Spring 2014

# Speciation dynamics of an agent-based evolution model in phenotype space

Adam David Scott



Department: Physics

# SPECIATION DYNAMICS OF AN AGENT-BASED EVOLUTION MODEL IN PHENOTYPE SPACE

by

ADAM DAVID SCOTT

A DISSERTATION

Presented to the Faculty of the Graduate Schools of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

and

UNIVERSITY OF MISSOURI AT ST. LOUIS

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

PHYSICS

2014

Approved by
Sonya Bahar, Advisor
Eric Majzoub
Paul Parris
Thomas Vojta
Istvan Kiss

## ABSTRACT

This dissertation is an exploration of phase transition behavior and clustering of populations of organisms in an agent-based model of evolutionary dynamics. The agents in the model are organisms, described as branching-coalescing random walkers, which are characterized by their coordinates in a two-dimensional phenotype space. Neutral evolutionary conditions are assumed, such that no organism has a fitness advantage regardless of its phenotype location. Lineages of organisms evolve by limiting the maximum possible offspring distance from their parent(s) (mutability, which is the only heritable trait) along each coordinate in phenotype space. As mutability is varied, a non-equilibrium phase transition is shown to occur for populations reproducing by assortative mating and asexual fission. Furthermore, mutability is also shown to change the clustering behavior of populations. Random mating is shown to destroy both phase transition behavior and clustering. The phase transition behavior is characterized in the asexual fission case. By demonstrating that the populations near criticality collapse to universal scaling functions with appropriate critical exponents, this case is shown to belong to the directed percolation universality class. Finally, lineage behavior is explored for both organisms and clusters. The lineage lifetimes of the initial population of organisms are found to have a power-law probability density which scales with the correlation length exponent near critical mutability. The cluster centroid step-sizes obey a probability density function that is bimodal for all mutability values, and the average displays a linear dependence upon mutability in the supercritical range. Cluster lineage tree structures are shown to have Kingman's coalescent universal tree structure at the directed percolation phase transition despite more complicated lineage structures.

## ACKNOWLEDGMENTS

I would like to thank Dr. Sonya Bahar for the opportunity to pursue this work. She has provided me superb guidance and support both in the lab and the "real world". I have been extremely lucky to work with her and to become her friend. I would like to thank Dr. Eric Majzoub, Dr. Paul Parris, Dr. Thomas Vojta, and Dr. Istvan Kiss, for their fine guidance. Each has provided me with helpful insights and very useful suggestions for this work. I would also like to thank Dr. Nevena Marić for many conversations and the resulting collaborations that helped make this dissertation possible.

I would like to thank the graduated students from the Center for Neurodynamics, Dr. Daisuke Takeshita and Dr. Kaushalya Premachandra, for answering so many questions and working with me when I started working on this dissertation. I would also like to thank to Dr. Nathan Dees for laying the groundwork for the research presented here. I would also like to thank Dawn King for being a huge help with this work and for being a great friend.

My parents deserve mountains of gratitude. I would like to thank my dad for giving me many opportunities to explore the world around me and for pointing out objects in the night sky. I would like to thank my mom for reminding me that all things in life are connected, and piecing together the puzzle of those connections is a meaningful and worthy pursuit. They inspired me to study the natural sciences, and I certainly would not be here without them and their support.

To my wife, Staci, I thank her for all of her love and support, even as she has worked on her own dissertation. No matter what, she gives me strength to overcome any challenge. I must also thank, and perhaps apologize to, my wife and close friends for being patient with me as I used them as sounding boards. Without their ear and comments I could not have developed many of the ideas throughout this dissertation. I would be absolutely lost without them.

Finally, I would like to thank the funding source for this work, the James S. McDonnell Foundation.

# TABLE OF CONTENTS

# LIST OF ILLUSTRATIONS

# LIST OF TABLES

# NOMENCLATURE

| Symbol | Description |
|---|---|
| $\mu$ | Mutability |
| $\kappa$ | Competition Limit |
| $\delta$ | Random Death Probability |
| $c_b$ | Offspring Organism Coordinate |
| $c_r$ | Reference Organism Coordinate |
| $c_m$ | Mate Organism Coordinate |
| L | Linear Landscape Size |
| R | Nearest-Neighbor Index |
| s | Number of Organisms in a Cluster |
| $\Delta$ | Off-Critical Measure |
| t | Time |
| $\alpha$ | Critical Exponent - Population Density Decay |
| $\nu_{\parallel}$ | Critical Exponent - Correlation Time |
| $\beta$ | Critical Exponent - Population Density |
| z | Critical Exponent - Dynamical Exponent |
| $\nu_{\perp}$ | Critical Exponent - Correlation Length |
| $\rho$ | Population Density |
| $\xi_{\parallel}$ | Correlation Time |
| $\xi_{\perp}$ | Correlation Length |
| P | Survival Probability |
| $\tilde{\rho}$ | Population Density Universal Scaling Function |
| $P_1$ | Goodness of Fit - Critical-Quench |
| $P_2$ | Goodness of Fit - Finite-Size |
| $C_b$ | Offspring Cluster Coordinate |
| $C_r$ | Parent Cluster Coordinate |
| $\tau$ | Lineage Lifetime |
| F | Maximum Cluster Fitness |

| | |
|---|---|
| $\Gamma$ | Cluster Centroid Step-Size |
| $T_n$ | Time to Most Recent Common Ancestor of n-Lineages |

# 1. INTRODUCTION

> "I made certain very simple, but not very inaccurate, suppositions, concerning average fertility, and I worked to the nearest integer, starting with 10,000 persons, but the computation became intolerably tedious after a few steps, and I had to abandon it."
>
> (Francis Galton, F.R.S. 1875)

The natural world is immensely complicated, and with simplifying models, one can navigate and understand its complexity. With the use of stochastic computational simulations, systems that are inherently noisy, such as biological systems, may be studied with greater ease. Where mathematicians seek the most basic, logical understanding of such systems, biologists seek a grander picture. The modeling approach of a physicist is then to find the happy medium between simplicity and complexity. This happiness is the motivation for the approach taken here.

## 1.1. NEUTRAL THEORY

All of the work within this dissertation operates under the assumption of neutral theory. This is in contrast to the foundation on which modern evolutionary biology is based, the theory of natural selection as described by Charles Darwin (1859). Natural selection arises through differential fitness, and describes how species adapt to changes in their environment and allows them to have a successful continuation of heritable traits. Fisher (1930) ushered the theory of natural selection from studies based only on natural history to statistical modeling. However, it wasn't until the work of Kimura that notions of evolution without selection (without effects of differential fitness on survival) were explored (Kimura and Crow 1964; Kimura 1968, 1983). In his work, Kimura described genetic drift, a neutral theory of evolution that provides a "null hypothesis" for the formation of species by assuming that different genetic mutations have equal fitness benefits for the organisms which carry them. Recently, Hubbell (2001) described a neutral theory of ecology which also predicts clustering of spatial patterns of flora that occur from drift alone and without any species bias about where each plant or tree may

grow within a local area. Both of these neutral models suggest theoretical possibilities, but they do not imply prevalence.

There has been contention regarding possibility vs. prevalence of speciation by genetic drift and of neutral clustering in ecology (Ricklefs 2006), and this is compounded by the rarity of sympatric speciation. Several example species have been shown to have formed from sympatric speciation including sticklebacks (Schluter 1994), snails (Johannesson, Rolan-Alvarez, & Ekendahl 1995), and anolis lizards (Losos, et al. 1998). Even microorganisms such as bacteria have been shown to develop via sympatric speciation (Cadillo-Quiroz et al. 2012). Dieckmann and Doebeli (1999) showed that in an agent-based model of sympatric, assortative mating (explained below) organisms undergoing genetic drift can speciate when competition for resources is included in their dynamics. Their model was predicted to apply to recently colonized habitats, and for trait-biased mating species that rely upon ecological traits such as size or coloring (Dieckmann & Doebeli 1999).

The set of genes found in the DNA of an organism defines its genotype. The physical manifestation of the instruction from a gene is protein expression. The expression of a protein can, by itself or in concert with other proteins, be manifested as an external trait in an organism. The phenotype is then the set of traits which are observed from the genotype. Genetic drift explores the set of possible genotypes, or genotype space. The set of possible phenotypes then describes a phenotype space. Natural selection acts upon phenotypes, therefore a selected phenotype corresponds to a particular genotype. When a mutation occurs somewhere in a genotype, there can be a corresponding mutation to the phenotype. Neutral theory then describes how mutations offer no survival advantage over the original genotype (or phenotype) or any other mutated genotype. The work presented throughout exists in a phenotype space with no consideration of genotype space or physical space.

## 1.2. PHASE TRANSITIONS

The purpose of studying phase transitions is to seek how a control parameter drives fundamental changes in the dynamics of an order parameter which is a measure of the overall behavior of a system. Here, the focus is on continuous, non-equilibrium phase

transitions. “Continuous” refers to the continuous first derivative of the order parameters as the control parameter is varied, and it is accompanied by diverging variations of the order parameter as the control parameter approaches the critical point. “Non-equilibrium” refers to the inability of system to transition equally between phases. In the supercritical state the system is active and fluctuating, whereas, subcritically, the system forever stays in an absorbing, inactive state. There are a variety of dynamical characteristics which go along with such a transition near and at a critical point. These include scale-free behavior of the order parameters (having no characteristic scaling), large variations of the order parameter, and asymptotically long decay times. The overarching objective of this dissertation is to report on the behavior of a specific neutral phenotype evolution model in the presence of a continuous, non-equilibrium phase transition.

## 1.3. BASICS OF THE PHENOTYPE EVOLUTION MODELS

The models used throughout this dissertation are modifications of a previously-described evolution model with rugged and changing fitness landscapes (Dees & Bahar 2010). As in the original model, organisms are described here by independent and arbitrary trait values (coordinates) in a continuous, two-dimensional phenotype space. Interactions of organisms with each other and their environment are based upon four considerations.

**1.3.1. Organism Lifecycle.** First, to reproduce, if one imagines a sexually reproducing organism that is free to choose its mate, it will more likely choose one that has similar traits, referred to as assortative mating (Kondrashov & Shpak 1998; de Cara, Barton, & Kirkpatrick 2008; Otto, Servedio, & Nuismer 2008). The offspring of the mating pair will exhibit a phenotype that is some combination of the parental phenotypes, with the incorporation of an additional amount of variation due to mutation. Note that this type of “blending inheritance” in phenotype space does not imply a blending of genotypes, which cannot occur in a biological system (Ridley 2004). Offspring which share too similar traits are likely to compete for the same set of resources, so when offspring compete, one of them will die. Finally, not all organisms are permitted the opportunity to survive before they can reproduce.

For the model considered here, in contrast to that of Dees and Bahar (2010), at every stage in the lifecycle of organisms, there is no selection preference. The phenotype coordinates of organisms do not convey any advantage or disadvantage in terms of fitness, which is defined as the number of offspring produced, competition, or luck in escaping from random, mortal events. These processes will be detailed in the following sections.

**1.3.2. Clustering and Species.** Since there is no explicit representation of geographic distance in the model, clustering of organisms in the phenotype space corresponds to sympatric speciation. Clustering is roughly based on the biological species concept, in the case of assortative mating and random mating, and on the phenetic species concept in the case of asexual fission. To clarify the biological terminology, sympatric speciation is the formation of species found in the same ("sym-") physical area ("patric"). The biological species concept defines a species as a closed set of reproducing organisms that produce viable offspring. The concept of phenetic species describes bacterial species according to shared or similar phenotypes.

## 1.4. SUMMARY

The role of mutability in controlling phase transition behavior and cluster formation will be studied. In Section 2, observation of phase transitions and clustering behaviors in the case of assortative mating and asexual fission are discussed in contrast with random mating. In Section 3, it is shown that the asexual fission model undergoes a phase transition which belongs to the directed percolation universality class. Statistical distributions and branching dynamics of the lineages of organisms and clusters in the asexual fission model are then discussed in Section 4. In addition, multilevel selection is discussed among possible implications of cluster fitness, and is put into context by possible universal structure measures of average time to most recent common ancestor. Suggestions for future studies branching from this work, such as implications of multilevel selection leading to scale-free selection, explosive increases in biological diversity, and comparisons with biological morphology data are discussed in Section 5.

# 2. OBSERVATIONS OF CLUSTERING AND PHASE TRANSITIONS

## 2.1. INTRODUCTION

Speciation under neutral conditions was initially studied by Kimura in the case of genetic drift (Kimura & Crow 1964; Kimura 1968, 1983). A more recent, equivalent approach to clustering under neutral conditions was introduced by the work of Hubbell in the context of physical clustering in ecology (Hubbell 2001). How selection affects clustering in genotype and phenotype space or in physical space is of great concern to evolutionary biology and ecology, because species are generally identified by such clustering. Removing selection from evolution or from ecological pattern formation introduces important questions about clustering. However, these theories do not imply that they describe the prevalent evolutionary process of speciation, only that selection-free clustering is possible (Hubbell 2001).

Evolutionary systems have been simulated with the goal of determining why clustering occurs for interacting organisms under neutral conditions. In particular, clustering was observed in a neutral model of organisms described by their spatial locations and genotypes (de Aguiar, Baranger, Baptestini, Kaufman, Bar-Yam 2009). Organisms underwent assortative mating by finding mates nearby in both physical and genotype spaces. It was concluded that assortative mating was the essential element to achieve speciation, and both physical and genotype spaces were necessary for clustering. Species abundance curves were produced which matched the predictions of Hubbell's neutral theory of biodiversity, so it was counted as strong support for Hubbell's theory (Banavar & Maritan 2009).

A related mathematical class of clustering systems to the de Aguiar et al. model is branching-coalescing random walks (BCRW). Much of the properties of such models are still being uncovered (Dutta, Panduragan, Rajaraman, and Roche 2013; Cooper et al. 2012; Arthreya & Swart 2005). However, clustering in BCRW models has seen little attention in the context of computational evolutionary biology with at least one exception, that of Hubbell (2005).

Zhang, Serva, and Polikarpov (1990) observed clustering in a population of BCRW agents reproducing by asexual fission. Each organism had the same probability

for reproduction, mimicking neutral conditions. Another fission model, with organisms described by binary digit string genotypes, was studied by Derrida and Peliti (1991). They produced mathematical results such as the probability of observing specific genealogies under neutral conditions.

Meyer, Havlin, and Bunde (1996) produced a modified version of the Zhang et al. model in which they investigated the clustering of organisms under neutral-like conditions: equal birth and death rates. They determined that clustering arose from asymmetry in the birth and death processes. That is, offspring are born near their parents, whereas death kills off organisms regardless of their location in the space. Young, Roberts, and Stuhne (2001) introduced a "Brownian bugs" model of organisms reproducing by fission and concluded again that the minimum criterion for clustering in such models is a spatial asymmetry in the birth and death processes. They also noted that continuous, fluid-like population undergoing a diffusion process with equal birth and death rates does not yield clustering, so only agent-based models can exhibit clustering. In the works of Houchmandzadeh (2002) and Houchmandzadeh and Vallade (2003), mathematical properties of BCRWs were studied directly in the context of Hubbell's neutral theory of biodiversity. Similarly, Lawson and Jensen (2008) studied a neutral model of phenotype evolution with populations undergoing BCRW.

Recently, it was shown that for several different types of rugged fitness phenotype landscapes, mutability could optimize clustering (Dees & Bahar 2010). Here, as below, mutability is defined as the maximum phenotypic distance an organism can be from its parents. Static, rugged landscapes (a grid of fitness values varying between one and four, where fitness is the number of offspring an organism can produce), moving rugged landscapes, and landscapes with feedback (reducing fitness with growing local population density, increasing fitness with lesser local population density) all showed that populations of organisms could cluster. Populations were also observed to undergo non-equilibrium phase transition behavior (transitioning from an absorbing state of extinction to indefinite survival) for slightly smaller values of mutability on each landscape. Near the suspected critical mutability, large variations in the ensemble populations were also observed, with some simulations ending relatively quickly while others thrived for the duration of the simulation.

The work presented here extends the results of the assortative mating phenotype space model in two ways. A neutral fitness landscape is assumed throughout, and phase transition behaviors and clustering are also observed for reproduction by fission. It is also shown that reproduction by random mate selection destroys both the phase transition and the formation of clustering. A measure for the quality of clustering is provided to compare clustering across a range of mutability values. The results demonstrate that sympatric speciation of organisms, described only by phenotypes, undergoes clustering in an evolution model under neutral conditions.

