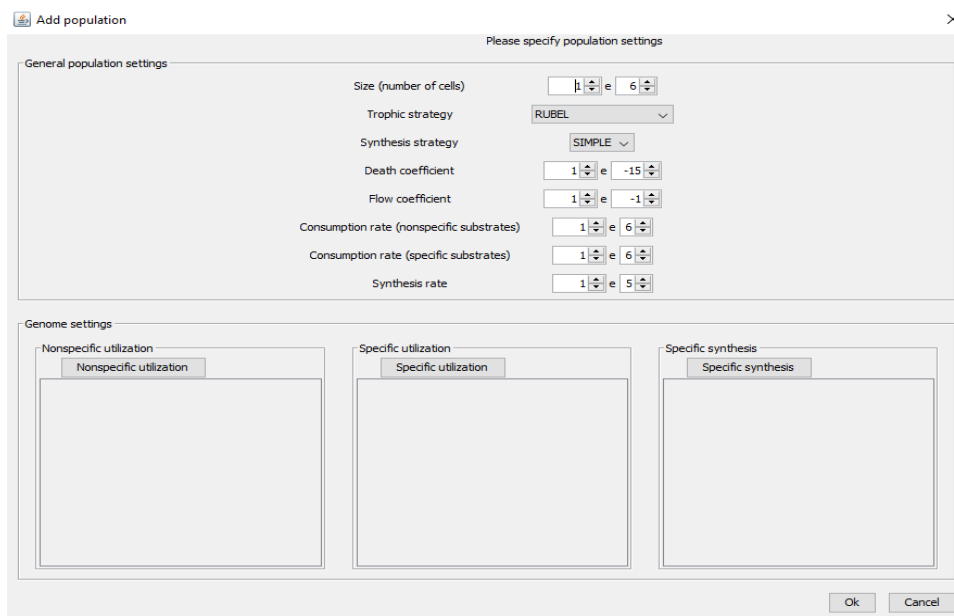


Figure 7. Initializing population specific variables. Snapshot taken from Haploid Evolutionary Constructor Software developed by Sergey A. Lashin (Lashin, 2014).

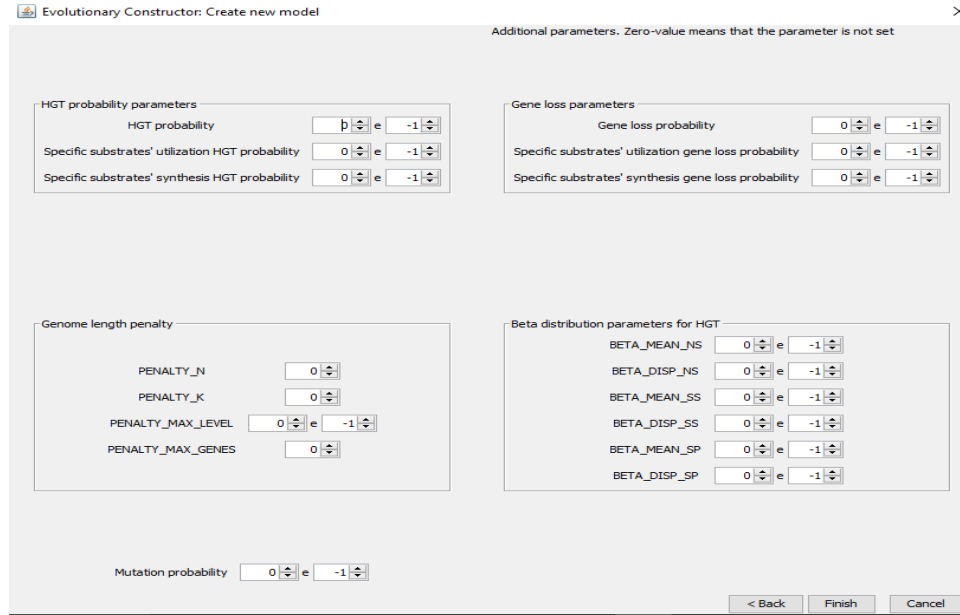


Figure 8. Initializing additional parameters. This setting includes the setting for HGT and Gene loss. Snapshot taken from Haploid Evolutionary Constructor Software developed by Sergey A. Lashin (Lashin, 2014).



Figure 9. Initial model script and population trophic rings at the end of generation. Snapshot taken from Haploid Evolutionary Constructor Software developed by Sergey A. Lashin (Lashin, 2014).

In summary, HEC software tool allows the comprehensive study of the microbial community model analyzing the dynamics of changes in allelic frequency, population size,

concentration of metabolites, community trophic structure and its evolution, including stochastic genetic factors.

4. METHODOLOGY AND EXPERIMENTS

In this paper, four experiments were conducted based on two models. For each model, two scripts, one without the possibility of the mutational process through HGT and gene loss and the other with the possibility of HGT and gene loss, were prepared. Due to computational limitation, each simulation runs only for 1000 generations. In each script, the list of specific and non-specific substrates is enumerated beginning from one in increments of one until one reaches the total number of specific and non-specific substrates, respectively. For instance, non-specific substrates in a population are enumerated as N1, N2, N3, etc. and specific substrates are enumerated as S1, S2, S3, etc. In each model, there are different simulations with their respective codes, which are provided in Appendix A. The first part of each model's experiment entails examining behaviors of studies of populations with different combinations of specific and non-specific substrates. However, in the second part of the experiments the possibility of mutation through horizontal gene transfer and gene loss is applied stochastically throughout the entire simulation at a rate of 0.01 (1% of the whole generation) and the behavior of the populations is analyzed accordingly.

As discussed in the previous section, the HEC model script consists of two different sections:

- 1) Model declaration section: this part requires provision of global variables that are common to the populations in play. This script starts with a *declare* and ends with an *end declare* statement. Between the declare and end declare statements, model declaration statements to initialize the variables such as: *environmental volume*, *mutation probability*, *number of specific* and *non-specific substrates*, as well as population declaration statements to initialize the population specific variables such as: *population size*, *mortality rate*, *flow rate*, *adding new gene of specific* and *non-specific substrates* are listed.

- 2) Calculation section: starts immediately after end declare statement and contains different model calculation statements. For example, the possibility of *horizontal gene transfer (HGT)*, *gene loss* and *number of generations* a model runs are among several statements in this section of the script.

The experimental models, Model 1 and Model 2 are listed in tabular format, alongside the input substrates they take and the output substrates they produce. The given list of input and output (both SS and NS) are chosen in such a way to see if the different populations exhibit some level of cooperation or competition based on their type and concentration of inputs taken or outputs produced. The dependence of different populations among one another and the competition that may exist to get the needed substrates in their ecosystem is supposed to play a vital role in the overall state of the community. In circumstances where symbiotic relationships exist, the presence of keystone species can maintain the balance and survival of the populations; hence it is very beneficial to the microbial community.

Once the number and concentration of substrates and all the populations specific parameters are declared, the from the beginning to the end of the generations, the populations use only these substrates. Introduction of new substrates from outside initially created community is not implemented in the HEC software tool.

The first model in this paper consists of three different populations with different specific and non-specific substrates composition.

Table 1. Model 1 population composition

	Population 1	Population 2	Population 3
Input	ABCD	ACDF	ADFG
Output	FGHI	GHIJ	HIJK

Table 1 shows the composition of the first model's list of specific and non-specific substrates together with the number of populations in the ecosystem. In this model, all populations have the same amount of substrate concentration with all other characteristics such as their reproduction rate, mortality rate, flow rate, and environmental volume being the same.

List of non-specific substrates: AD

List of specific substrates: BCFGHIJK

As mentioned in the first paragraph of the methodology section, the representation for the non-specific substrates in the HEC script will be:

GENE_N=1, GENE_N=2

For the specific substrates, their representation in the HEC script will be:

GENE_S=1, GENE_S=2, GENE_S=3, GENE_S=4, GENE_S=5, GENE_S=6, GENE_S=7, GENE_S=8

GENE_P represents the substrates produced in the environment

The second model in this paper also consists three different populations with different specific and non-specific substrates.

Table 2. Model 2 population composition

	Population 1	Population 2	Population 3
Input	ABCD	AbcDFG	AbcDHI
Output	FGHI	fgH	iJK

Table 2 shows the composition of the second model's list of specific and non-specific substrates together with the number of population in the ecosystem. In this second experimental model, like the first all the populations have the same amount of substrate concentration with all other characteristics such as their reproduction rate, mortality rate, flow rate, and volume being the same. However, the allele values for the substrates represented with upper case is 3.5, and those represented with smaller case is 1.5. Substrates b, c, f and g of Population 2 as well as b, c, i of Population 3 have allele values of 1.5 each. The amount of input of two of the four Non-specific Substrates (NS) i.e. b and c is lower in both Population 2 and Population 3 compared to Population 1. This experiment emphasizes on the study of behavior of different populations that contain different amount of NS.

List of non-specific substrates: ABCD

List of specific substrates: FGHIJK

As mentioned earlier, the first part of each model excludes the possibility of HGT and gene loss and experiments the populations' stimulus towards the availability of substrates in their environment. The second part of each model introduces the possibility of HGT and gene loss property in the declaration section of the script with the rest of the parameters (global and population specific) being the same as in the first part. The close association of microbes enhances the probability of horizontal gene transfer which enables organisms to develop different new features over the course of their generation (Gogarten, 2005). HGT is a vital mutational process that is implemented in HEC software. Hence, in the second part each of the given two models the populations will have the ability to use this vital process to develop novel features to adapt the availability of the resources in the environment.

4.1. Experiment 1.1

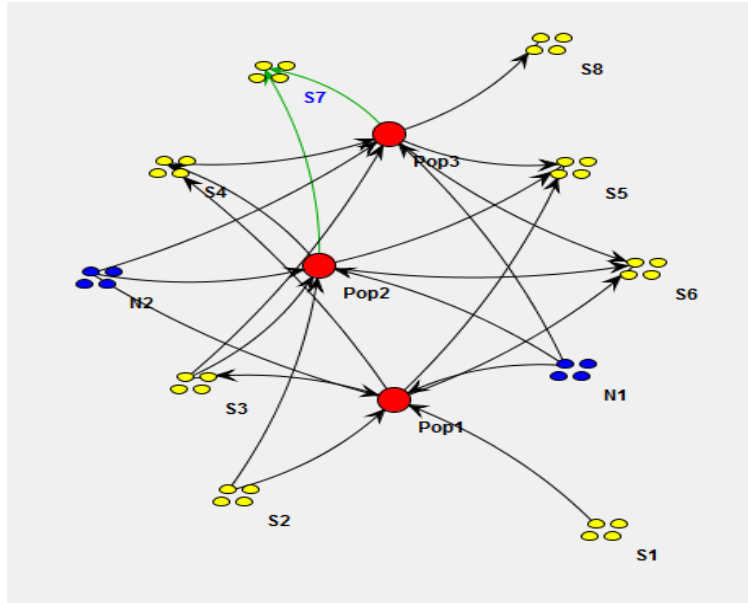


Figure 10. Scheme of Model 1 substrates at the end of the given evolution.

Figure 10 shows the interaction between different populations and their substrate intake as well. The blue circles of a scheme represent the non-specific substrates available in the environment N1, N2, which are representations for the non-specific substrates explained in Model 1 (A and D). The yellow circles are representation of specific substrates present in the environment S1, S2, ..., S8. These were also mentioned as the 8 specific substrates of Model 1 available in the environment (BCFGHIJK). The red circles represent the number of population pop1, pop2 and pop3. The arrows represent the intakes of specific and/or non-specific substrate by a population.

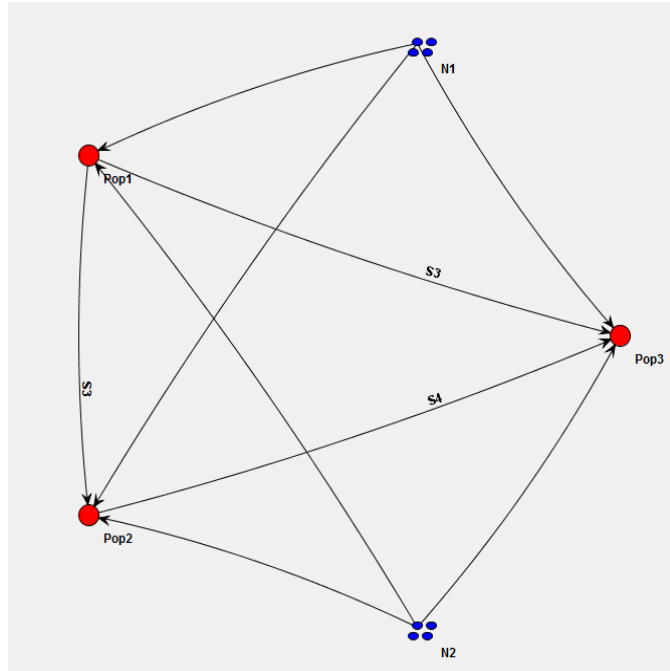


Figure 11. Trophic rings of Model 1 at the end of the given evolution.

Symbiotic relationships, which are special types of interaction between species, are widespread in ecosystems and especially specific to prokaryotic communities. These communities are characterized by a complicated trophic chain with extent of closeness. Specific traits of prokaryotes, such as diet, small cell size and dispersal limitation, lead to organism's inability to escape from the surrounding environment quickly and to the biogenic desert problem in which organisms that inhabit the center of a high-density population starve due to lack of nutrients. Large majority of prokaryotes in nature exist as a part of communities with a complicated structure whose common metabolite pool often forms a complete cycle. The presence of such cycles optimizes community members' metabolism (Lashin, 2014). This can be clearly explained through Figure 11, which shows the trophic rings of Model 1 in which the metabolic products of one population can be utilized by another population. As shown in Table 1, specific substrate F, which is a metabolic product of Population 1 is used as an input by Population 3. This process is represented

in Figure 11 through the edge going from Population 1 to Population 3 with a label S3 which is a variable representing specific substrate F.

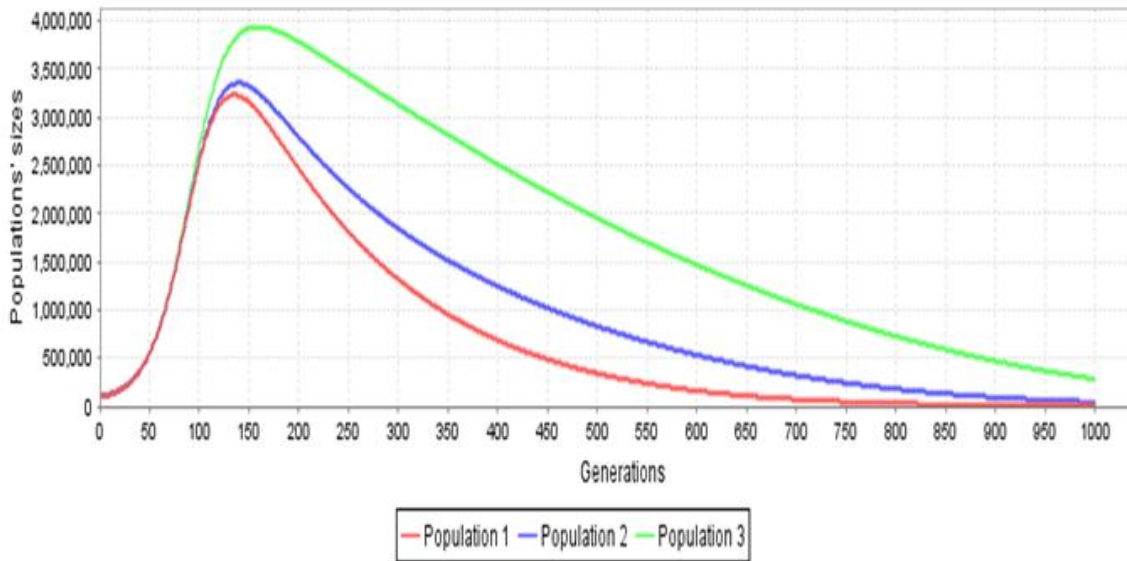


Figure 12. Population size dynamics of Model 1. The population size at the beginning and the end of the evolution shown with both Population 1 and Population 2 become extinct and ultimately die after the 1000th generation.