## 2.2. METHODS

The models described in this section simulated three different sizes of phenotype space, 21x21, 45x45, and 77x77. Each phenotype space was associated with a neutral fitness landscape (with fitness=2) such that every phenotype allowed no selection bias through fitness differences. In all simulations, 300 organisms were initialized according to a uniform random distribution within the phenotype space and given the same mutability, $\mu$. Simulations were performed for five runs for $\mu = 0.30$ to $1.50$ with increments of 0.02 units. The lifecycle of organisms was generational, as described above, and began with each organism producing two offspring, as is dictated by the neutral fitness landscape. Once all organisms reproduced, the parents were removed, leaving only their offspring. The offspring underwent competition death, random death, and death by absorbing boundaries. This process of birth and death was repeated until either the populations fell below three organisms (extinction; three organisms is the minimum necessary to determine a cluster, as will be discussed in more detail below) or the simulations reached the maximum time limit of 2000 generations. Simulations were performed in MATLAB (The MathWorks) on PCs using a Windows 7 operating system.

**2.2.1. Reproduction Schemes.** Three different reproduction schemes were simulated. *Assortative mating* between organisms was determined by measuring distances between organisms. Having no gender distinction among organisms, mates were chosen to be the nearest-neighbor (NN) organism to the reproducing organism, also known as the reference organism. Once mates were established, offspring were generated for each reference organism according to an area about the mating parents. This area is referred to

as the birth region. Offspring were uniformly distributed within the birth region whose opposing corners are defined by the reference organism and its mate. The region was then extended in each direction by an addition of the mutability of the reference parent (Figure 2.1). The offspring locations are ($c_{bx}$, $c_{by}$) with the reference parent location ($c_{rx}$, $c_{ry}$) and mate location ($c_{mx}$, $c_{my}$):

$$c_{bx} = \min(c_{rx}, c_{mx}) - \mu + [(\max(c_{rx}, c_{mx}) + \mu) - (\min(c_{rx}, c_{mx}) - \mu)]r_x \quad (1a)$$

$$c_{by} = \min(c_{ry}, c_{my}) - \mu + [(\max(c_{ry}, c_{my}) + \mu) - (\min(c_{ry}, c_{my}) - \mu)]r_y \quad (1b)$$

The mutability, $\mu$, is determined by the reference parent, and uniformly distributed random numbers, $r_x$ and $r_y$, were generated by MATLAB's Mersenne twister pseudorandom number generator on the interval [0,1]. An example of assortative mating simulations is shown in Figure 2.2.

The *asexual fission* algorithm is similar to that for assortative mating, except that the rectangular birth region around two parents is reduced to a square with sides of length $2\mu$, centered on the single reference parent. The coordinates of each offspring ($c_{bx}$, $c_{bx}$) were determined as

$$c_{bx} = c_{rx} - \mu + 2r\mu \quad (2a)$$

$$c_{bx} = c_{ry} - \mu + 2r\mu \quad (2b)$$

where the parent coordinates are ($c_{rx}$, $c_{ry}$), mutability is $\mu$, and r is some random number drawn from a uniform distribution on the interval [0,1].

Random mating modifies the assortative mating algorithm by having no NN calculation. Instead, mates were chosen at random from the population, regardless of their phenotype location. For every reference organism, all other organisms were equally likely to be chosen as a mate. Random mating is used as a "null condition" to identify the importance of local birth to clustering and phase transition behaviors. An example of random mating simulations is shown in Figure 2.3.

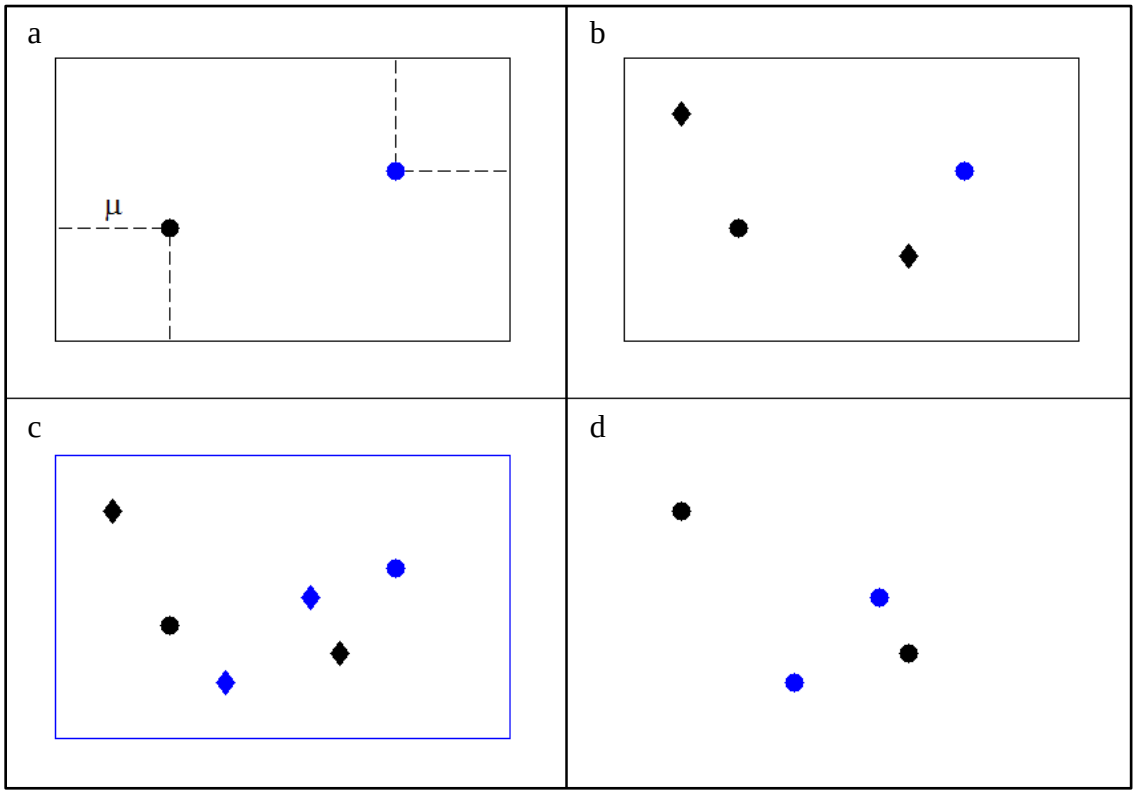


Figure 2.1. Assortative Mating. (a) A parent (black circle) and its mate (blue circle); $\mu$ shows the mutability of the reference parent and the black rectangle its birth region. (b) The offspring of the reference parent (black diamonds). (c) Assuming the mate of the blue parent is the black parent, its birth region is defined by the blue rectangle. Since all organisms in a simulation shares the same $\mu$, the birth region is identical to that of the black parent shown in panel a. (d) After parents reproduce, they are removed, leaving the offspring to undergo a battery of death processes.

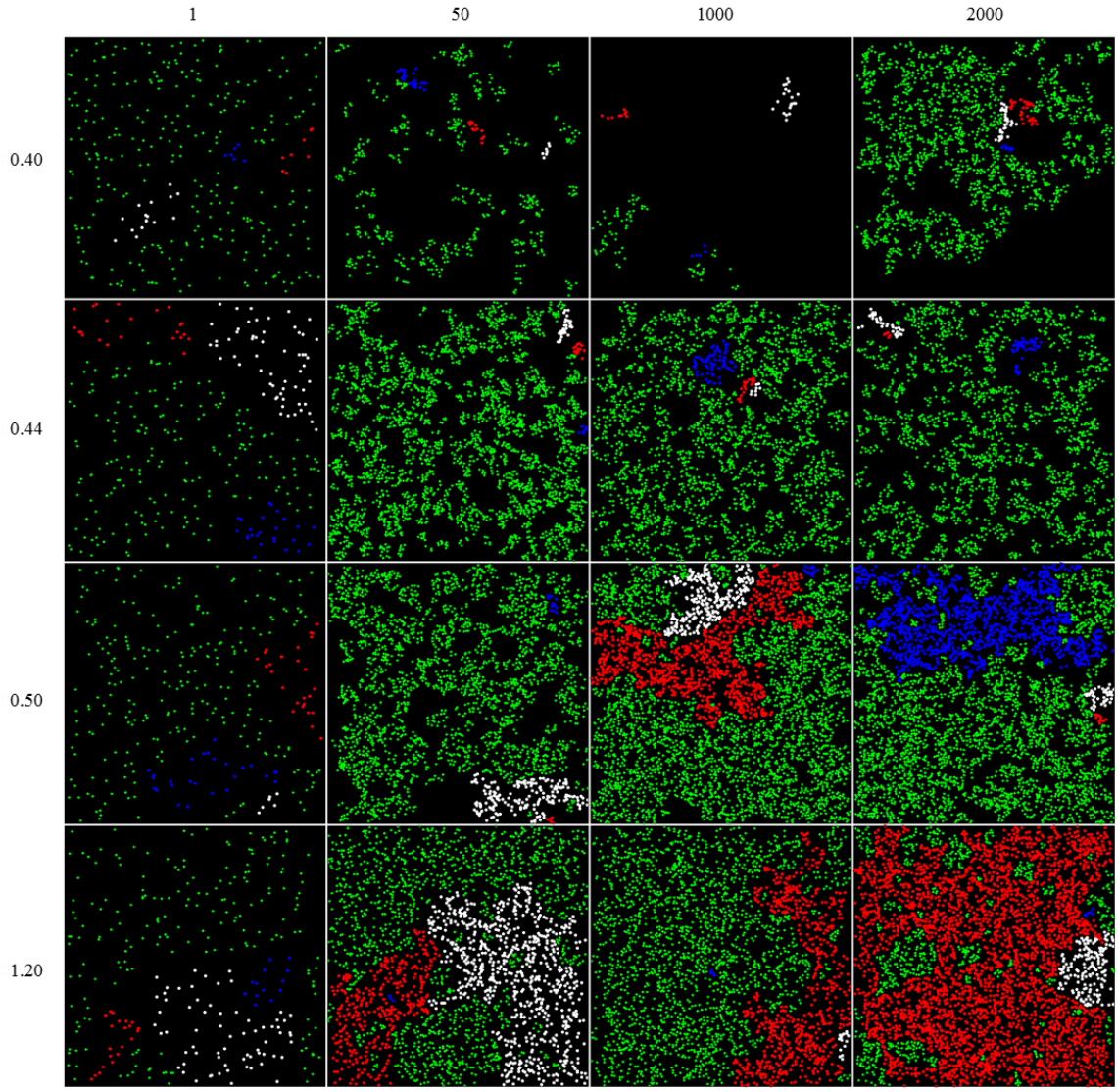


Figure 2.2. Generation Snapshots – Assortative Mating. Generational snapshots (horizontal axis) for different values of $\mu$ (vertical axis). The general population of organisms is shown in green, and example clusters are colored by red, white, or blue.

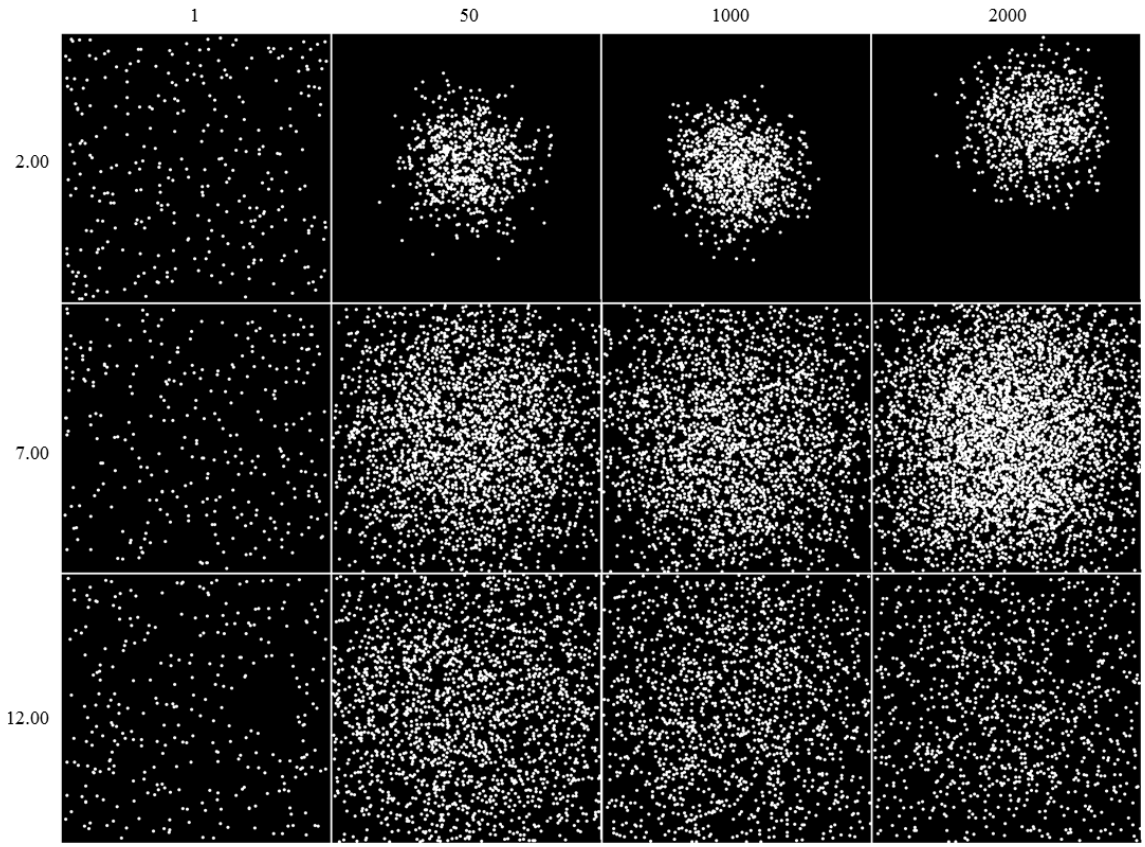


Figure 2.3. Generation Snapshots – Random Mating. Generational snapshots (horizontal axis) for different values of $\mu$ (vertical axis). The general population of organisms is colored white. Typically, only a single cluster existed at every generation for any random mating simulation.

**2.2.2. Deaths.** Three death processes were modeled. *Competitive death* occurred for any offspring within the competition radius, or limit, of $\kappa = 0.25$. Both offspring were equally likely to be chosen for death to ensure neutral conditions, giving no survival bias to any offspring. Removing only one offspring corresponds to the particle process of coalescence. The *random death* process removed a uniformly distributed random percentage of the population that was capped at 70% of the surviving offspring population after the competitive death process had been completed. Finally, any offspring found outside the phenotype space boundaries were removed, giving rise to *absorbing boundaries.* Periodic boundary conditions were not used as they are biologically unrealistic: a large phenotype cannot map onto a smaller value of the same phenotype.

**2.2.3. Clustering.** Determining the clustering of organisms is a post simulation process. For each resulting population in every generation, clusters were determined according to NN and next-nearest-neighbor (NNN) connections, in the cases of assortative mating and asexual fission. For assortative mating, clusters correspond to the biological species concept as described above, with the addition of the NNN, also defined as alternates. Finding closed sets based only on NN mates is precisely analogous to the biological species concept. However, since mating occurs deterministically (the reference organism always mates with its NN), it could be reasoned that the alternate might also be a viable mate, even though it is never chosen as one for offspring production. In the asexual fission case, clusters defined with phenotypic NN and NNN correspond to the phenetic species concept. For both assortative mating and asexual fission, a cluster "seed" was determined for each reference organism by its NN and NNN. Clusters are the mathematical union of seeds; an iterative process then found closed, disjoint sets (see Figure 2.4).

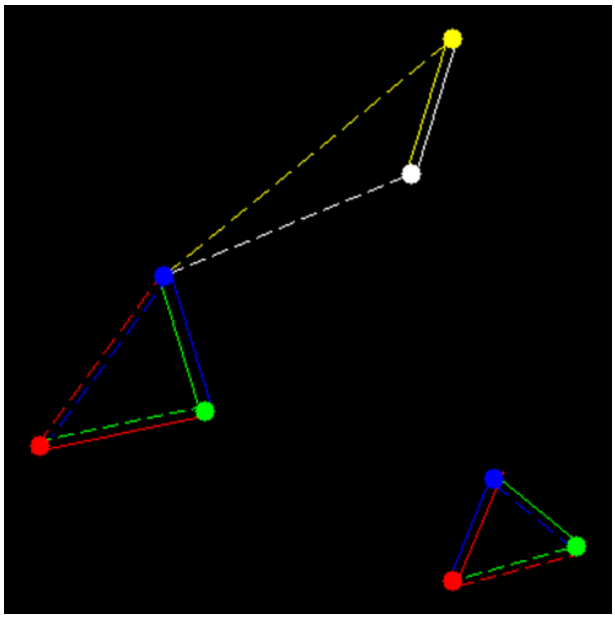

Figure 2.4. Clustering Algorithm. Each organism (colored circles) has a NN connection (solid line) and a NNN connection (hashed line). For example, white's NN is the connected yellow circle according to the solid white line, and white's NNN is the connected blue circle indicated by the hashed white line.