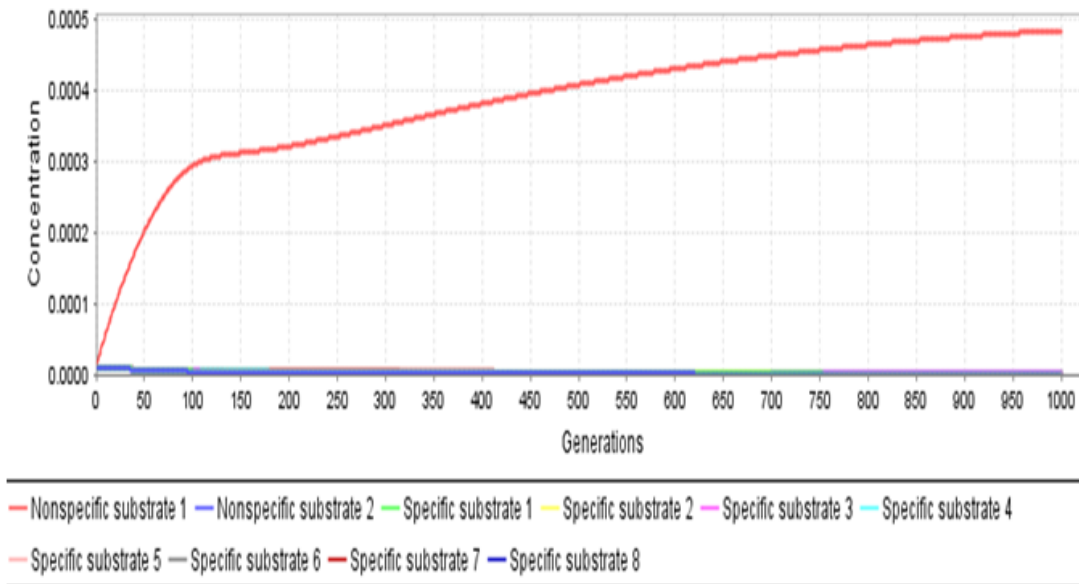


Figure 13. Substrate (NS and SS) concentration dynamics of Model 1.

For this model, all the generations start with the list of specific and non-specific concentrations given in the first column of Table 3. However, with increasing generations all the

substrates, except for NS 1 (Non-specific Substrate A), kept on being depleted resulting in the death of all the three populations. Figure 13 shows the dynamics of both NS and SS throughout the entire evolution. Overall, NS 1 has a higher concentration level compared to the rest of NS and SS. However, regardless of its abundance, all the three populations cannot sustain the entire evolution and eventually die as shown in Figure 12. The resulting concentration of both the NS and SS is given in the second column of Table 3.

Table 3. Prior and Posterior concentration of NS and SS of Experiment 1.1.

Starting concentration of NS and SS	Final concentration of NS and SS
substrates_ns = 1e-5, 1e-5	substrates_ns = 4.840374e-4, 6.252886e-11
substrates_ss = 1e-5, 1e-5, 1e-5, 1e-5, 1e-5, 1e-5, 1e-5, 1e-5	substrates_ss = 2.8760216e-10, 1.6455386e-10, 1.138414e-8, 7.430727e-8, 3.7280655e-7, 3.7280655e-7, 3.591879e-7, 2.9192617e-7

Table 3 shows a snippet of input and output script showing the concentration of input and output substrates of Experiment 1.1. The table shows the concentration of NS and SS at the start of the generation (values on the left-hand side) and at the end of the generation (values on the right-hand side). As shown in the table, the amount of input taken from the three populations has decreased as generations goes on, hence all the populations are on the verge of extinction at around 1000th generation.

4.2. Experiment 1.2

Unlike the schema and trophic rings of Model 1 shown in Figure 10 and Figure 11, respectively, the schema and trophic rings of this experiment shown in Figure 14 and Figure 15, respectively, is more intricate. The number and concentration of inputs is still the same in both the experiments; however, the inclusion of Horizontal Gene Transfer (HGT) in this experiment results in creating an enhanced population that takes the minimum concentration of NS and/or SS.

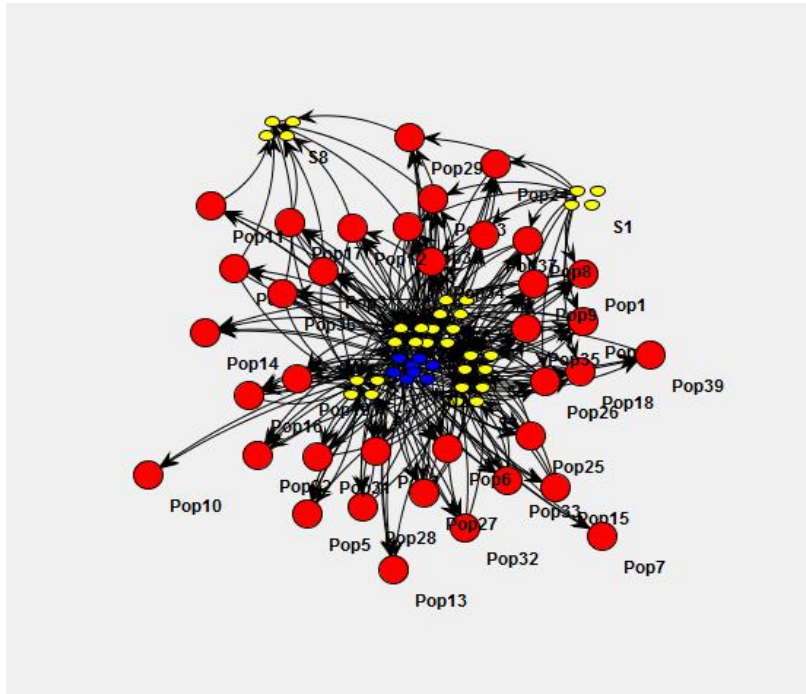


Figure 14. Scheme of Model 1 substrates at the end of the given evolution (HGT included).

The complex directed network in Figure 14 shows the input and output substrates of the original populations introduced at the start (pop1, pop2 and pop3) and the newly generated populations (pop4 through pop39). The indegrees of a population represent the number of input substrates taken by the population and the outdegrees represent the number of output substrates produced by the population.