In the random mating case, clusters were found differently. Similar to randomly choosing mates, alternates were randomly chosen from the rest of the population. Although local clustering is lost, this choice of clustering conforms to the intended biological modeling. If a mate can be randomly chosen, then so might its alternate. This follows the mate and alternate clustering analogy with assortative mating. Figure 2.3 shows only a single cluster in each snapshot

## 2.3. RESULTS

In the example assortative and random mating simulations shown above in Figures 2.2 and 2.3, it can be seen that populations with different $\mu$ values cluster and fill the phenotype space quite differently. These observations drive the following analysis.

Measuring the time average of the population and number of clusters within each run and then taking the sample average over the time averages, the mean population, <Population>, and mean number of clusters, <Clusters>, are shown for each value of simulated $\mu$ in the assortative mating case in Figure 2.5. Small $\mu$ simulations result in small values in each since populations go extinct before 2000 generations. With increasing $\mu$, <Population> and <Clusters> rise with increasing error bars, which were determined from standard deviations of the sample averages. Further increasing $\mu$, error bars reduce in size, and <Population> levels out while <Clusters> reaches a peak and then plateaus. As the phenotype space area is increased, the slopes of rising <Population> and <Clusters> also increase. However, the $\mu$ for which populations consistently survive, the assumed critical point $\mu_c$, varies very little, with a value of $\mu_c = 0.40$ identified for each landscape. The value of $\mu$ which gives a peak in <Clusters>, $\mu_p = 0.46$, is also consistent for each landscape. At very large $\mu$, populations steadily decline, due to increased death by absorbing boundaries. All of these behaviors are also observed for asexual fission (see Figure 2.6), but with the assumed critical point varying slightly, $\mu_c = 0.36$ for 21x21 and $\mu_c = 0.34$ for 45x45 and 77x77 landscapes. The peak in <Clusters> occurs at $\mu_p = 0.38$ for each landscape in the asexual fission model.

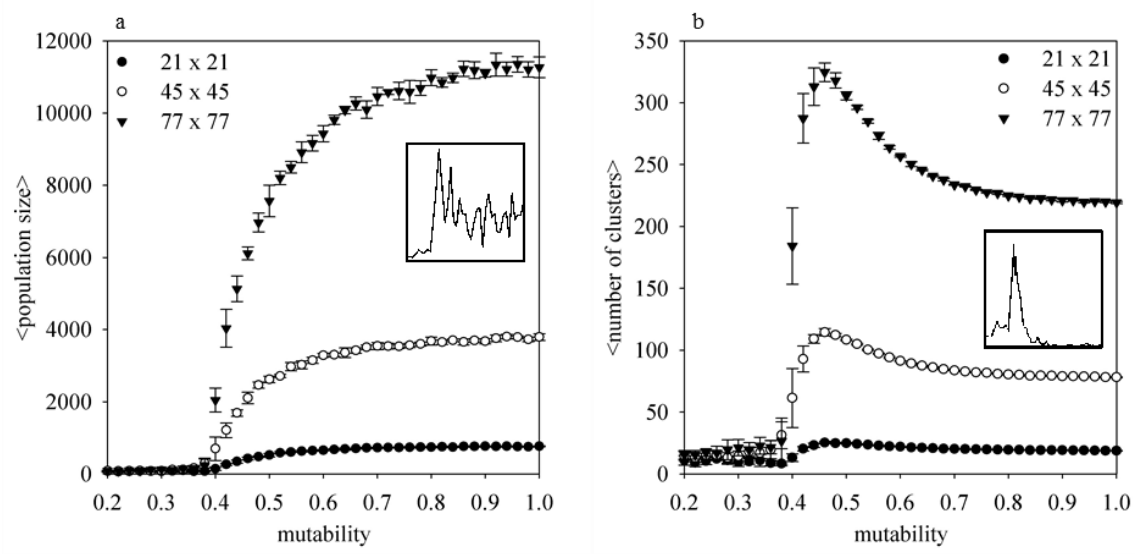


Figure 2.5. Sample Averages of Population and Clusters – Assortative Mating. (a) Sample population averages for each $\mu$. (b) Sample number of clusters averages for each $\mu$. Insets show standard deviation of the 77x77 landscape data for each panel as a function of mutability (adapted from Scott et al. 2013).

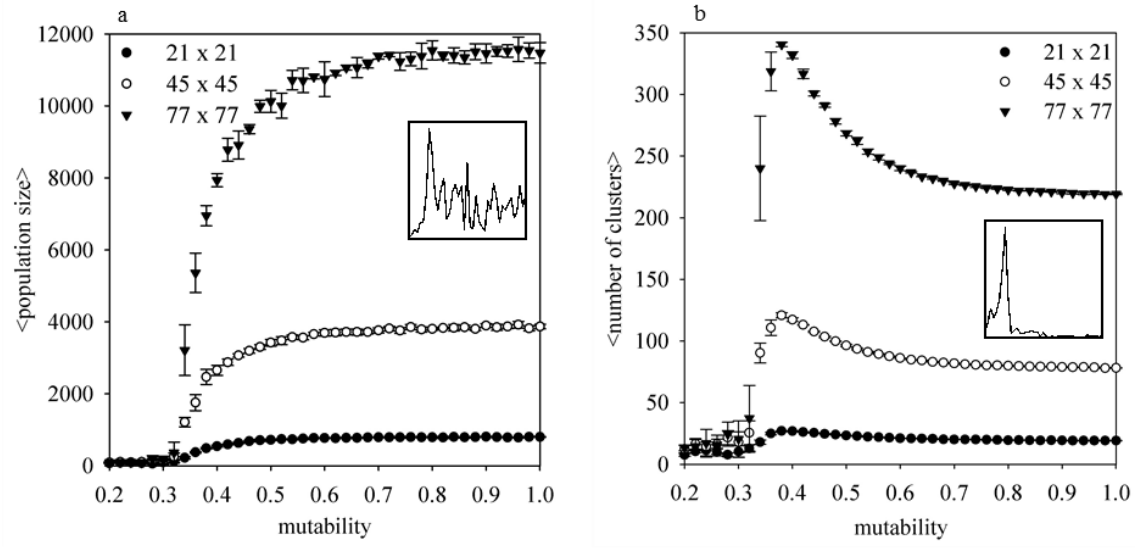


Figure 2.6. Sample Averages of Population and Clusters – Asexual Fission. (a) Sample population averages for each $\mu$. (b) Sample number of clusters averages for each $\mu$. Insets show standard deviation of the 77x77 landscape data for each panel as a function of mutability (adapted from Scott et al. 2013).

The transition of <Population> from extinction to survival with increasing $\mu$ was also observed in the system lifetimes. Histograms of system lifetimes for the 45x45 landscape are shown in Figure 2.7. Elongation in the histogram tail grew more power-law like as $\mu$ approached $\mu_c$. For $\mu = 0.42$ and above, system lifetimes were observed to last indefinitely with no sign of decay, and were stopped manually after a few million generations (not shown).

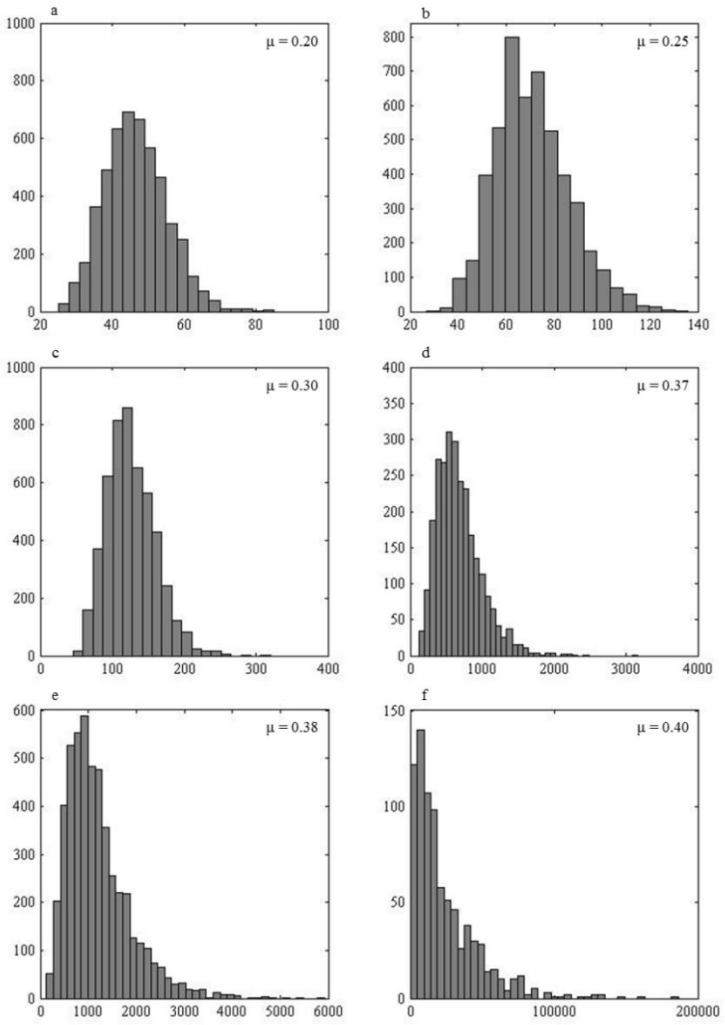


Figure 2.7. Histograms of System Lifetimes – Assortative Mating. Many more simulations were run for $\mu = 0.20$ (5000 runs), $\mu = 0.25$ (5000 runs), $\mu = 0.30$ (5000 runs), $\mu = 0.37$ (2662 runs), $\mu = 0.38$ (5017 runs), and $\mu = 0.40$ (828 runs) (adapted from Scott et al. 2013).

Temporal behavior of the population densities about $\mu_c$ were observed for the asexual fission model and are shown in Figure 2.8. For the values of $\mu$ shown, the average population density in each generation was calculated from 100 runs. Below the assumed $\mu_c$, these time series show the population decaying to extinction. As $\mu$ approaches $\mu_c$, the rate of decay becomes progressively slower until, at $\mu_c$, the system reached a surviving state.

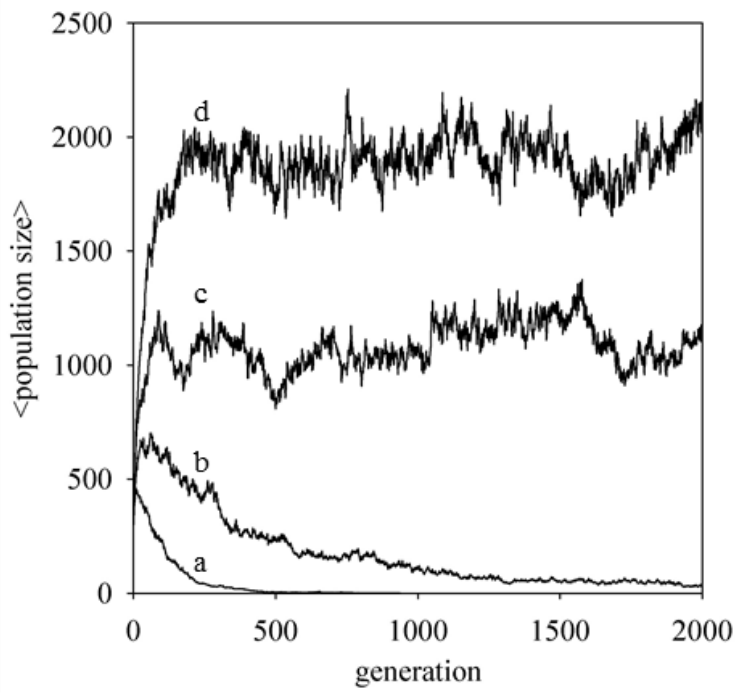


Figure 2.8. Time-Series Average Populations Near Criticality – Asexual Fission. (a) $\mu = 0.30$, (b) $\mu = 0.32$, (c) $\mu = 0.34$, (d) $\mu = 0.36$ (adapted from Scott et al. 2013).

The rise in <Population> along with error bars was not observed for the random mating model. Instead, <Population> gradually increased over a much larger range of $\mu$, as shown in Figure 2.9. Furthermore, no distinct clustering was observed since <Clusters> was approximately one for all $\mu$.

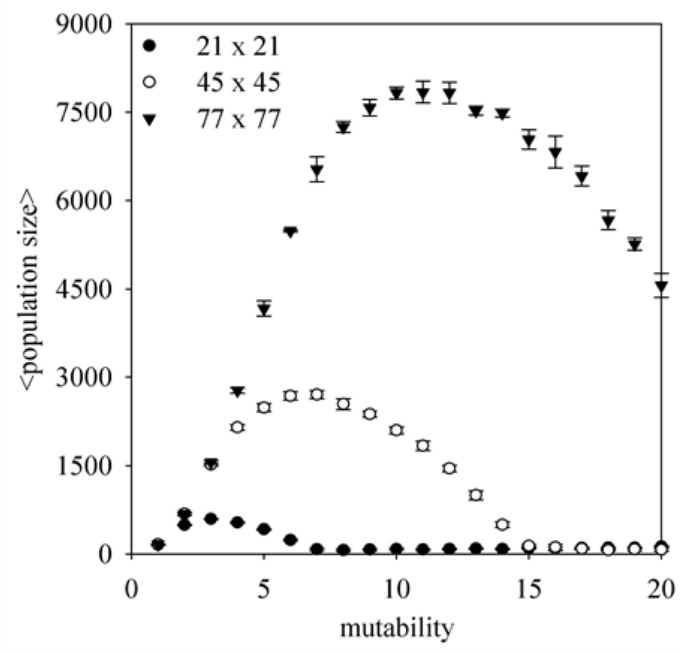


Figure 2.9. Sample Averages of Populations – Random Mating.
(adapted from Scott et al. 2013).

The quality of clustering was measured for assortative mating and asexual fission using the Clark and Evans (1954) nearest-neighbor index, $R$ (see Eq. 3 for a square area)

$$R = \frac{2\sqrt{N}\sum_{i=1}^{N}\left(\left(c_{r_i x}-c_{m_i x}\right)^2+\left(c_{r_i y}-c_{m_i y}\right)^2\right)^{1/2}}{NL}. \quad (3)$$

Here, N is the population of a given generation, $c_{r_i}$and $c_{m_i}$ are the locations of a reference organism, $i$, and its NN, respectively, and $L$ is the linear landscape size (21, 45, or 77). This index measures the average NN distances for each population and compares it to a purely random distribution of NN measures given a specific area. An index of $R = 1$ corresponds to populations whose NN distances are purely random. For $R < 1$, populations are more aggregated, and, for $R > 1$, populations are more uniformly distributed. Clark and Evans introduced a significance measure to determine if populations may be considered aggregated or uniformly distributed when $R$ is near 1. The

sample average, $\langle R \rangle$, is shown in Figure 2.10 for both assortative mating and asexual fission.