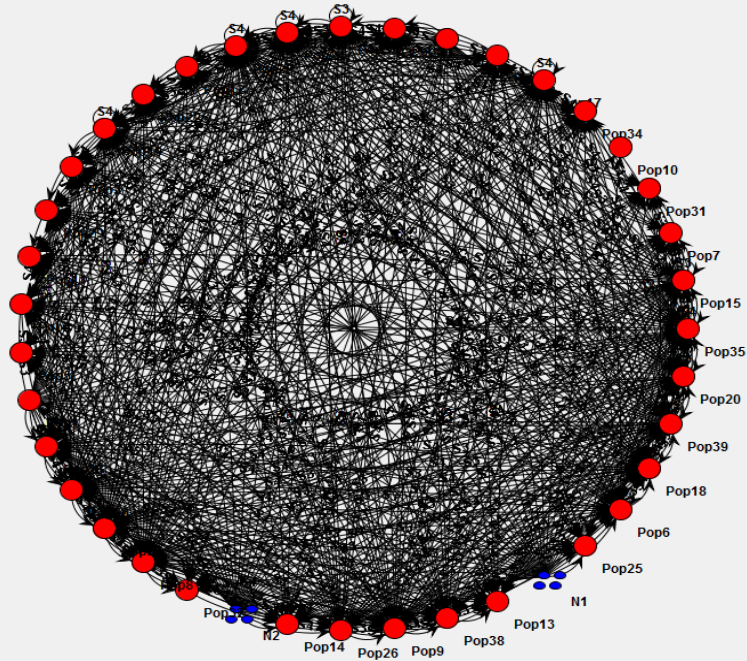


Figure 15. Trophic rings of Model 1 at the end of the given evolution (HGT included).

The intricate web shown in Figure 15 represents the output substrates of a given population taken as input by another population. The large number of edges are indications of the high dependency that exists between different number of populations. In the trophic rings network, the indegree of a population represents the number of NS taken by the population as an input, and the outdegree of a population represent its output that are taken as input by another population.

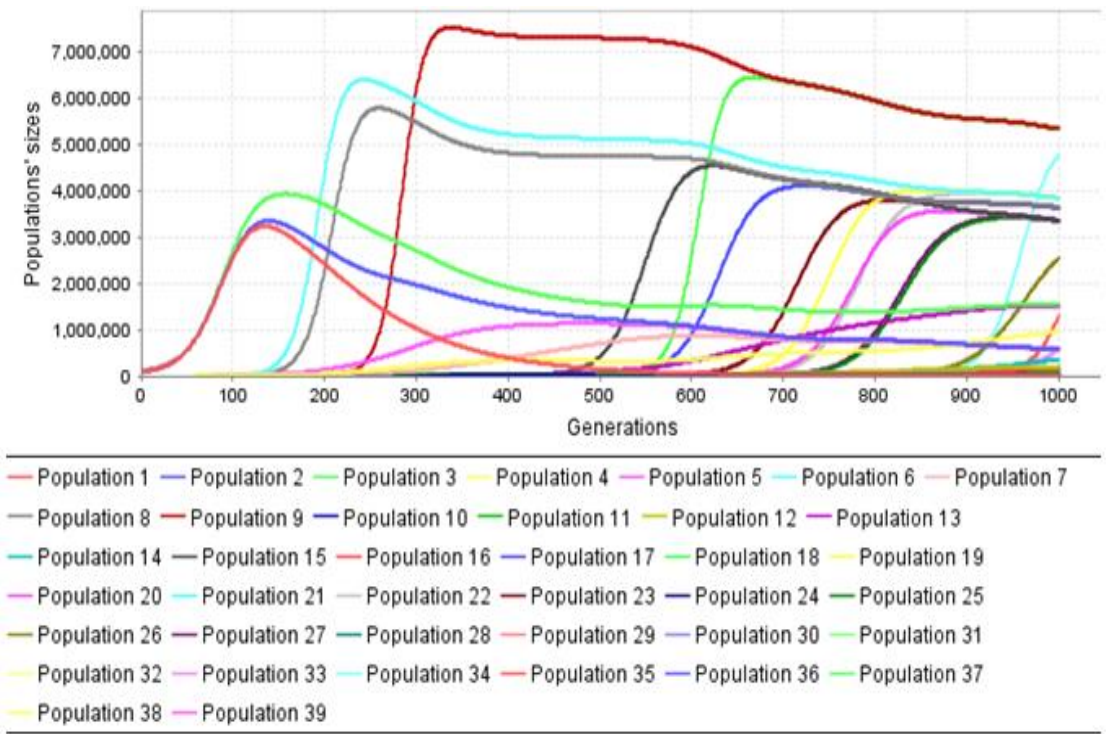


Figure 16. Population size dynamics of Model 1 (HGT included).

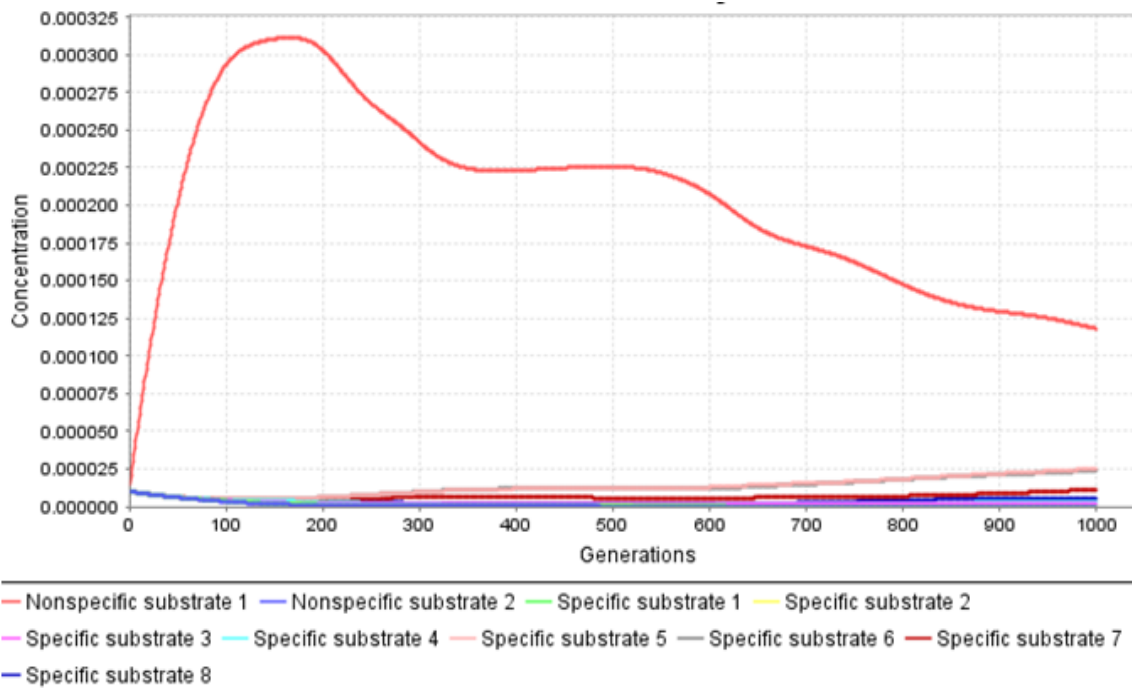


Figure 17. Substrate (NS and SS) concentration dynamics of Model 1 (HGT included).

As shown in Figure 16, when the HGT is enabled in the script of Model 1, although we started with only 3 populations we end up with an overall 39 populations. All the population specific variables like mortality rate, flow rate, synthesis and utilization of specific and non-specific substrates remain the same at the end of the simulation as in the initial simulation. Regardless of the immutability of these values, new populations emerged with increase in the number of generations. The emergence of new populations was due to organisms' ability to utilize the most abundant substrates found in their environment and/or their ability to utilize their own output as an input in the next generation. Less competition of certain substrates also benefits survival of certain organisms that use these substrates. Figure 17 shows the overall number and concentration of NS and SS after HGT was introduced to the script. The following two paragraphs explain trophic chains and status of particularly Population 4 and Population 9.

Population 4:

Population 4 takes three inputs in which two are the most abundantly used non-specific substrates, A and D, and a third specific substrate, G. This population in turn produces three substrates in which two of them, J and K, are among the least produced and least needed substrates. Hence, as soon as this population got created around 50th generation, it continues its modest survival tendency until the end of the 1000th generation.

Population 9:

This population takes as an input NS substrates A and D and SS B, C, F and G. regardless of many substrate requirements for this population, the inputs A, D, C and G are found in abundance in the environment; hence, after population 9 came to existence around the 200th generation (shown in Figure 16), after only few generations its size increased to around seven

million and continues to dominate the entire population until the end of the 1000th generation where the population size drops by around a million.

4.3. Experiment 2.1

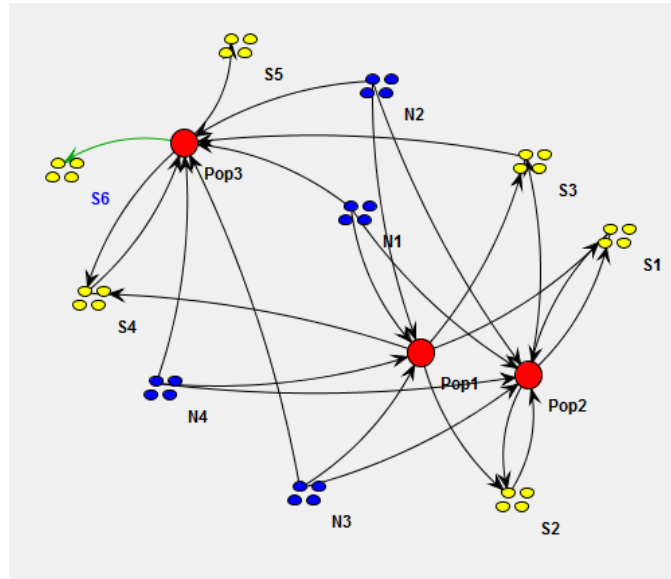


Figure 18. Scheme of Model 2 substrates at the end of the given evolution. Yellow circles represent non-specific substrates, blue circles represent specific substrates and red circles represent. Arrows represent the input and output of specific substrates and/or non-specific substrate by each population.

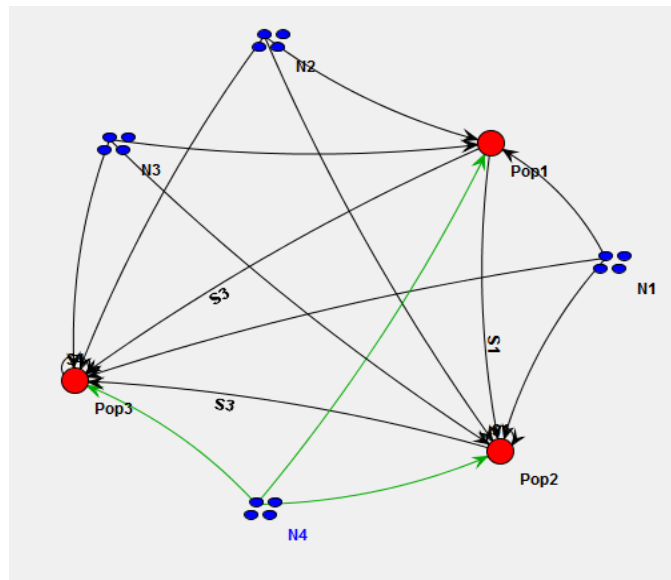


Figure 19. Trophic rings of Model 2 at the end of the given evolution. Arrows going from one population to other represent the specific substrates that are output of a one population utilized as an input by other population. Arrow going from blue dots to a population represent that the non-specific substrates is used by that population.

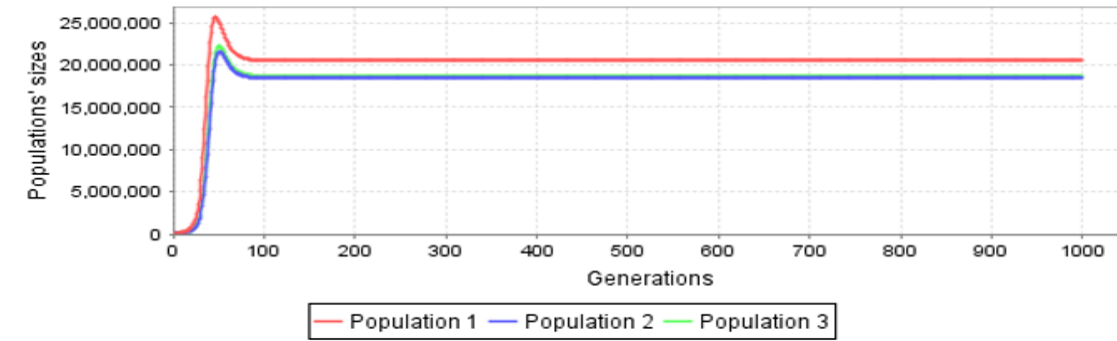


Figure 20. Population size dynamics of model 2.

As shown in Figure 20, the three populations managed to survive the entire generation. Apart from population 1, which depends only on the NS present in the environment, the other two populations need additional specific substrates as inputs. In this experiment, at the end of the simulation, population 1 managed to survive throughout the entire generation solely because its inputs came from the environment and doesn't depend on substrates produced from the other two populations. According to Figure 21, the concentration of NS 1, one of the basic substrates needed for population 1 to survive, is very large throughout the entire generation which gives the population better chance of survival. Not only population 1 but also Population 2 and population 3 survived the entire generation as well. The reasons for the survival of population 2 and population 3 are because:

1. Both population 2 and 3 utilize the abundant substrates found in the environment as their input. Utilizing abundant resources from the environment gives them a good chance of survival.
2. Both population 2 and 3 use their outputs from previous generation as an input for the current generation. This avoids the competition of vital resources and improves their chance of survival.

3. Population 2 and 3 greatly benefit from the substrates that are produced from population 1. Products of population 1, substrates F and G are used as inputs by population 2 while substrates H and I are used as inputs by population 3.

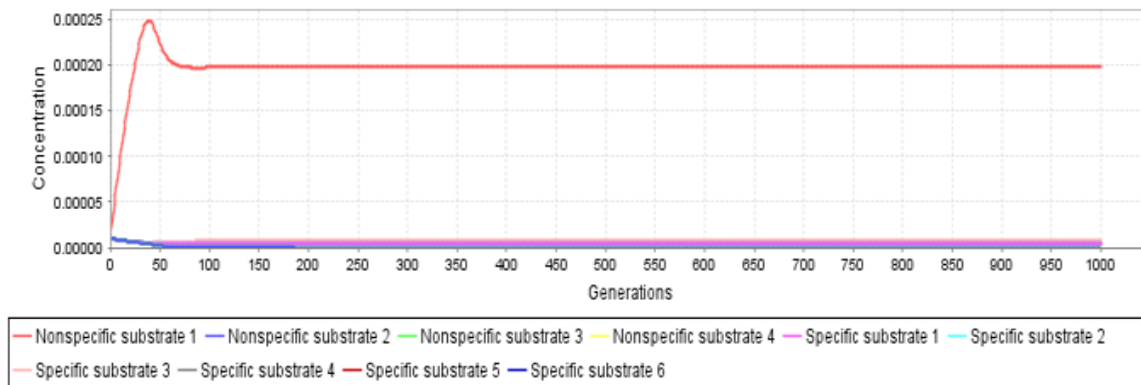


Figure 21. Substrate (NS and SS) concentration dynamics of Model 2.