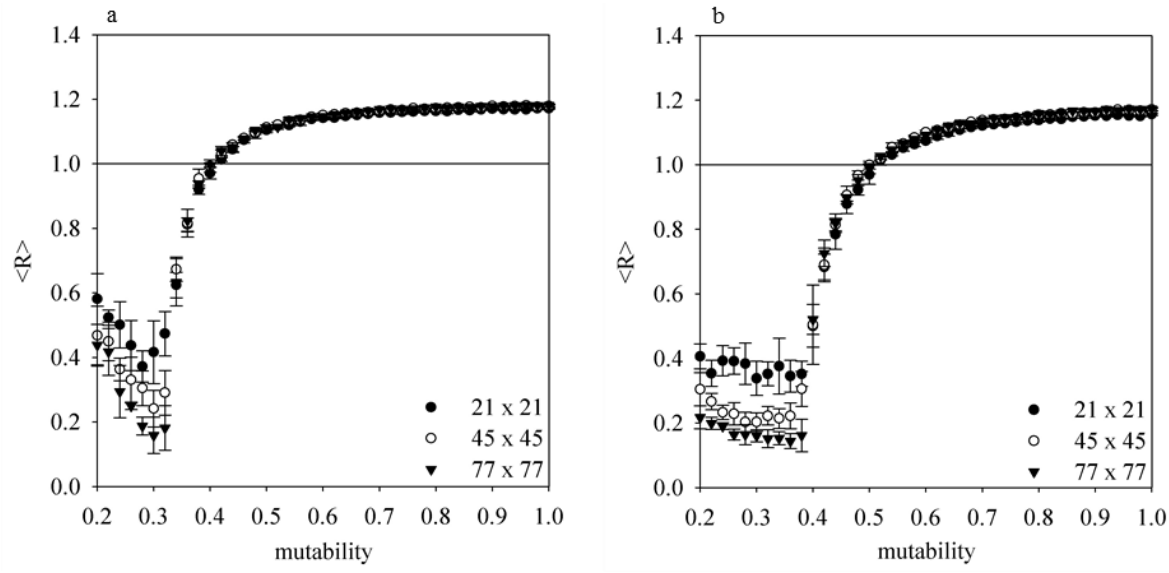


Figure 2.10. Sample Nearest-Neighbor Index – Populations. (a) assortative mating and (b) asexual fission (adapted from Scott et al. 2013).

Assortative mating populations were found to have purely random distributions (within 1% significance) for $0.48 \leq \mu \leq 0.54$ on the 21x21 and 45x45 landscapes, and for 0.48 to 0.50 for the 77x77 landscape. Below each range, populations were found to be significantly aggregated, and above each range, populations were more uniformly distributed. Similarly, asexual fission populations were found to be distributed according to a purely random distribution for $0.38 \leq \mu \leq 0.40$ on the 21x21 and 45x45 landscapes. Populations were significantly aggregated for $\mu \leq 0.38$ and uniformly distributed for $\mu \geq 0.40$ on the 77x77 landscape.

The distributions of cluster-size, known as species abundance in biology, are shown in Figure 2.11 for values of $\mu$ near $\mu_c$ for assortative mating. A cluster size is measured as the number of organisms in a cluster. The double logarithmic plots in Figure

2.11 show a change in concavity as $\mu$ increases from below $\mu_c$ to above $\mu_c$. Nearest $\mu_c$, power-law behavior is observed in the tail of the distribution, as emphasized by the best-fit line from a minimization of a chi-square linear-fit to the data.

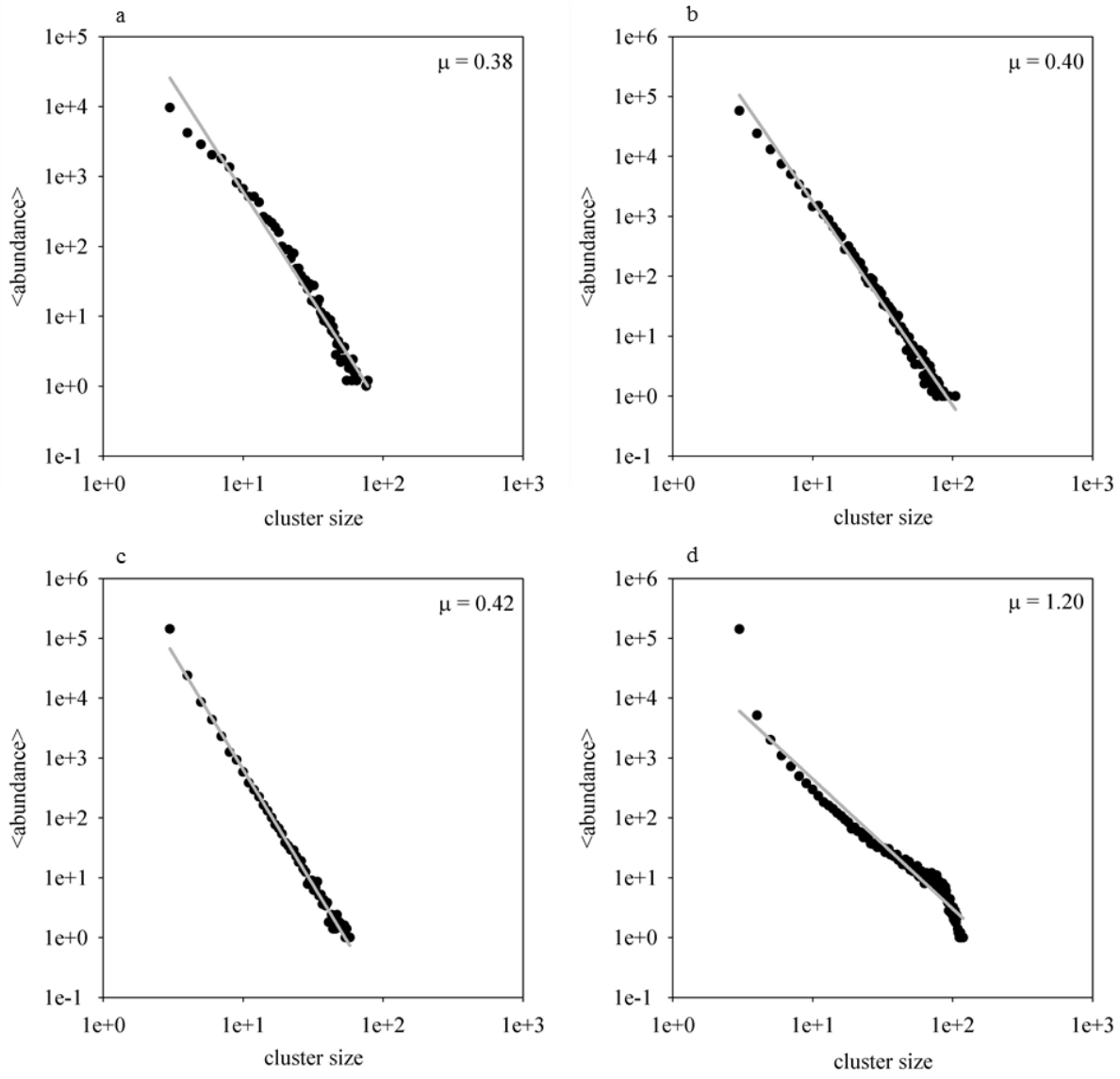


Figure 2.11. Cluster-Size Distributions Near Criticality. Note the concavity change in the distribution from $\mu = 0.40$ (b) to $\mu = 0.42$ (c) (adapted from Scott et al. 2013).

## 2.4. DISCUSSION

It is demonstrated above that clustering of organisms can occur without physical or genotype space and on a neutral landscape. Clustering in this model is representative of sympatric speciation and occurs for both assortative mating and asexual fission. In contrast, the random mating model almost always produces one large cluster. These findings are in agreement with other studies of neutral fission models that characterize clustering (Zhang et al. 1990; Derrida & Peliti 1991; Meyer, Havlin, & Bunde 1996; Young et al. 2001; Houchmandzadeh 2002; Houchmandzadeh & Vallade 2003; Lawson & Jensen 2002), but are in contrast with the conclusions reached by de Aguiar et al. (2009) which concluded that only assortative mating of organisms described in a genotype space and physical space were the minimum requirement for clustering.

Derrida and Peliti (1991) investigated lineages of individuals defined by genomes of spin-like alleles within fixed population sizes. The genomes were subject to a constant mutation rate, and individuals underwent asexual fission on a neutral fitness landscape. Serva and Peliti (1991) extended the model to include random mating, and Higgs and Derrida (1992) modified it for mating between individuals with some amount of genomic overlap. Speciation occurred for mating with the requirement of genome overlap. In contrast to these models, the models described in this dissertation are not tied to specific population sizes, and this allows far richer dynamics of emergent speciation and "biodiversity".

The results presented in this work are related to BCRWs in a discrete space as studied by Athreya and Swart (2005). Particles in their model perform independent random walks, undergo binary splitting, and experience coalescence and random death. The Markov process in their work is similar to the one presented here with the primary difference being that the agents in this work exist in a continuous space with absorbing boundaries. In the thermodynamic limit of an infinite landscape, the critical value of $\mu$ may be smaller than reported (Marić, in preparation). As a result, increasing the landscape size to much larger areas may produce a different critical point. Even with the landscape sizes used here, since the critical range in both assortative mating (~0.40) and asexual fission (~0.34) is relatively close to the competition limit (0.25), this effect is not detectable from the simulations. The smaller landscape size (or much larger mutabilities)

induces a greater amount of killing from absorbing boundaries as compared with larger landscapes (or smaller mutabilities). A measure of the proportion of deaths due to competition, random, and absorbing boundaries shows that the vast majority of deaths come from competition and random death over a large range of $\mu$, well above both the critical point and the peak in <Clusters>. The lack of system size effects may be due to the correlation length being much less than the dimensions of the phenotype space; this was shown to be the case for critical parameters in both Privman (1990) and Toral and Tessone (2007). At very large values of $\mu$, much greater than $\mu_p$, system size effects become apparent due to a greater portion of deaths from absorbing boundaries.

Spatial asymmetry between birth and death processes was shown to be the primary requirement for clustering in the "Brownian bugs" model of Young et al. (2001). Fuentes, Kuperman, and Kenkre (2003) also found that some nonlocal competition is a prerequisite for pattern formation. Through various forms of an "influence function", nonlocal competition was controlled for some key parameters such as the linear system size and width of the influence function. Figure 4 in their work suggests phase transition behavior in the ratio of the influence function width to the linear system size (Fuentes et al. 2003). The importance of spatial asymmetry was confirmed to be necessary as demonstrated by the loss of clustering in the random mating variant of the neutral model described in this chapter. Selecting mates arbitrarily in phenotype space causes a loss in local birth, and causes birth regions to cover larger areas of the phenotype space, matching the scale of the random death process.

The relative values of $\mu$, the competition radius, the linear system size, and the amount of random death likely define the critical behavior of the neutral model. If the competition radius goes to zero, the critical mutability will likely go to zero as well, becoming highly dependent on the amount of random death and potentially losing the phase transition entirely. A similar case occurred in Fuentes et al. (2003) where all population structure was lost in the extreme local limit.

Steep increases in the order parameters, such as <Population> and <Clusters> are not atypical for this system, since they were observed on rugged fitness landscapes for similar intermediate values of $\mu$ (Dees & Bahar 2010). The neutral model shows that the NN index $R$ also experiences a steep increase on the interval between $\mu_c$ and $\mu_p$. The

control case of random mating on the neutral landscape is in stark contrast, showing a much more gradual increase in <Population> over a much greater range of $\mu$. A sharp increase in the population size, separating the extinction-survival transition across $\mu_c$, along with large error bars near the critical point, and divergence of the system lifetime, are indicative of a continuous non-equilibrium phase transition. Finite-size effects were not detected in the critical value of $\mu$, as discussed above, but increasing steepness in the transition was observed with increasing landscape sizes. This effect is typical for a directed percolation phase transition (Solé 2011).

Although phase transition-like behavior has been observed in other evolutionary models (Luz-Burgoa, Moss de Oliveira, Schwämmle, Sá Martins 2006), to the author's knowledge, no other study has shown a transition due to varying a mutation-related parameter. Continuous phase transitions are accompanied by scale-free behavior at and near the critical point, and this behavior can be detected in the log-log plots of cluster-size distributions on the critical mutability range. Power-law behavior in the tails of the distributions can be detected from the linearity of log-log plots, indicating no characteristic scale of the cluster sizes. Observing a concavity change from "down" to "up" in Figure 2.11 with steadily increasing $\mu$ near $\mu_c$ indicates a power-law in the distributions. This suggests that the phase transition is continuous, and since the critical point separates subcritical populations going to the absorbing state of extinction from supercritical populations going to a fluctuating active state, the transition is also clearly a non-equilibrium one.

By characterizing the filling of the space, the NN index $R$ measures the quality of clustering in populations for different values of $\mu$. For $R < 1$, the phenotype space is less well filled, with populations being more often clumped together. As $R$ increases and becomes greater than one, populations become more uniformly distributed and clusters are less well-defined. The existing clusters for $R > 1$ correspond biologically to less well-defined species. Given sufficient knowledge of genetic lineages, species could still be defined, and the clustering algorithm can still be used to determine clusters. The peak observed for <Clusters> corresponds to the same range of $\mu$ for which populations are associated with $R \sim 1$. This effectively means that $R$ transitions from clumpy populations for $\mu < \mu_p$ to uniformly distributed populations for $\mu > \mu_p$. The decline of <Clusters>

is then tied to the more space-filling populations. Populations with $\mu_c < \mu < \mu_p$ were more aggregated and exhibited relatively similar $R$ values across the landscape sizes, but, for $\mu < \mu_c$, $R$ splits among the landscape sizes. This behavior is likely due to smaller population sizes, since the simulations tended to extinction for $\mu < \mu_c$. This effect can be explained by the increase in linear system size which is incorporated into the $R$ measure via an inverse relationship for square landscapes (Eq. 3).

The neutral models described here could certainly be extended to include higher phenotype dimensionality. Indeed, in some paleobiological works such as Foote (1990) and Abe and Libermann (2012), nearest-neighbor methods are used to discern species of trilobites according to more than ten morphological traits. From renormalization group methods, it is known that the dimensionality of a space determines critical behavior, where above a certain upper critical dimension (depending on the universality class of the phase transition), the phase transition behavior conforms to mean-field theoretical predictions (Hinrichsen 2000). The dimensionality is suggested to affect phenotype clustering in Lawson and Jensen (2008), so the relationship between biodiversity and higher dimensional models is a recommended course of future study.

# 3. CHARACTERISATION OF A PHASE TRANSITION

## 3.1. INTRODUCTION

A given phase transition is characterized by universal dynamics which describe the system's behavior; the exponents which describe the system's scaling in the neighborhood of the phase transition define its universality class (Henkel, Hinrichsen, Lübeck 2009). At the critical point, the correlation length of the system diverges, and the critical behavior of the system can then be analyzed through renormalization group methods (Fisher 1998; Täuber, Howard, & Vollmayr-Lee 2005; Lesne 1998). In non-equilibrium phase transitions, a system undergoes an irreversible transition from an active fluctuating state into an absorbing state from which it cannot escape (Henkel et al. 2009, Hinrichsen 2000, Ódor 2004). Universality classes can be described by reaction-diffusion models of unary and binary particle processes describing birth, death, and particle interactions such as coalescence and annihilation (Ódor 2004; Hinrichsen 2003; Täuber et al. 2005). There are many different universality classes of non-equilibrium phase transitions such as pair contact processes with and without diffusion (PCP and PCPD), parity-conserving (PC), conserved threshold transfer process (CTTP), and perhaps the most important class, directed percolation (DP) (Henkel et al. 2009; Ódor 2004, Hinrichsen 2000). For DP, the processes involved are birth (for example, $A \rightarrow 2A$), death ($A \rightarrow 0$), and coalescence (for example, $2A \rightarrow A$) (Täuber et al. 2005). The spatial and temporal dimensions in which these processes take place (discrete or continuous) do not significantly affect the critical dynamics, since DP models have been shown to occur with discrete time steps on lattices with varied geometries and off-lattice in a continuum or even on a lattice with continuous time (Oborny, Meszéna, Szabó 2005; Grimmett 2008). The dimensionality of the spatial component determines different universal behaviors for a universality class up to an upper critical dimension, above which the universal behaviors coincide with the mean field description (Henkel et al. 2009, Hinrichsen 2000, Ódor 2004). The upper critical dimension of DP is four (Hinrichsen 2000; Ódor 2004; Henkel et al. 2009).

In the context of biological evolution, DP models have included cellular automata such as the Domany-Kinzel (DK) model and the contact process (CP); however, these

models are usually modified by additional biological detail so that the universal behavior is modified (Laventrovich, Korolev, Nelson 2013; Lipowski, Ferreira, Wendykier 2012; Kuhr, Leisner, Frey 2011; Oborny et al. 2005; Lipowski & Lopata 1999). Generally, mathematical models of biological processes are defined in physical space and/or genetic space and are commonly restricted to constant population sizes (Derrida & Peliti 1991; Brunet & Derrida 2009; Tran, Hofrichter, Jost 2013; de Aguiar et al. 2009). More uncommon are models occurring in phenotype space (Lawson & Jensen 2008). In this work, the asexual fission model on a neutral landscape is investigated; this model resembles a spatial-branching process with coalescence in phenotype space, where the particle (organism) dynamics are birth (A→2A), random death (A→0), and coalescence (2A→A). In the previous section, it was shown that the model transitioned from an active surviving state to an absorbing state of extinction as the maximum phenotype mutation size was varied (Scott et al. 2013). The observed system behavior was suggestive of the DP universality class. Here, it is demonstrated, by measuring the critical exponents and universal functions, that the model does indeed belong to the DP universality class. Biological implications regarding experimental comparisons of this model to paleobiological morphology studies and disorder problems that abiogenesis theories are discussed in Section 3.6.