It is true that the three mentioned factors equip population 2 and population 3 with the necessary inputs and give them a better chance of survival in the ecosystem. However, the third factor is more vital because population 1 behaves as a keystone species in the ecosystem. The products of this population are important inputs to the other two populations (i.e. population 2 and population 3). Moreover, as shown in figure 22, removal of population 1 from the original model disrupts the ecosystem by creating shortage of supply of non-specific inputs H and I. As a result, the size of population 3 declines which can ultimately result in the collapse of the population. As shown in figure 22 population 2 (which represents population 3 before removal of population 1) only decreases in size until 100,000 generation. Running the simulation beyond that can give a clear picture of what happens to both populations. Hence, another experiment using small-scale representation of both populations showed that population 3 ultimately collapses when population

1 is removed from the environment. Hence, it can be concluded that population 1 is playing the role of keystone species as it balances the survival of other populations in the ecosystem.

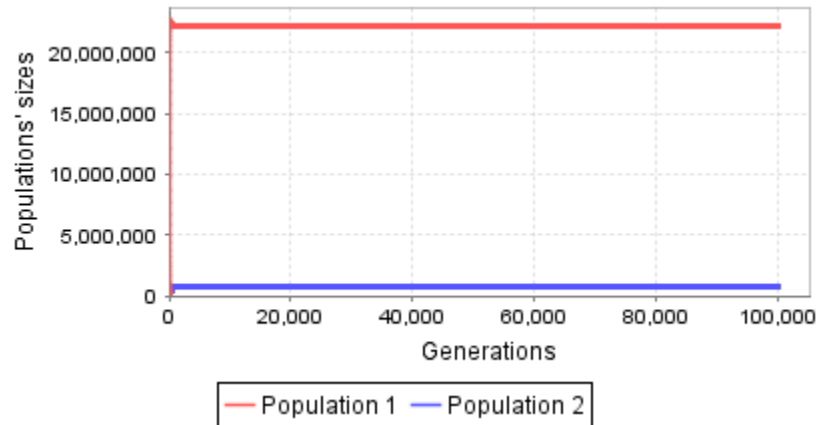


Figure 22. Population size dynamics after removal of population 1 from the original Model 2.

4.4. Experiment 2.2

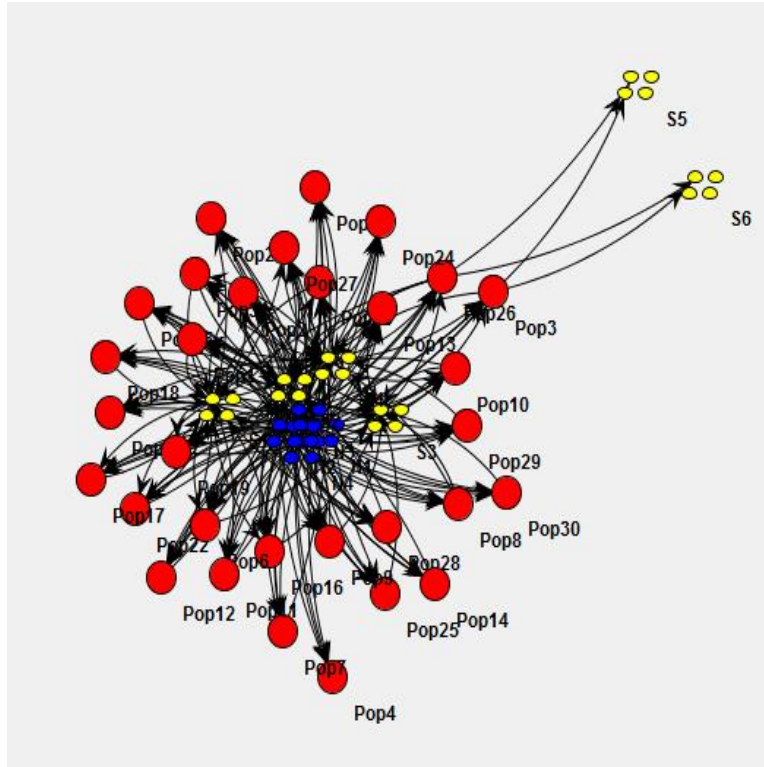


Figure 23. Scheme of Model 2 substrate at the end of the given evolution (HGT included).

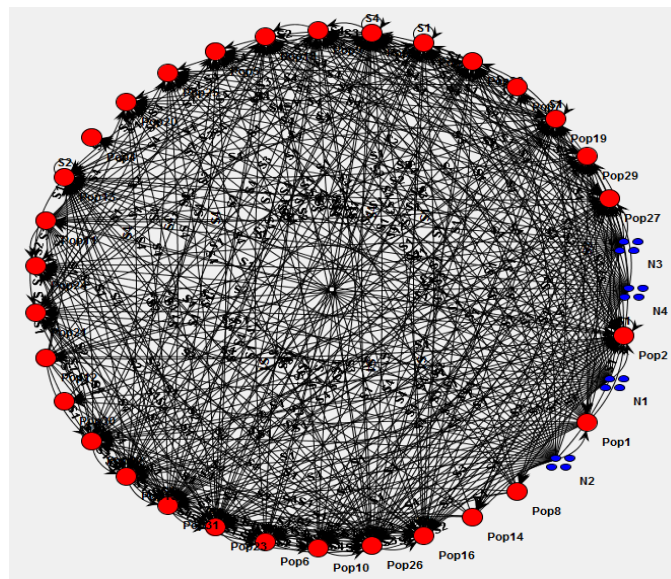


Figure 24. Trophic rings of Model 2 at the end of the given evolution (HGT included).

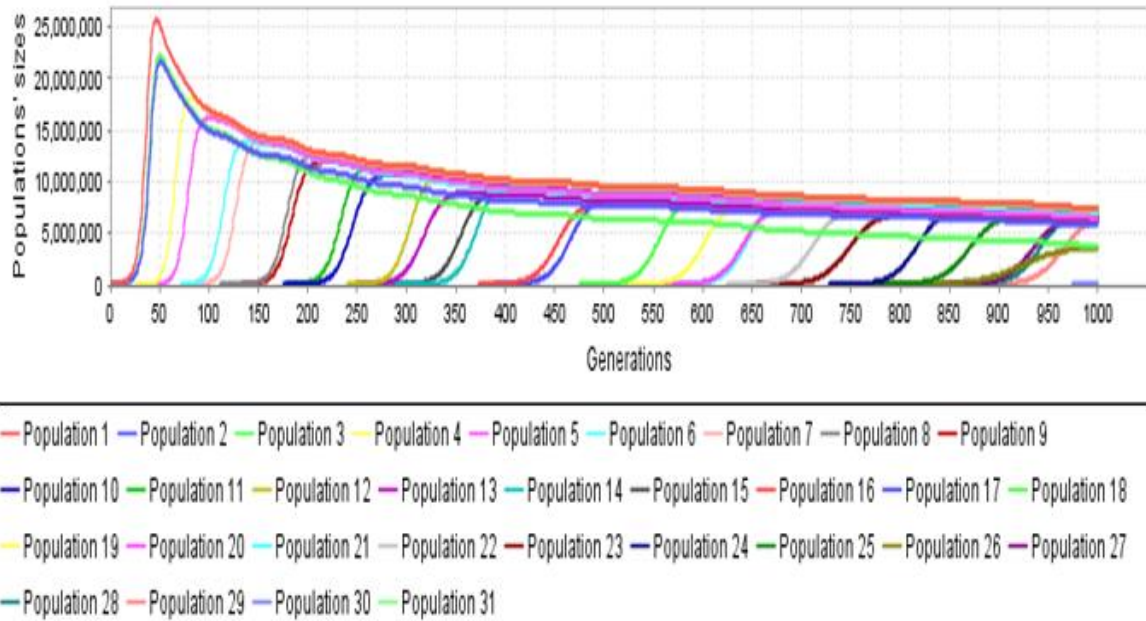


Figure 25. Population size dynamics of Model 2 (HGT included).