### 3.2. CRITICAL EXPONENTS AND UNIVERSAL FUNCTIONS

Universality classes of non-equilibrium phase transitions are in general defined by four fundamental exponents. Barring any special symmetries (two of the exponents are identical in DP), the four exponents are independent, and each characterizes the scaling behavior, in the neighborhood of the phase transition, of an order parameter which depends on the distance between the control parameter and the critical point. The control parameter of interest here is the mutability, $\mu$. Let $\Delta$ be the off-critical measure, the absolute difference in mutability from the critical point

$$\Delta = |\mu - \mu_c|. \tag{4}$$

Order parameters such as the population density (steady-state) are known to depend upon the off-critical measure ($\Delta$), time ($t$), and in the case of finite systems, the linear landscape size ($L$). Near the critical point, these dependencies are given by a homogenous function of the population density (can factor out a scaling constant raised to some power), known as a finite-size universal scaling equation (Laventrovich, Korolev, & Nelson 2013; Henkel et al. 2009)

$$\rho(\Delta, t, L) = b^{-\beta}\tilde{\rho}(tb^{-\nu_\parallel}, \Delta b, L^z b^{-\nu_\parallel}). \tag{5}$$

The parameter b is a dimensionless scale factor to be chosen, and the function $\tilde{\rho}$ is the universal scaling function. Choosing $b = t^{1/\nu_\parallel}$ gives

$$\rho(\Delta, t, L) = t^{-\beta/\nu_\parallel}\tilde{\rho}\left(1, \Delta t^{1/\nu_\parallel}, \frac{L^z}{t}\right). \tag{6}$$

The first dependency of $\tilde{\rho}$ indicates that the density decays as a power law. The second dependency gives the off-critical scaling behavior, and the third gives the finite-size scaling behavior. From Eq. 6, it is known that the basic set of critical exponents (other exponents can be derived from them) can also be found from scaling relationships with the control parameter $\Delta$ (Hinrichsen 2000).

$$\rho \propto \Delta^{\beta} \tag{7}$$

$$\xi_\parallel \propto \Delta^{-\nu_\parallel} \tag{8}$$

$$\xi_\perp \propto \Delta^{-\nu_\perp} \tag{9}$$

$$P \propto \Delta^{\beta'} \tag{10}$$

where $\rho$ is the population density (steady-state), $\xi_\parallel$ is the correlation time, $\xi_\perp$ is the correlation length, and $P$ is the survival probability (Hinrichsen 2000).

Due to the time-reversal symmetry (growth from single seed simulations yield survival probabilities which scale identically to density decays from critical-quench simulations) of directed percolation, the population density scales exactly the same as

survival probability, so $\beta = \beta'$ (Hinrichsen 2000). The above scaling measures assume an infinite system size, so due to finite system size, simply using the scaling relations above is expected to produce inaccurate exponents (Hinrichsen 2000). Investigation of the exponents from the data-collapse technique of critical-quench simulations was used.

Data-collapse is a technique in which time-series generated from different parameter values are fitted to some function through rescaling (from modifications of Eqs. 10 and 12). For example, Eq. 11 focuses on one of the dependencies in Eq. 6 which shows how the population density depends upon time ($t$), and the off-critical measure ($\Delta$). By rescaling the time and off-critical measure with the same exponents, $\alpha$ and $\nu_{\parallel}$, for each time-series, the original data "collapses" onto the same curve. Graphically, there are two parts of the universal function described by Eq. 11, one for subcritical and the other for supercritical data. Examples of the desired universal scaling functions for DP systems can be found for CP simulations in Hinrichsen (2000), the radial DK model in Laventrovich, Korolev, and Nelson (2013), a tumor growth model with highly unusual DP dynamics in Lipowski, Ferreira, and Wendykier (2012), and in an experimental system of liquid crystals in Takeuchi, Kuroda, Chaté, and Sano (2009).

Without altering the universal properties, the arguments to the universal scaling function may be rescaled. Changing the off-critical argument of the universal scaling by $\left(\Delta t^{1/\nu_{\parallel}}\right)^{\nu_{\parallel}} = t\Delta^{\nu_{\parallel}}$ gives

$$\rho \sim t^{-\alpha}\tilde{\rho}(t\Delta^{\nu_{\parallel}}) \tag{11}$$

with

$$\alpha = \frac{\beta}{\nu_{\parallel}}. \tag{12}$$

This allows for the determination of $\nu_{\parallel}$ and, in turn, of $\beta$ from the relationship with $\alpha$ as given in Eq. 12. The $\nu_{\parallel}$ above and below the critical point should be equal, and in turn, so should $\beta$.

The finite-size argument of the universal scaling function may also be changed by the rescaling $(L^z/t)^{-d/z} = t^{d/z}/L^d$,

$$\rho \sim t^{-\alpha}\tilde{\rho}(t^{d/z}/L^d) \tag{13}$$

where $d$ is the spatial dimension and

$$z = \frac{\upsilon_{\parallel}}{\upsilon_{\perp}} \tag{14}$$

which relates $\upsilon_{\parallel}$ and $\upsilon_{\perp}$.

### 3.3. MODEL

Simulations were performed identical to the asexual fission case described in Section 2. However, the three landscape sizes and initial populations are different. The landscape sizes were 29x29, 37x37, and 45x45, and had initial population sizes of 12474, 20290, and 30000 organisms, respectively. The initial population sizes reflect the same initial population density of organisms for each landscape, equivalent to filling approximately 79.9% of sites on a hexagonal lattice within the landscape boundaries (see corresponding Methods and Appendix A). The initial populations were seeded as before across the space with a uniform random distribution.

### 3.4. METHODS

**3.4.1. Critical-Quench Simulations.** Critical-quench experiments on lattices are experiments which start with every lattice site occupied in the active state and are used to determine the critical point of a system by noting changes in the population density decay behavior (population density is a linear rescaling of actual population size, so the logarithmic decay is equivalent for both). Due to the continuous space in this model, the maximum population capacity was taken to correspond to the greatest density of organisms fitted into the landscape area where the minimum spacing between organisms is the competition radius $\kappa$. This means that organisms can be considered as hard circles of radius κ centered about the coordinate of each organism. The greatest density of circles

(organisms) in a square landscape is the hexagonal lattice (Steinhaus 1999). From this the maximum population was determined for a given phenotype space area (see Appendix A). Due to computational constraints for the larger landscapes starting populations were set at 79.9% of the maximum capacity, giving initial populations of 12474, 20290, and 30000 for landscape sizes 29x29, 37x37, and 45x45, respectively. Although not a "full" initial condition, the starting populations are much greater than the average populations observed to occur for the same mutabilities in Scott et al. (2013). The maximum time allowed for each simulation was $T = 10^5$ generations. There were 100 simulations performed at each value of $\mu$ over the range 0.29 to 0.37, at 0.01 increments. The simulations were run in MATLAB with the Parallel Processing Toolbox on PCs running Windows 7 as well as the UMSL Bortas 24-node Oracle server with MATLAB and the Distributed Computing Server Toolbox.

**3.4.2. Critical Point and Decay Rate.** The decay from large starting populations was observed for a range of mutabilities about the hypothesized critical point from ($\mu = 0.34$) Scott et al. (2013). Simulations with $\mu = 0.34$ appeared to be well into the surviving regime, whereas simulations with $\mu = 0.32$ were well into the extinction regime. The simulated critical point, $\mu_{cs}$, was determined as the mutability which minimized the $\chi^2$ linear fit of double logarithmic plots of the population density vs. time for the range of mutabilities described above. Only generations 10-1000 were used because of the manifestation of noticeably different decay dynamics around 2000 generations. Finite-size effects incur a penalty on simulations in the supercritical range, when system lifetimes should have diverged to infinity at the critical point in an infinite landscape. A non-zero probability exists in this case that finite-size simulations can go to the absorbing state in the supercritical regime (Henkel et al. 2009). This caused the density decays to fall into the absorbing state for very long lifetimes, which seemed to begin around 2000 generations in this case.

Noting that the exponent $\alpha$ changed with $\mu$ and that the rate of change decreased near $\mu_{cs}$, cubic interpolation of $\alpha$ was performed with 0.0025 increments of $\mu$ to determine an estimated critical mutability and decay rate, $\mu_{ce}$ and $\alpha_e$. The estimated critical point, $\mu_{ce}$, was determined to occur at the $\mu$ for which the change in interpolated $\alpha_e$ was steepest (second derivative peak). An estimated $\mu_{ce}$ and $\alpha_e$ were determined for

each of the three landscapes. In the following, *subcritical* refers to $\mu < \mu_c$ and *supercritical* refers to $\mu > \mu_c$. The values of $\alpha_e$ and $\mu_{ce}$ were used in the data-collapse approach, described below, in order to determine $\upsilon_{\parallel}$.

**3.4.3. Data Collapse.** Using the estimated critical values $\mu_{ce}$ and $\alpha_e$ on the time range of 1-1000 generations, with a rearrangement of Eq. 11 the subcritical and supercritical data-collapses according to the universal function

$$\rho t^{\alpha} \sim f(t\Delta^{\upsilon_{\parallel}}). \tag{15}$$

The minimization of the goodness of fit (see below) of overlapping sections of the rescaled time series determined separate subcritical and supercritical values of $\upsilon_{\parallel}$ for each landscape. By the exponent relation in Eq. 12, the exponent $\beta$ was then obtained for both subcritical and supercritical regimes.

To obtain $\upsilon_{\perp}$, the finite-size universal scaling function was plotted from a rearrangement of Eq. 13 for generations 1-5000. The universal scaling function at each landscape size was based on the average population density $\rho$ for the supercritical $\mu = 0.33$.

$$\rho t^{\alpha} \sim f(t^{d/z}/L^{d}) \tag{16}$$

By using the same method of goodness of fit that was performed for $\upsilon_{\parallel}$, the exponent $z$ was determined, which led to a determination of $\upsilon_{\perp}$ from Eq. 9. The value of $\alpha$ used for the determination of $z$ was averaged from the $\alpha_e$ from each landscape size. Similarly, $\upsilon_{\parallel}$ was averaged over the six values corresponding to the sub- and supercritical values of each landscape.

**3.4.4. Goodness of Fit.** To determine the exponents which produce the best data-collapse, the goodness of fit measure outlined by Bhattacharjee and Seno (2001) was used. However, instead of searching a two-dimensional exponent space, only values of $\upsilon_{\parallel}$ were scanned while $\alpha$ was held constant. The goodness of fit measures collectively, for each overlapping pair of rescaled data, how well the data sets collapse onto each other, according to the scaling functions in Eqs. 15 and 16. The two measures for data-collapse, $P_1$ and $P_2$, are

$$P_1 = \left[\frac{1}{N_{overlap}} \sum_{p=1}^{n} \sum_{j \neq p}^{n} \sum_{i=1}^{N_{overlap}} |\rho_j \, t_{i,j}^{\alpha} - \mathcal{E}_p(\Delta_j \, t_{i,j}^{\nu_{\parallel}})|^q\right]^{1/q}, \tag{17}$$

$$P_2 = \left[\frac{1}{N_{overlap}} \sum_{p=1}^{n} \sum_{j \neq p}^{n} \sum_{i=1}^{N_{overlap}} |\rho_j \, t_{i,j}^{\alpha} - \mathcal{E}_p(t_{i,j} \, L_j^{-z})|^q\right]^{1/q}. \tag{18}$$

The measures only consider overlapping regions of data points from the *n* curves, indexed by $j$ and $p$, so data is interpolated by the function $\mathcal{E}_p$. For example, $P_1$ is calculated separately for subcritical and supercritical data sets, so $n$ = 4 for subcritical, and $n$ = 5 for supercritical. There are three different landscape sizes, and so $n$ = 3 for calculating $P_2$. The $p^{\text{th}}$ interpolation curve is over the shared scaled time span of the $j^{\text{th}}$ curve. There are $N_{overlap}$ interpolated points that are indexed by $i$. The value $q$ is defined as a positive integer which may be chosen arbitrarily, and is taken here as $q$ = 1. A wide range of values for $\upsilon_{\parallel}$ and $\upsilon_{\perp}$ ranging from 0.25 to 3.00 in 0.001 increments were used to scan for the minimum in the goodness of fit measures. This was done to ensure that other potential universality classes and a mixture of different exponents could be ruled out (see Lübeck 2004 for a list of other universality class exponents; Lipowski et al. 2012). The minimum of $P_1$ was found separately for subcritical and supercritical, so $n$ = 4 for the four subcritical mutabilities and $n$ = 5 for the five supercritical mutabilities. For $P_2$ the value of $n$ was 3, since that is the number of supercritical curves at $\mu = 0.33$ from each of the landscapes simulated.

The estimated error bars for $\upsilon_{\parallel}$ and $\upsilon_{\perp}$ were calculated from the width of the goodness of fit measure about the minimum of the $P_1$ and $P_2$curves (Eqs. 17 and 18), giving

$$\Delta\nu_{\parallel} = \eta\nu_{\parallel 0}\left[2ln\frac{P_1(\nu_{\parallel 0}\pm\eta\nu_{\parallel 0})}{P_1(\nu_{\parallel 0})}\right]^{-1/2}, \quad (19)$$

$$\Delta\nu_{\perp} = \eta\nu_{\perp 0}\left[2ln\frac{P_2(\nu_{\perp 0}\pm\eta\nu_{\perp 0})}{P_2(\nu_{\perp 0})}\right]^{-1/2}. \quad (20)$$

Here, $\eta$ represents the desired size of the error bars to be measured and was taken to be 1% for the results below.

### 3.5. RESULTS

The amount of concavity change in the population density time series about the critical point was visually seen to vary slightly between the landscapes 29x29 and 45x45. However, the most linear population decay on the generation range and the lower and upper bound range for the critical point described above (minimization of $\chi^2$ linear fit) consistently occurred for $\mu = 0.33 = \mu_{cs}$ for each landscape (Figures 3.1, 3.2, and 3.3). The corresponding $\alpha_s$'s were 0.4010±0.1377, 0.3389±0.1381, and 0.3403±0.1404 for landscapes 29x29, 37x37, and 45x45, respectively (error bars were calculated from the standard deviation of the population density data from the $\chi^2$ linear fit test found in Garcia 2000). By interpolating $\alpha$ vs. $\mu$, the steepest change in $\alpha$ marked the estimated critical point, and was the same for each landscape at $\mu_{ce} = 0.3275$. The corresponding values of $\alpha_e$ are summarized in Table 3.1 and the best-fit lines from $\alpha_e$ are shown in Figures 3.1, 3.2, and 3.3. From the minimization of the Bhattacharjee and Seno (2001) goodness of fit measure, $P_1$ from Eq. 17, $\upsilon_{\parallel}$ was determined for each landscape size separately below ($\upsilon_{\parallel}{}^{-}$) and above ($\upsilon_{\parallel}{}^{+}$) $\mu_{ce}$, using $\alpha_e$. These values are summarized in Table 3.1 and the corresponding best-fit off-critical data-collapse plots are given in Figures 3.4, 3.5, and 3.6. Using Eq. 12 $\beta$ was obtained below ($\beta^{-}$) and above ($\beta^{+}$) the transition along with its estimated error bars as propagated from $\alpha_e$ and each corresponding $\upsilon_{\parallel}$. These results are listed in Table 3.1. The average of the six values for $\upsilon_{\parallel}$ (subcritical and supercritical for each landscape) and the resulting calculation for $\beta$ are found in Table 3.2.