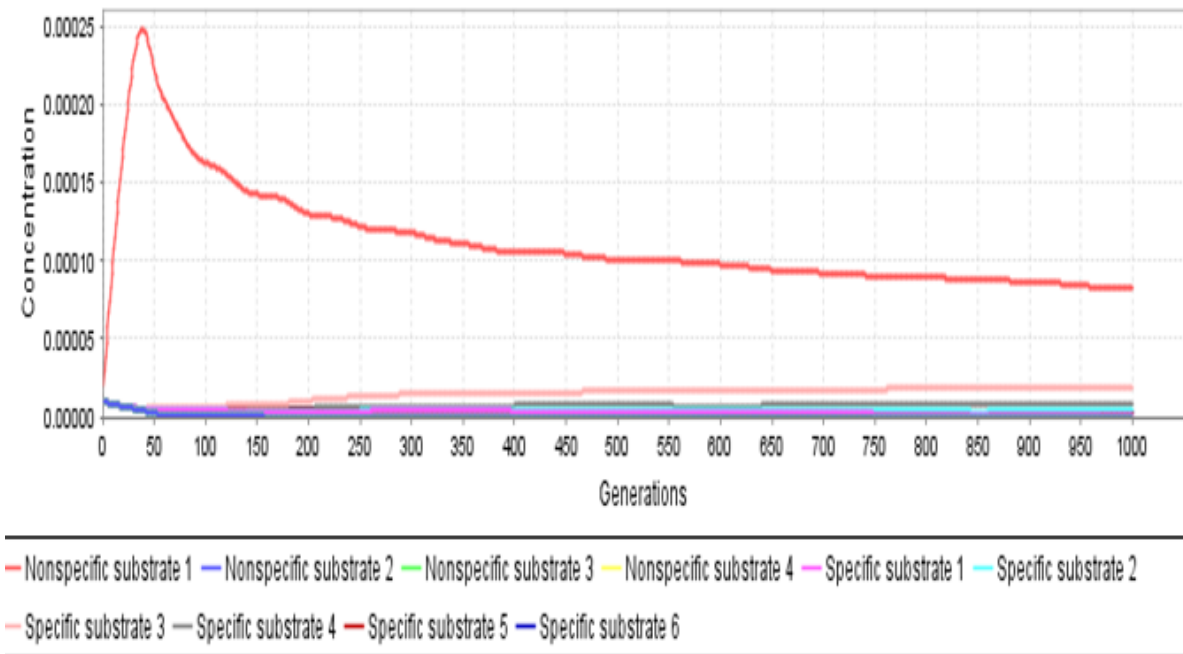


Figure 26. Substrate (NS and SS) concentration dynamics of Model 2 (HGT included).

After introduction of HGT and gene loss to model 2, unlike the population size dynamics of Model 1 with HGT, a similar survival trend between the initial three populations and the emerging ones can be observed in Figure 18. The difference in the concentration levels of the inputs has created a similarity of size between the different populations by eliminating competitive populations and supporting mutualism in the community. Hence regardless of the difference of concentration of NS in this model, populations have a better survival rate and relatively similar population sizes. Introduction of HGT helped populations to specialize on their input by focusing on the abundant substrates found in the environment and through mutualism among one another. As shown in Figure 25, after introduction of HGT to Model 2, a total of 28 populations were created in addition to the initial three populations. Organisms that have common substrates as well as using and producing same substrates are categorized under same population.

In HEC, the model environment is a bounded flow system of fixed volume containing all populations and substrates. Hence, besides the declared number of substrates and their concentration, additional substrates cannot be introduced from outside the environment. nonspecific substrates are reproduced by the in-flow; whereas specific substrates are running into the environment through consumption and production. HEC simulation tool does not implement the prioritization technique of substrates for organisms to consume them. Hence, organisms may consume and utilize substrates in non-prioritized fashion and then secrete products into the environment, which inversely can be used by other organisms as substrates. However, unused substrates can arise in the ecosystem, which remains buried or removed by the flow. Furthermore, accumulations of non-specific substrate are often formed when the inflow of non-specific substrates are used in miniscule amount compared to the other substrates.

5. CONCLUSION

A keystone species has a huge impact in maintaining the structure and balance of an ecological community (Scott, 1993). The HEC software was helpful to create different models with varying parameters and analyze their effect in the environment. From the two simple models explained in this paper, it can be argued that among the presented substrates there were some influential populations that hugely influenced their environment and survival of the other populations.

As explained in Experiment 1.1, when different populations use a substrate at an equal amount with no population producing that substrate, then the populations may die as the generations go on. Populations have the worst chance of survival when there is competition for vital substrates between the populations and when there is strong inter-dependence within the population. In the absence of keystone species, the existence of both competition and inter-dependence, with no one in the ecosystem to balance it, can lead to the demise of the community. Moreover, the usage of a population's products as an input by any population can prolong the existence of some populations. With the horizontal gene transfer (HGT) property introduced within the populations, some tend to evolve and produce new populations that rely mainly on the abundant substrates in the environment or do not heavily depend on the ultimately diminishing substrates. Through HGT, organisms can enhance their survival rate in a competitive and resource-scarce ecosystem. It was also noted that populations that take their own output as input in the next generation have a better chance of survival.

From the above experiments it was shown that dominant populations mostly depend on the substrates that are abundant in the community. Furthermore, they avoid depending on substrates produced from other population's, thereby having less inter-dependence with other populations.

These factors can increase the chance of survival of the population in an ecosystem. However, absorption of essential substrates by these dominant species can lead to the depletion of resources and ultimately lead to the collapse of the ecosystem. The presence of keystone species, as seen in Model 2, balances the substrate utilization that and enhances the chance of survival of all the populations in the ecosystem. As shown in figure 22, the removal of keystone species (population 1) results in the demise of the inferior population.

Agent-based modeling simulations have led to a novel approach of modeling systems composed of autonomous, interacting agents (Charles, 2009). They have been used in an array of fields like health, social sciences and economics in decision making and understanding complex phenomenon. The Haploid Evolutionary Constructor tool captures and explains ideas related to microbial ecosystem. However, the fact that it does not allow the production of non-specific substrates is a big shortcoming. Moreover, HEC doesn't allow the introduction of new substrates outside the previously declared setting.

The models described above are simple representations of real-life examples that exhibit certain symbiotic relationships. In the presence of actual microbial samples, it would be interesting to continue studying and analyzing samples to understand the behavior of microbial populations competing for shared substrates and identifying keystone populations that support the survival of the ecosystem. Furthermore, the above models run until 1000 generation only because of the limit of the computing capacity. Understanding the behavior of these models beyond 1000 generations would be very helpful to have a better picture of the populations but that requires high-performance computing (HPC).

REFERENCES

- Abdou, M., Hamill, L., & Gilbert, N. (2012). Designing and Building an Agent-Based Model. In C. A. Heppenstall AJ, Agent-Based Models of Geographical Systems (pp. 141-165). Springer.
- An G., F. B. (2016). Optimization and Control of Agent-Based Models in Biology: A Perspective. PMC, 63-87.
- An, D.-M. M. (2013). Agent-based models in translational systems biology. PMC, 159-171.
- Castiglione. (2006). Agent Based Modeling. Scholarpedia.
- Charles, M. M. (2009). Agent-based Modeling and Simulation. Proceedings of the 2009 Winter Simulation Conference (WSC) (pp. 86-98). Austin: IEEE.
- Epstein, J. M., & Axtell, R. (1996). Social Science from the Bottom Up. Washington, D.C.: The Brookings Institute.
- Garcia, R. (2005). Uses of Agent-Based Modeling in Innovation/New product Development Research. Product Innovation Management, 380-389.
- Gogarten, J. T. (2005). Horizontal gene transfer, genome innovation and evolution. Nature Publishing Group, 679-687.
- Hatna, E. B. (2012). The Schelling Model of Ethnic Residential Dynamics: Beyond the Integrated - Segregated Dichotomy of Patterns. Journal of Artificial Societies and Social Simulation, 6.
- Lashin, S. M. (2014). Evolution, Biodiversity and Ecology in Microbial Communities: Mathematical Modeling and Simulation with the “Haploid Evolutionary Constructor” Software Tool. In Oscar G, Biodiversity - The Dynamic Balance of the Planet (pp. 111-136). InTech.
- Murphy, W. D. (2008, April). An agent based approach to modelling microbial ecosystems. Retrieved from ERCIM News: <https://ercim-news.ercim.eu/en73/rd/an-agent-based-approach-to-modelling-microbial-ecosystems>
- Politopoulos. (2007, September 11). Review and Analysis of Agent Based Models in Biology. Retrieved from Web Archive: <https://web.archive.org/web/20110727085744/http://www.csc.liv.ac.uk/research/techreports/tr2007/ulcs-07-021.pdf>
- Rahman, A. S. (2007). An Analysis To Wealth Distribution Based On Sugarscape Model In An Artificial Society. International Journal of Engineering, 211-223.

Scott, L. M. (1993). The keystone species concept in ecology and conservation. *Biosciences*, 219-224.

Shashkova, P. T. (2016). Agent Based Modeling of Human Gut Microbiome Interactions and Perturbations. *PLOS*.

Van Dyke Parunak, S. R. (1998). Agent Based Modeling vs Equation Based Modeling: A case study of users' guide. *Proceedings of Multi-agent systems and Agent-based Simulation* (pp. 10-25). Berlin: Springer.

APPENDIX A. HEC SCRIPT FOR MODEL 1 (HGT NOT INCLUDED)

```
DECLARE
volume = 0.01
nonspec = 2
spec = 8
flow = 0.01
substrates_ns = 1e-5, 1e-5
substrates_ss = 1e-5, 1e-5, 1e-5, 1e-5, 1e-5, 1e-5, 1e-5, 1e-5
comsub = 5e-4

POP=1
  SIZE=1e+5
  INCREASER= rubel
  K_DEATH=1E-8
  K_FLOW=0.02
  CPROD=1E+4
  CCONS_NSP=1E+7
  CCONS_SP=1e+7
  // genome
  GENE_N=1;allele_values: 3.5; allele_concentrations: 1.0
  GENE_N=2;allele_values: 3.5; allele_concentrations: 1.0
  GENE_S=1;allele_values: 3.5; allele_concentrations: 1.0
  GENE_S=2;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=3;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=4;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=5;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=6;allele_values: 3.5; allele_concentrations: 1.0
END POP(1)

POP=2
  SIZE=1e+5
  INCREASER= rubel
  K_DEATH=1E-8
  K_FLOW=0.02
  CPROD=1E+4
  CCONS_NSP=1E+7
  CCONS_SP=1e+7
  // genome
  GENE_N=1;allele_values: 3.5; allele_concentrations: 1.0
  GENE_N=2;allele_values: 3.5; allele_concentrations: 1.0
  GENE_S=2;allele_values: 3.5; allele_concentrations: 1.0
  GENE_S=3;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=4;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=5;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=6;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=7;allele_values: 3.5; allele_concentrations: 1.0
END POP(2)
```



```
POP=3
  SIZE=1e+5
  INCREASER= rubel
  K_DEATH=1E-8
  K_FLOW=0.02
  CPROD=1E+4
  CCONS_NSP=1E+7
  CCONS_SP=1e+7
  // genome
  GENE_N=1;allele_values: 3.5; allele_concentrations: 1.0
  GENE_N=2;allele_values: 3.5; allele_concentrations: 1.0
  GENE_S=3;allele_values: 3.5; allele_concentrations: 1.0
  GENE_S=4;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=5;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=6;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=7;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=8;allele_values: 3.5; allele_concentrations: 1.0
END POP(3)

END DECLARE

STOCHASTIC=1000
```

APPENDIX B. HEC SCRIPT FOR MODEL 2 (HGT NOT INCLUDED)

```
DECLARE
volume = 0.01
nonspec = 4
spec = 6
flow = 0.02
substrates_ns = 1e-5, 1e-5, 1e-5, 1e-5
substrates_ss = 1e-5, 1e-5, 1e-5,1e-5, 1e-5,1e-5
comsub = 5e-4

POP=1
  SIZE=1e+5
  INCREASER= rubel
  K_DEATH=1E-8
  K_FLOW=0.02
  CPROD=1E+4
  CCONS_NSP=1E+7
  CCONS_SP=1e+7
  // genome
  GENE_N=1;allele_values: 3.5; allele_concentrations: 1.0
  GENE_N=2;allele_values: 3.5; allele_concentrations: 1.0
  GENE_N=3;allele_values: 3.5; allele_concentrations: 1.0
  GENE_N=4;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=1;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=2;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=3;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=4;allele_values: 3.5; allele_concentrations: 1.0
END POP(1)

POP=2
  SIZE=1e+5
  INCREASER= inhibitory
  K_DEATH=1E-8
  K_FLOW=0.02
  CPROD=1E+4
  CCONS_NSP=1E+7
  CCONS_SP=1e+7
  // genome
  GENE_N=1;allele_values: 3.5; allele_concentrations: 1.0
  GENE_N=2;allele_values: 1.5; allele_concentrations: 1.0
  GENE_N=3;allele_values: 1.5; allele_concentrations: 1.0
  GENE_N=4;allele_values: 3.5; allele_concentrations: 1.0
  GENE_S=1;allele_values: 3.5; allele_concentrations: 1.0
  GENE_S=2;allele_values: 3.5; allele_concentrations: 1.0
  GENE_P=1;allele_values: 1.5; allele_concentrations: 1.0
  GENE_P=2;allele_values: 1.5; allele_concentrations: 1.0
```

```
    GENE_P=3;allele_values: 3.5; allele_concentrations: 1.0
    //GENE_P=4;allele_values: 3.5; allele_concentrations: 1.0
END POP(2)
```

```
POP=3
```

```
    SIZE=1e+5
    INCREASER= inhibitory
    K_DEATH=1E-8
    K_FLOW=0.02
    CPROD=1E+4
    CCONS_NSP=1E+7
    CCONS_SP=1e+7
    // genome
    GENE_N=1;allele_values: 3.5; allele_concentrations: 1.0
    GENE_N=2;allele_values: 1.5; allele_concentrations: 1.0
    GENE_N=3;allele_values: 1.5; allele_concentrations: 1.0
    GENE_N=4;allele_values: 3.5; allele_concentrations: 1.0
    GENE_S=3;allele_values: 3.5; allele_concentrations: 1.0
    GENE_S=4;allele_values: 3.5; allele_concentrations: 1.0
    GENE_P=4;allele_values: 1.5; allele_concentrations: 1.0
    GENE_P=5;allele_values: 3.5; allele_concentrations: 1.0
    GENE_P=6;allele_values: 3.5; allele_concentrations: 1.0
END POP(3)
```

```
END DECLARE
```

```
STOCHASTIC=1000
```