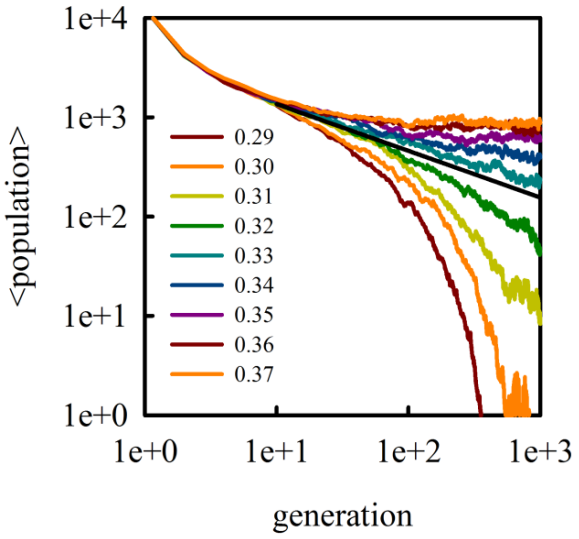


Figure 3.1. Critical-Quench Average Populations – 29x29 Landscape. The estimated line (solid black) has a slope of $\alpha_e$ = 0.47.

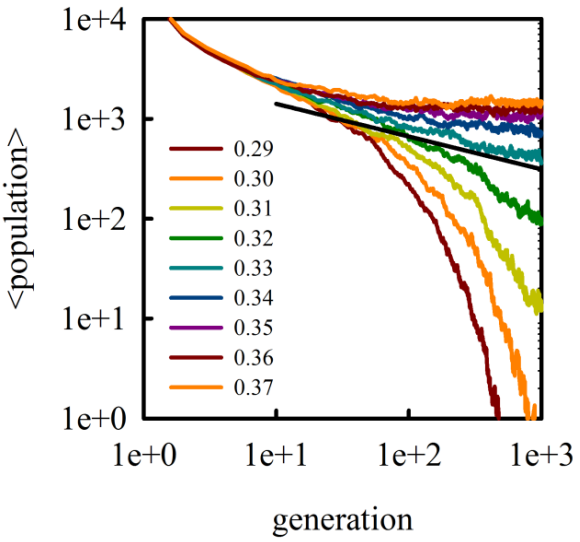


Figure 3.2. Critical-Quench Average Population – 37x37 Landscape. The estimated line (solid black) has a slope of $\alpha_e$ = 0.42.

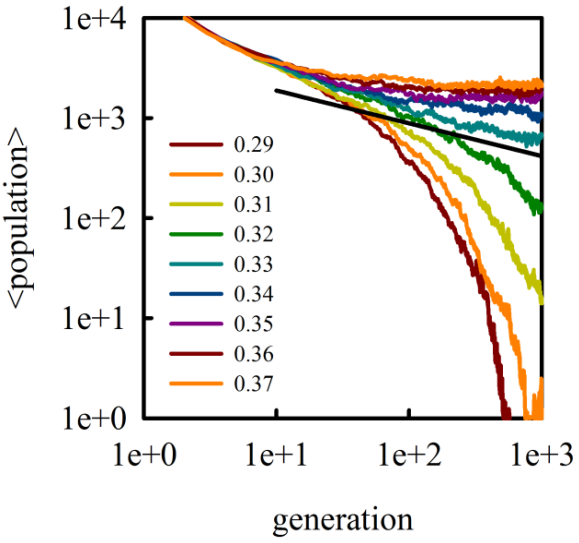


Figure 3.3. Critical-Quench Average Populations – 45x45 Landscape. The estimated line (solid black) has a slope of $\alpha_e$ = 0.42.

Table 3.1. Estimated and Best-fit Critical Exponents – Critical-Quench.

| $L$ | $\mu_{ce}$ | $\alpha_e$ | $\upsilon_{\parallel}{}^{-}$ | $\upsilon_{\parallel}{}^{+}$ | $\beta^{-}$ | $\beta^{+}$ |
|---|---|---|---|---|---|---|
| 29 | 0.3275 | 0.47(14) | $1.372^{+0.0.93}_{-0.095}$ | $1.349^{+0.083}_{-0.121}$ | $0.645^{+0.233}_{-0.234}$ | $0.634^{+0.225}_{-0.243}$ |
| 37 | 0.3275 | 0.42(14) | $1.313^{+0.083}_{-0.117}$ | $1.247^{+0.073}_{-0.078}$ | $0.548^{+0.216}_{-0.230}$ | $0.521^{+0.203}_{-0.205}$ |
| 45 | 0.3275 | 0.42(14) | $1.309^{+0.108}_{-0.063}$ | $1.214^{+0.092}_{-0.094}$ | $0.552^{+0.229}_{-0.210}$ | $0.512^{+0.209}_{-0.210}$ |

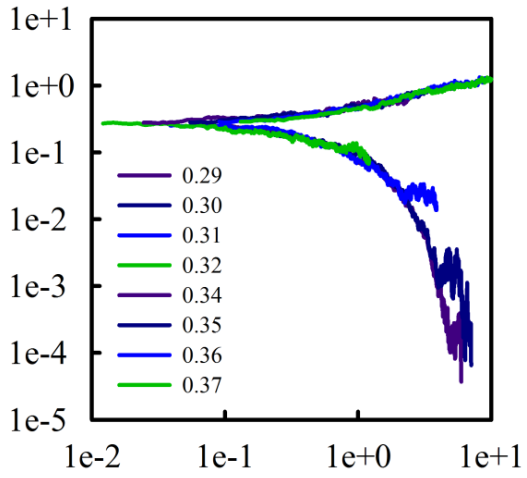


Figure 3.4. Off-Critical Data-Collapse – 29x29 Landscape. The data is plotted according to the function of Eq. 15. Subcritical values in the bottom universal curve were for $\mu \leq 0.32$, and supercritical values in the top curve were for $\mu \geq 0.34$.

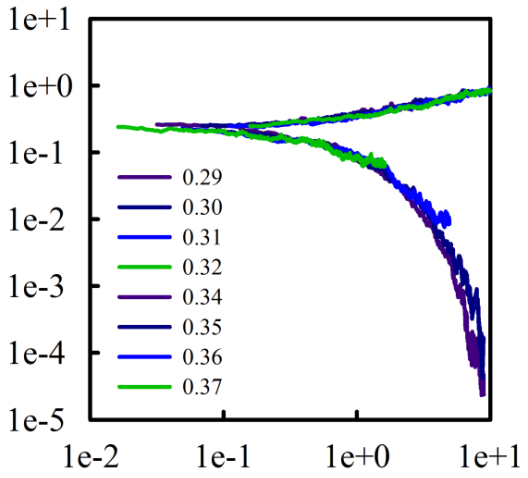


Figure 3.5. Off-Critical Data-Collapse – 37x37 Landscape. The data is plotted according to the function of Eq. 15. Subcritical values in the bottom universal curve were for $\mu \leq 0.32$, and supercritical values in the top curve were for $\mu \geq 0.34$.

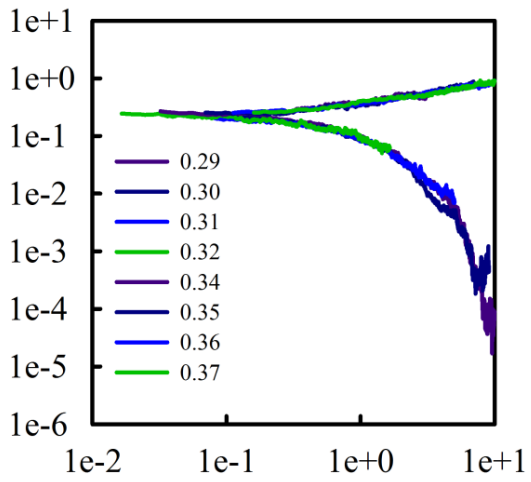


Figure 3.6. Off-Critical Data-Collapse – 45x45 Landscape. The data is plotted according to the function of Eq. 15. Subcritical values in the bottom universal curve were for $\mu \leq 0.32$, and supercritical values in the top curve were for $\mu \geq 0.34$.

The $z$ exponent for the universal function of finite-size data-collapse was determined from the minimization of the goodness of fit measure, $P_2$ from Eq. 18. This was done for curves from each landscape size at a supercritical value of $\mu = 0.33$, and using $\langle \alpha_e \rangle$ and $\langle \upsilon_{\parallel} \rangle$ for the $\alpha$ and $\upsilon_{\parallel}$ values in Eq. 18. The value of $z$ was determined, and the resulting value for $\upsilon_{\perp}$ was calculated from Eq. 14. These results are summarized in Table 3.2. Figure 3.7 shows the corresponding finite-size universal scaling function.

Table 3.2. Average and Best-Fit Critical Exponents – Finite-Size.

| | $\langle \alpha_e \rangle$ | $\langle \upsilon_{\parallel} \rangle$ | $\langle \beta \rangle$ | $\upsilon_{\perp}$ | $z$ |
|---|---|---|---|---|---|
| Calculated | 0.44(14) | $1.301^{+0.089}_{-0.095}$ | $0.568^{+0.222}_{-0.222}$ | $0.799^{+0.103}_{-0.138}$ | $1.63^{+0.18}_{-0.26}$ |
| Theoretical | 0.454(1) | 1.295(6) | 0.584(4) | 0.734(4) | 1.76(3) |

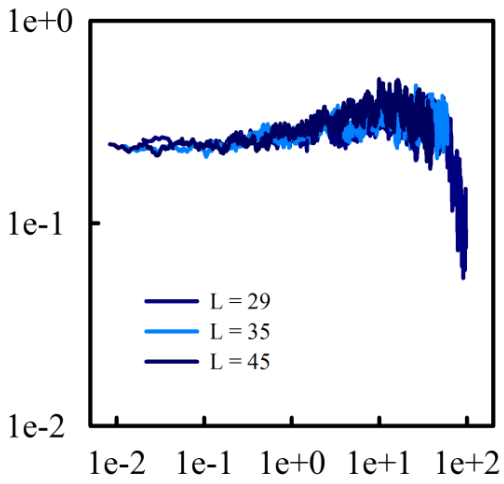


Figure 3.7. Finite-Size Data-Collapse – $\mu = 0.33$. The data is plotted according to the function of Eq. 16.

### 3.6. DISCUSSION

The results above clearly indicate that the non-equilibrium phase transition in this continuous space, spatial branching process model of phenotype evolution belongs to the directed percolation universality class. This system is a special case of evolution where there is no variation in the fitness landscape, reproduction is asexual, and organisms exist in sympatry where they are ensured local interactions. Neutrality, asexual reproduction, and sympatry should be tantamount to the universality classification because of the reaction-diffusion system they invoke. For example, if organisms were able to reproduce with some alternative scheme such as random mating, then it follows from the results of Section 3 that the phase transition behavior is lost. In the case of assortative mating, the particle process of reproduction changes to 2A→2A, and its universality classification should belong to PCPD (Park 2011). Täuber et al. (2005) state that directed percolation can exist with either the annihilation (2A→0) or coalescence processes. With the exchangeability between annihilation and coalescence, if given some other neutral fitness level with an odd number of offspring (for example, three), then the variation of this system would contain the particle operation A→3A or A→(m+1)A, m = 2, and therefore belong to the PC or BARWe universality class (branching-annihilating random walk with

even offspring as denoted by m) (Täuber et al. 2005). Although similar phase transition behavior was observed with rough fitness landscapes in Dees and Bahar (2010), the DP transition is likely lost through the form of spatial disorder (discussed in more detail below) (Dobramysl & Täuber 2008). Another aspect of the phenotype space is the interaction with the boundaries. Should organisms be considered to develop new phenotypes or absorb into the boundaries due to insufficient protein production as could be considered here? If the phase transition behavior were driven by boundary activity, then new dynamical features can arise (Henkel & Schütz 1994; Fröjdh, Howard, & Lauritsen 2001; Barato, Bonachela, Fiore, Hinrichsen, Muñoz 2009).

The model described here model may be used as a null hypothesis, much like other studies where the neutral fitness or genetic drift is a null hypothesis (Kimura 1983; Dieckmann & Doebeli 1999). Applications may be in paleobiological systems where the only data known from fossil records are morphology (Abe & Liebermann 2012; Foote 1990). The model could easily be scaled up to account for more phenotypes, which would increase the dimensionality of the space. Universality classes are allowed different accessible dimensions for the agents, but the exponents are known to converge on the mean-field theoretical predictions when the dimension of the space reaches the upper-critical dimension. Above the upper critical dimension, the exponents and universal behavior conform to mean-field predictions. For DP, the upper-critical dimension is four, and thus in the trilobite studies by Abe and Liebermann (2012), as well as in Foote (1990), where twelve or more phenotypes are considered, the dimension of their spaces would be well above the upper-critical dimension for DP and likely most other plausible universality classes. However, other issues arise when working with available fossil data. As noted by Abe and Liebermann (2012), the exact line of descent of any ancient species is not known in general. In addition, some radiations lack temporal resolution to assign parent-offspring species. These issues do not arise in this model and the species lineages are considered in Section 4.

For both goodness of fit measures, the minimum wells were asymmetric and therefore resulted in asymmetric error bars for both below and above the transition (Table 3.1). Even at 1% estimated error, the error bars on the exponents reported in the tables above are somewhat large in comparison with those found in other DP studies. Especially

in the reported $\langle\beta\rangle$, error bars are nearly half the average value. This is in part due to error propagation from $\langle\alpha_e\rangle$ and $\langle\upsilon_\parallel\rangle$ from which it was calculated (Eq. 12). All of the exponents indicate that they are theoretical DP exponents within a 1% margin of error, as measured from the goodness of fit.

The decision to cut off the generation ranges for the off-critical data-collapses at 1000 was due to the observation that for much later generations, population density decays in some mutability curves showed unusual changes in the decay behavior. This effect is likely the result of finite-size effects where the probability of survival is never absolute even in the survival regime, unlike what is theorized in the thermodynamic limit of a truly infinite space (Henkel et al. 2009). This causes the steady-states of supercritical populations, which would otherwise stay surviving, to reach extinction. The relaxations were accurately observed since the generation range used was greater than calculated correlation times (on the order of $10^1$-$10^2$ generations) for each landscape. This can be calculated with the measured average correlation time exponent as given in Eq. 3.3. Likewise, from the power-law scaling of correlation length given $\upsilon_\perp$ in Eq. 3.4, the curves for $\mu = 0.33$ allowed accurate observation.

Only recently DP was reliably observed in an experimental setting (Takeuchi, Kuroda, Chaté, & Sano 2007 & 2009). The lessons from observing DP have been that short generational time scales are needed in order to observe long-time system dynamics and to have a reliable system free of noisy absorbing states. There are numerous potential biological models, ranging from predator-prey models, to species coexistence models, to calcium waves among cells which may belong to the DP universality class (Park 2011; Reinhardt, Bohm, Drossel, & Hinrichsen 2006; Timofeeva & Coombes 2004). In a tumor growth model, Lipowski et al. (2012) observed an unusual set of critical exponents appearing as mean-field DP, even though universal scaling functions and some critical exponents clearly exhibited non-mean-field DP, suggesting some biological systems can unexpectedly challenge or enhance universality theories. Populations of organisms undergoing reaction-diffusion processes could also be potential model systems from which to observe DP. However, DP may be observable only in small-scale organisms such as viruses, bacteria or yeast, since their reproductive cycle can be relatively short. An absorbing state of biological populations, extinction, does not suffer from natural

noisiness common to many other suggested DP systems which may suffer from quenched disorder (there are no zombies). The reaction-diffusion processes of some organisms can be modeled in physical space, genetic space, or phenotype space.

One must be careful to note any inhomogeneity in a model when considering its universality class. The DP conjecture and Reggeon Field Theory, from which DP was originally derived, require that a system must not contain spatial or temporal quenched disorders. Spatially quenched disorder occurs in many systems such as catalytic reactions or epidemics (Hinrichsen 2000). The absorbing state may be reached everywhere, but localized regions may not be able to reach the absorbing state or are returned to the active state in the presence of noise at the absorbing state. For example, in catalytic reactions under appropriate conditions, such as the Ziff-Gulari-Barshad (ZGB) catalytic reaction model of $CO+O \rightarrow CO_2$ on a platinum surface, belongs to DP (Ehsasi, Matloch, Frank, Block, Christmann, Rys, & Hirschwald 1989). However, experimental verification in Ehsasi et al. (1989) did not reveal DP behavior. A variety of suggested reasons exist, such as possible defects in the catalytic reactions (Zambelli, Wintterlin, Trost, & Ertl 1996; Hinrichsen 2000). Similar issues in real systems make it difficult to observe DP, and delayed its experimental verification (Takeuchi et al. 2007; Hinrichsen 2000). In a biological system such as endangered species, this problem is unlikely to occur, since extinct species cannot produce new species. Yet, if the system were the entirety of the tree of life on earth, then such a system may not be subject to extinction. If the environment after a global extinction could support the production of life through some recipe of a "primordial soup, crêpe, or pizza", then extinction no longer describes a true absorbing state (Bernal, Oparin, Mueller, Haldane, & Synge 1967; von Kiedrowski 1996; Smith & Szathmáry 1995; Ferreira & Fontanari 2002). The "recipes" are theories of abiogenesis describing how the first self-replicating life forms came from naturally occurring chemical processes on ancient earth. Dynamically, abiogenesis takes the form of spatially quenched disorder, so the dynamics of life on earth as a whole cannot possibly belong to DP. However, life would almost certainly take an alternative evolutionary path like in the thought experiment of "replaying the tape" (Gould 1990).

Temporal disorder occurs for systems where the dynamics are not consistent through time. Temporal disorder has been studied by Jensen (1996) where a 1D lattice

was subjected to the usual DP reaction-diffusion process. However, as the probability that a row of sites all propagate with certainty between two time steps was increased, the critical exponents were observed to deviate from the DP case where this absolute survival probability went to zero. The model presented in this work may incorporate this type of disorder, because, while the random death is calculated in the same way at every time step, the percentage of organisms subject to random death varies from one generation to the next. However, this disorder does not appear to have a significant effect, since the measured exponents show minimal deviation from the theoretical exponents. For comparison, Jensen (1996) saw changes from known DP exponents by around 6-14% when there was a 25% chance that no random death occurred in a time step. Percent differences of the critical exponents reported here are on similar order of change.

If the population-dependent random death occurred before, rather than after, the competitive death process, different population fluctuations could arise due to the non-commutative nature of the death processes. In a variant of the neutral model, where the amount of random death is not population dependent and with assortative mating, results obtained are similar to those described above (having critical behavior and clustering) and the order of death processes are commutable (King et al. in preparation). The results of the critical exponents may also be closer to the theoretical DP values due to the assurance that temporal disorder is no longer a murky issue as it is here. However, an investigation of the critical exponents and scaling behavior of the related model is needed to confirm the universality class. The related model is mathematically more tractable than the one focused on here, and the governing partial differential equation closely resembles a typical Langevin equation with the caveat that the coalescent coefficient varies as a function of mutability. Despite the complexity of the model presented here with regard to the random death process, the dynamics are within the realm of systems belonging to the DP universality class. The variable stochasticity of random death may cause the critical point to exist as a small interval which may therefore have uncorrelated critical behavior of the usual quantities of interest when performing the universality determination. This is perhaps an enlightening notion for renormalization group theory, since the model presented appears to have a unique style of particle processes. Likely, the mathematical intractability is to blame, but the mathematics can still be written out and future

computational studies of related mathematically intractable systems may shed even more light on the universal behavior.

A potential use of a phenotype evolution model belonging to DP is the possibility to model genus or species decay/formation. In phenotype data, one might infer that mutability of populations of some animals may not be great enough to sustain populations in their environments. The phase transition indicators from the populations of this model are robust with regard to variations of the fitness landscape, the noisy local process of reproduction, and the implementation of random death (Dees & Bahar 2010; Scott et al. 2013; King et al. in preparation). The expectations to observe supercritical mutability in populations may be translated for a variety of situations involving different phenotype space dimensions and scales. The phase transition behavior, having no characteristic scale, could allow for investigation of multiple levels of taxonomy. It is to the problem of multilevel selection that is the central theme of the next section.

# 4. ORGANISM AND CLUSTER LINEAGE DYNAMICS

## 4.1. INTRODUCTION

In this section, some properties of organism and cluster lineages are investigated. Relationships between parents and their offspring are naturally defined by temporal bonds, and can be mapped onto problems such as DP. In DP, measurements on the emerging infinite cluster (along the temporal direction) provide information on correlation times. Scaling of the correlation time (Eq. 8) is determined by the off-critical measure and the correlation time critical exponent (Hinrichsen 2000). The correlation time exponent may also determine the scaling distribution of lineage lifetimes. Lineage structures such as splitting events (offspring production) relate to fitness measures. Although organisms are allowed no more than two offspring, clusters may produce many more clusters. By measuring offspring cluster production, a measure of cluster fitness will be analyzed as a precursor to the possibility of multilevel selection in the model.

**4.1.1. Cluster Interactions**. Temporal, or directed, clusters have been studied in systems such as directed lattice animals (Bousquet-Mélou 1996; Marckert 2012; Bacher 2013), turbulent puffs and slugs (Sipos & Goldenfeld 2011), and in the context of multifractal graph-like structures (Feder 1988; Norton & Tandy 1999; Moon 1992; Berestycki 2003). A lattice animal is described in Stauffer and Aharony (1992) as a cluster of lattice sites, which can take different forms for a given number of sites in the cluster. For example, a cluster of three sites on a two-dimensional square lattice can take the shape of two unique lattice animals, where one is a line of the sites, and the other is in the shape of an elbow. A directed lattice animal (DLA) exists on an oriented lattice where it has at least one root vertex from which connected sites along a path in the preferred direction can be accessed from the root (Bousquet-Mélou 1996). The interaction between directed clusters (lineages) must be specified, because differences in the interaction rules produce different scaling properties. In the case of non-interacting DLA, the asymptotic scaling of lifetimes followed a power-law according to the DLA area, or number of sites in the DLA (Bousquet-Mélou 1996). Knežević and Vannimenus (2002) also observed critical behavior of interacting DLA. In particular, they measured critical exponents of the temporal and spatial asymptotic power laws of measured times between a root site

and its connected sites and the spatial distances between connected sites. The interacting DLA problem studied was framed as a bond percolation problem, and critical exponents of the interacting DLAs were shown to fit expectations of bond percolation.

In another system with interacting clusters, cluster lifetimes were found to scale superexponentially (Sipos & Goldenfeld 2011). They considered a fluid flowing through a pipe, and studied the lifetimes of different turbulence formations as the viscosity parameter, Reynolds number, was varied. Clusters of turbulence that did not span the pipe geometry are called puffs, and turbulent clusters that spanned the pipe are called slugs. Puffs could merge or split and the puff lifetime scaling depended upon the Reynolds number.

It is argued here that directed lattice animals represent organism lineages best; whereas turbulent puffs are more similar to cluster lineages. Individual lineages in the asexual fission model are non-interacting in the sense that no two lineages may merge (see Figure 4.1). This does not change the behaviors noted in previous sections, since the organisms still undergo the same dynamical processes within the phenotype space. Cluster lineages are more complex, since they can split and merge with other cluster lineages (Figure 4.1b). These points are emphasized in greater detail below in Subsection 4.1.3.

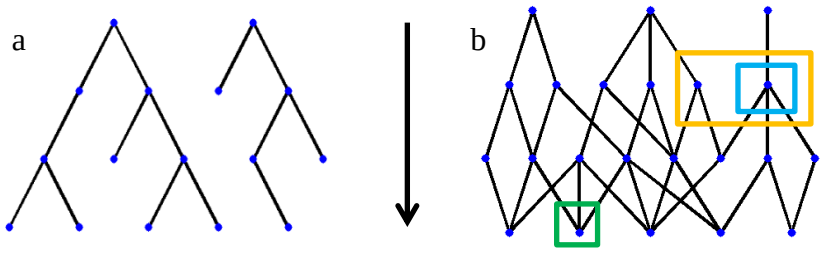


Figure 4.1. Examples of Organism and Cluster Lineages. (a) Organism lineages cannot interact (merge) and each parent has at most two offspring. (b) Cluster lineages can interact and each parent cluster can have many offspring clusters and each offspring cluster can have many parent clusters. The downward arrow indicates the forward-in-time direction. The boxes indicate examples of backward-in-time events: (blue box) multiple coalescence, (orange box) simultaneous multiple coalescence, and (green box) fragmentation (see Subsection 4.1.3).

It was shown in the preceding sections that the asexual fission model undergoes a DP phase transition as the mutability is increased to some critical value. From DP theory the population density, correlation length, and correlation time are known to exhibit scale-free behavior in the critical range of the control parameter consistent with DP. Dynamical properties of clusters about a non-equilibrium phase transition are of interest due to the scale-free behavior.

Biologically, this could model multiple levels (multilevel) of selection (Traulsen & Nowak 2006; Okasha 2009; Damore & Gore 2012). There are two descriptive classes of multilevel selection (MLS), and they describe the fitness of groups (species or clusters) with respect to the organisms within it or the clusters it produces (Damuth & Heisler 1988). The MLS1 class describes the cluster fitness as being dependent upon the collective fitness of its organisms, such as the average organism fitness. In contrast, the MLS2 class describes the cluster fitness as the number of offspring clusters produced by it. For the organisms, heritable traits, such as altruistic behavior, have been used to increase the survivability of some groups (Hamilton 1964a & 1964b). Evolving behavioral traits has been a contentious and intriguing topic among evolutionary biologists, because it deviates from the notion of organism competition by introducing organism evolutionary benefits from participation in a group (Okasha 2005). However, in the neutral model here, there is no heritability characteristic that can be maintained by clusters unlike the mutability held by the organisms. Instead, clusters of organisms are limited to measures such as cluster-size (number of organisms), centroid (center of mass, see below), or gyration radius. Even the cluster fitness depends upon the cluster-size and gyration radius (see below). The scale-free dynamics of the model paired with the concept of multilevel selection in evolutionary biology allows one to ask whether the scale-free behavior arising from the phase transition translates into scale-free selection? This is the primary motivation for this section.

**4.1.2. Graphs.** Lineages behave differently for organisms and clusters, so it is useful and important to define lineages according to known mathematical structures. Graph theory provides a mathematical description for such constructs.

Consider a set of nodes V, that may connected by a set of edges E. The collection of the set of nodes and edges defines a graph G = (V, E). A simple graph is one in which

no two nodes have multiple edges connecting them. The connections between nodes by the edges allows one to traverse the graph, and can be thought of as starting at one vertex, then jumping from vertex to vertex according to edge connections to a final vertex. Such traversals are called walks. A path is a type of walk where no vertex or edge is used more than once. A cycle is a walk whose starting vertex is also the final vertex, and an acyclic graph is a type of graph that contains no cycles. Simple graphs with cycles are sometimes referred to as pseudographs. Another specific type of graph that matches the structure of organism lineages is called a tree. Trees are graphs which are acyclic simple graphs. If a root of the tree is defined, then a tree whose edges lead away from the root is called a rooted tree. Finally, a set of disjoint trees is called a forest (Diestel 2012).

The graphs of interest in lineages are rooted trees for organism lineages, and rooted trees on which cycles are allowed for cluster lineages. If organisms are the node set and parent-offspring relationships define the edge set, then an organism lineage is just a rooted tree. Clusters interact in a more complicated manner; since they may merge and split by "trading" offspring organisms or causing divided groupings among offspring organisms (see Subsection 4.2.2).

**4.1.3. Coalescent Theory.** Coalescent theory provides a mathematical construction for lineage merging and splitting events, more commonly referred to coalescent-fragmentation processes (CFP) (see Berestycki 2009 for a review). The mathematical language is left arbitrary with respect to different organizational levels. That is, it could be applied for a single level such as for the genealogies of organisms or for species phylogenies. The theory is related to physical models of fragmentation and coalgulation such as the Smoluchowski equations and Marcus-Lushnikov process (Aldous 1999; Bertoin 2006). Aldous (1999) and Bertoin (2006) provide some linkages between more physical models and some more mathematical models of coalescent theory. The terminology used throughout will follow this convention: forward-in-time branching events describe *splitting* and colliding events describe *merging*; backward-in-time branching events describe *fragmentation* and colliding events describe *coalescence*.

The theory has its roots in the Galton-Watson (GW) branching process which was initially constructed to investigate the probabilities of extinctions of family surnames passed along a patriarchal line (Watson & Galton 1875). Their work initialized the

formulation of temporal measures on tree-like structures. About a century later, in a more rigorous construction of lineage coalescence times, Kingman (1982) described binary coalescence events occurring one at a time in haploid Wright-Fisher (WF) populations, called Kingman's n-coalescent, where n represents the number of randomly chosen descendants for whom one finds their most recent common ancestor (MRCA). WF models evolve in a different way than the BCRW populations of the present model such that, in the latter, offspring randomly choose their parents.

Kingman's coalescent was extended by Pitman (1999) to allow one instance of multiple coalescing lineages in a single time step, the Λ-coalescent. Independently, Schweinsberg (2000) and Möhle and Sagitov (2001) developed the theory to include simultaneous Λ-coalescent events in a single time step, the Ξ-coalescent. Berestycki (2004) studied the mathematical properties of Ξ-coalescents with fragmentation under the assumption that they were exchangeable in the case of coalescents and homogeneous in the case of fragmentation. This means that none of the properties of the lineages would influence the rates at which they undergo CFP. For example, if one assumes that the lineages in question are of clusters, then the number of agents within each cluster or any spatial extent that the cluster may cover will *not* affect the cluster lineage structure. Several descriptions of CFP have arisen to place dependencies on spatial descriptions, but the mathematical theory at present describes only spatial Λ-coalescents (Durrett & Limic 2002; Limic & Sturm 2006; Angel, Berestycki, & Limic 2009).

Upon an initial introduction to CFP, one might be quick to think that coalescence and fragmentation are dual processes (that the reverse of one has the same behavior as the other). However, it has been observed in some situations that duality is not a general property to CFP, especially in the case of the stochastic coalescent (Bertoin & Goldschmidt 2004; Aldous 1999). The stochastic coalescent was formulated upon cluster masses driving the coalescence rates, rather than the number of lineages as for Kingman's coalescent (Aldous 1999). More recently, universal tree structures have been explored by investigating the ratios of average times to MRCA with differing $n$ for Kingman and Bolthausen-Sznitman coalescents, a model for natural selection that is also a special case of Λ-coalescents (Brunet, Derrida, & Simon 2008; Brunet & Derrida 2013). The ratios also classify different universality classes of tree structures. Despite all of these

advancements, currently there is no mathematical description for spatial Ξ-coalescents with fragmentation (N. Berestycki, personal communication).

**4.1.4. Time to Most Recent Common Ancestor – First-Passage Time.** A common statistical measure regarding different CFP systems is the average time to MRCA from among $n$ starting descendants, $\langle T_n \rangle$. For stochastic processes, the mean first-passage time (MFPT) problem asks how long, on average, a trajectory takes to reach some intended target. For example, in the classic case of a drunkard walking down a street (whose stumbling walk mimics Brownian motion), one could ask how long, after leaving a bar it takes the inebriated person to get to the other side of the street. After averaging over the departing, clumsy patrons, one could then determine a MFPT. This is a type of MFPT problem, and an analytical solution is known for this type of diffusion called standard Brownian motion (continuous space and time), also known as the Weiner process, to reach a stationary spatial boundary in 1+1 dimensions (Lesne 1998).

A variety of systems have been considered regarding the type of target a walk reaches including: moving barriers (Tuckwell & Wan 1983), a specific finite sized target within the space (Bénichou & Voituriez 2014), and multiple finite-sized targets (Chevalier, Bénichou, Meyer, & Voituriez 2011). The space in which an object can maneuver is also important, but in certain cases, the distribution of FPT is known to exhibit universal features regardless of spatial properties, such as discrete, continuous, or fractal media (Condamin, Bénichou, Tejedor, Voituriez, & Klafter 2007; Bénichou, Chevalier, Klafter, Meyer, & Voituriez 2010; Bénichou & Voituriez 2014). Properties of MFPT have traditionally been explored with infinite spatial area or volume, but recent progress has been made for confined, finite volumes with reflecting boundaries (Bénichou, Chevalier, Klafter, Meyer, & Voituriez 2010; Bénichou & Voituriez 2014). The majority of the above systems involve random walkers without both branching and coalescing dynamics. The temporal properties of BCRW systems has only recently begun for systems involving BCRW by Dutta et al. (2013) that found coverage time measures (visiting all vertices of a graph) for a variety of graph structures.

**4.1.5. Summary.** The primary focus of this section will be to determine how $\mu$ changes the CFP for both organism and cluster lineages as it drives the system through the DP transition to the uniformly distributed populations. To do so the MFPT problem

on lineages will be addressed. Specifically, organism lineages in the model appear to represent a spatial Ξ-coalescent, whereas the cluster lineages appear to be structured as spatial Ξ-coalescents with fragmentations (Figure 4.1). To the knowledge of the author, the cluster lineages represent a unique problem not yet explored mathematically. The CFP of lineages do not contain the exchangeability and homogeneous properties of coalescents and fragmentations that have been previously studied, because the lineages should be dependent upon how populations fill the space and clustering properties (which can be measured by the NN index $R$).

The organism lineage lifetimes distribution will be determined for each value of $\mu$, and power-law scaling is expected at criticality due to the temporal nature of the infinite cluster in DP as well as from directed lattice animals. Fitness and spatial measures on the cluster centroids are taken to provide a better sense of how clusters might interact. The cluster centroid step-sizes will be measured as well as the NN index $R$ (Eq. 3) for cluster centroids. In an attempt to measure universal tree behavior, ratios of the average times to MRCA between n = 2, 3, and 4 will be measured as a function of $\mu$. These results will be discussed along with their evolutionary impact to multilevel selection.

## 4.2. METHODS

Simulations of the neutral asexual fission model were run in the same manner as in Section 2; however, lineage algorithms are introduced here. Data was collected from nine runs at each value of $\mu$ on a range encompassing the critical point, from 0.30 to 0.45 in increments of 0.01 units.

**4.2.1. Organism Lineages – Genealogies.** Parent-offspring relationships are relatively simple to define. Organism lineages, or genealogies, in the model are rooted binary trees. A parent will have at most two offspring according to the fitness landscape described above. Neither the parent nor its offspring will become related to any other organism from their respective populations (Figure 4.1). Even when two offspring from different parents undergo a competition event (coalesce), their lineages do not merge. Therefore, an organism only links to their parent from the previous generation and their offspring (if any) in the next generation. Given an organism from some generation (such

as the first generation, one of the first 300), its line of descent is extinguished when the last of its descendants die. Additionally, extinction of a population occurs when all lines of descent have gone to extinction.

Genealogy lifetimes, $\tau$, are measured from the original 300 organisms in the first generation. Probability density distributions of $\tau$ for different values of $\mu$ are measured from each of the initial populations from nine simulations, giving a sample size of 2700.

**4.2.2. Cluster Lineages – Phylogenies.** Cluster lines of descent, phylogenies, are much more complicated than for the organisms. This is due in part by the merging-splitting actions of clusters. Consider a cluster of organisms; each organism will produce offspring according to a Markov branching process. Let the cluster be of minimum size (three) and let those organisms in the cluster be labeled $\Psi_i$, where i is 1, 2, or 3. Let their offspring be labeled $\psi_{ij}$, where j can be 1 or 2. Assuming all of the offspring survive, suppose that three offspring, $\psi_{11}$, $\psi_{12}$, and $\psi_{32}$ form a cluster separate from the others (with $\psi_{21}$, $\psi_{22}$, and $\psi_{31}$ in the second cluster). In the resulting offspring generation, two clusters are formed from the splitting of the parent cluster. Depending on the number of organisms in a cluster, the maximum number of offspring clusters, F, can be greater than two, representing an expanded fitness limit on the cluster level. The maximum number of offspring clusters a parent cluster can possibly produce depends upon the minimum cluster size (three) and the parent cluster size, s

$$F(s) = floor(s/3). \tag{18}$$

The floor function is a rounding calculation that rounds down to the nearest integer. This means that the cluster fitness ranges from zero to F(s).

One could reason that if two clusters of the same size were spread out differently, such that the gyration radius of one cluster was significantly different from the other, then the cluster with a greater gyration radius will likely have more offspring organisms. More offspring could survive because they would be less likely to die from competitive death. Therefore, the gyration radius of clusters may also influence the number of offspring clusters between clusters of the same size.

Clusters also merge when offspring from separate parent clusters join into a single cluster. This allows for an offspring cluster to have one or many parent clusters. This type of interaction is absolutely not allowed at the organism level, since the coalescence of organisms results from competitive death. The coalescence event is therefore of a different flavor at the cluster level. There is no death of a cluster (removal) based on phenotypic proximity.

The probability density distribution of cluster splits and mergers were calculated for each value of $\mu$. Only clusters among the last half of surviving generations were used to generate the distributions. Additionally, average cluster fitness of the same set of clusters was calculated as a function of $\mu$.

**4.2.3. Cluster Centroid Step-Sizes.** Cluster centroids are calculated as the center of mass of constituent unit mass organisms and their locations

$$C_x = \frac{1}{s}\sum_{i=1}^{s} c_x, \tag{19a}$$

$$C_y = \frac{1}{s}\sum_{i=1}^{s} c_y. \tag{19b}$$

Step-sizes, Γ, are then defined as the distance between each parent cluster centroid ($C_p$) and each of its offspring cluster centroids ($C_b$):

$$\Gamma = \sqrt{\left(C_{px} - C_{bx}\right)^2 + \left(C_{py} - C_{by}\right)^2}. \tag{20}$$

Probability density distributions of Γ were created for values of $\mu$ = 0.30, 0.33, 0.36, and 0.39. Only the clusters from the last half of surviving generations were used, regardless of whether populations decayed to extinction. Average Γ vs. $\mu$ was also calculated based on the same set of clusters.

**4.2.4. Average Time to Most Recent Common Ancestor.** Given some number of randomly chosen clusters in the same generation, the average time to MRCA, of the chosen cluster lineages can be calculated. This is a backward in time calculation, starting from a later generation and following lineages back toward their first root ancestor. The first root ancestor (MRCA) can be considered the point at which all lineages of the

original selected clusters coalesce. Average times to MRCA were determined for four different initial groups of clusters with $n$ = 2, 3, and 4. Only the last three-quarters of surviving generations that also contained at least twenty clusters were considered for starting points. From among these, 1000 random generations were selected for sampling. This was done for each $n$ and for each value of $\mu$.

### 4.3. RESULTS

Genealogical lifetimes are shown as a normalized probability density distribution in Figure 4.2. Near the simulated critical point 0.33, an asymptotic power-law tail was observed with slope $-1.282 \pm 0.072$.

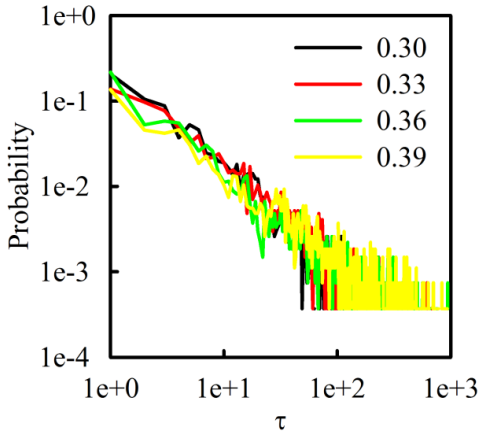


Figure 4.2. Probability Density Distributions of Organism Genealogy Lifetimes.

Cluster splitting behavior as it depends upon $\mu$ is visualized in Figure 4.3. The splitting events count how many offspring clusters were produced for each cluster, so Figure 4.3a also represents the cluster fitness distribution. Since clusters may not produce any offspring clusters, a zero splitting event, all events were shifted by one so that the double-logarithmic plot could show the probability of cluster death events. A power-law tail was observed for values of $\mu$ near the critical point. The average cluster fitness,

number of offspring clusters produced by each cluster, $\langle F \rangle$ is plotted with respect to $\mu$ in Figure 4.3b.

Cluster centroid step-size statistics are shown in Figure 4.4. The probability density distribution of the cluster centroid step-sizes is shown in Figure 4.4a. For all values of $\mu$, the distributions were found to be bimodal. Average $\Gamma$ as a function of $\mu$ is linear in Figure 4.4b.

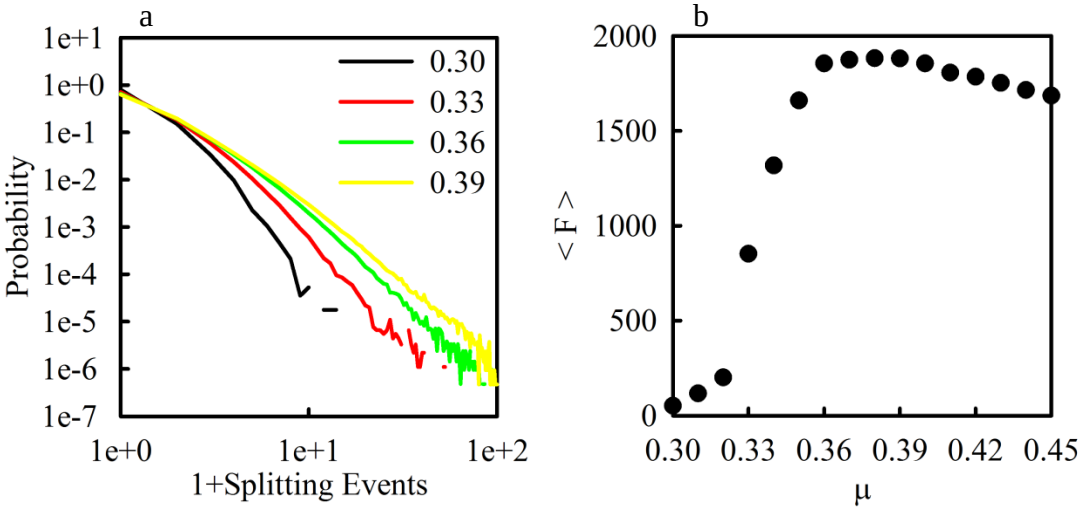


Figure 4.3. Cluster Splitting Events. (a) Probability density distributions of cluster splitting events are shown for $\mu$ = 0.30, 0.33, 0.36, and 0.39. To produce the double-logarithmic plot, one was added to the splitting events such that 1 represents the complete death of a cluster, 2 represents a cluster producing one offspring cluster, etc. (b) Average cluster splitting events (cluster fitness) is shown as a function of mutability.

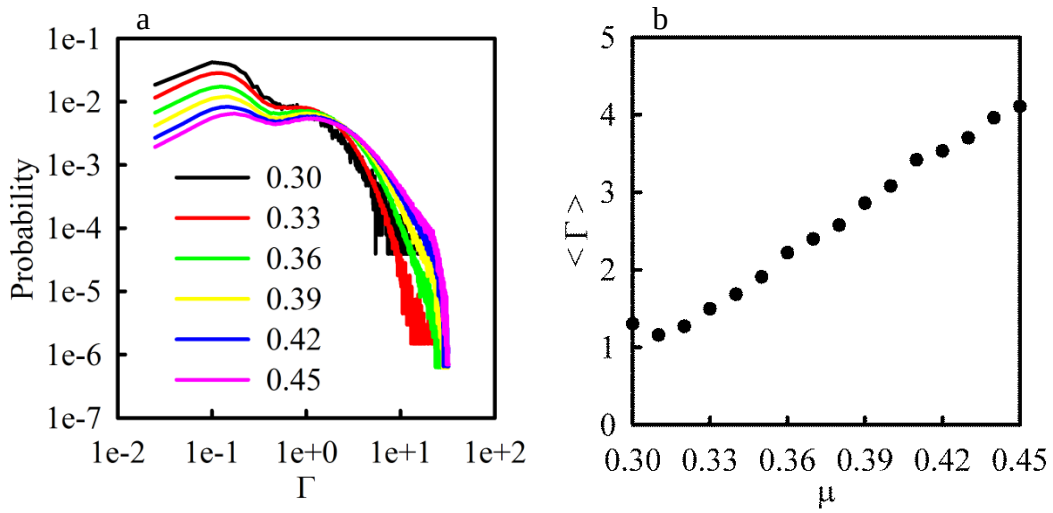


Figure 4.4. Cluster Centroid Step-Size. (a) Probability density distributions of cluster centroid step-sizes. (b) Average cluster centroid step-size as a function of mutability.

Cluster centroid NN index $R$ was calculated as a function of $\mu$ (Figure 4.5). The population of cluster centroids were found to be significantly aggregated only for $\mu < \mu_c$, and were significantly distributed according to a uniform distribution for $\mu > \mu_c$. At criticality, $\mu = \mu_c$, the population of cluster centroids were found to be distributed according to a purely random distribution. A peak was observed about the same $\mu_p$ as for the sample average number of clusters in Figure 2.6.

The ratios of average times to MRCA of the phylogenies is presented in Figure 4.6. The $\langle T_3 \rangle / \langle T_2 \rangle$ measure had a value of about 1.30 at criticality and decayed to roughly 1.1 at $\mu = 0.45$. The $\langle T_4 \rangle / \langle T_2 \rangle$ ratio had a value of 1.45 at criticality and decayed to approximately 1.1 at $\mu = 0.45$. The inset shows the standard deviation of each ratio; however, both curves are very close, with maxima at $\mu = 0.33$.

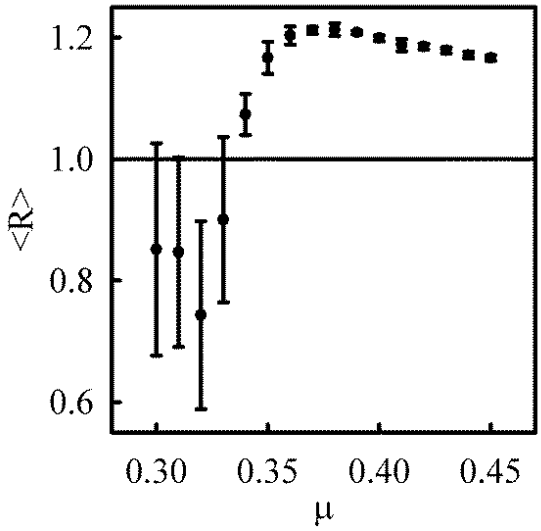


Figure 4.5. Sample Nearest-Neighbor Index – Cluster Centroids.

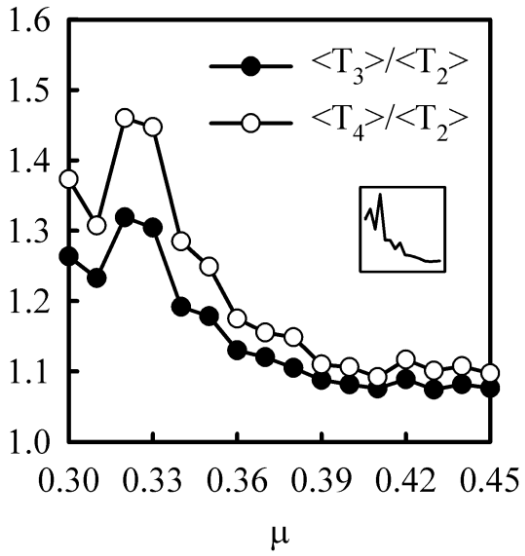


Figure 4.6. Ratios of Average Times to Most Recent Common Ancestor – Phylogenies. The dots indicate ratios calculated and the lines are added to aid the eye. The inset shows the standard deviation of the ratios. Both standard deviations fell almost identically onto each other, and they are of the same order of magnitude as the ratios. The peak standard deviation occurs for $\mu = 0.33$.

## 4.4. DISCUSSION

The above results indicate potential scale-free properties near the DP phase transition. Lineage lifetimes of the organism lineages suggest possible power-law scaling for the curve at $\mu = 0.33$. A best-fit on the tail from 10-100 generations showed the power-law scaling exponent to be $-1.282 \pm 0.072$. This result may point to the $\nu_{\|}$ exponent dictating the organism lifetimes (see Table 3.2). It is well known that the system lifetimes obeys this scaling at the critical point, and the correlation time exponent relates to the cluster size of directed lattice animals (Henkel et al. 2009; Park & Park 2011; Bousquet-Mélou 1996).

The splitting events observed in Figure 4.3 may also suggest scale-free properties. There does not seem to be a difference in the extinction probability regardless of $\mu$. Certainly, as $\mu$ increases, the populations begin to survive and grow in size. The number of clusters also increases, but one may think that cluster death, a zero splitting event, could reduce with increasing $\mu$. The results are contradictory to this notion, however.

With the power-law tail at $\mu = 0.33$, one could ask if this scale-free fitness demonstrates scale-free selection? In short, no. The reason for this is two-fold. First, there is no selection bias, no differential fitness, on the organism level. Clustering of organisms occurs, (Figure 2.10) but there is no selection for any particular traits in the phenotype space. Secondly, the clustering of clusters is on the order of the system size. The most basic unit in the system is an organism, so only two levels of evolving populations can be sufficiently addressed. Given genetic information driving the locations in phenotype space, as well as expanding the size of the space through dimensions or total volume could allow more levels of biological organization to be studied. Selection could then be investigated at the level of digital nucleotides, genotypes, organisms described by phenotypes, clusters (species), and super-clusters (genera). However, this endeavor would likely require much greater computing resources than those employed here.

Instead, a more appropriate question to ask is whether multilevel selection is present. In this case, maybe, multilevel selection of the MLS2 variety is indicated through clusters producing offspring clusters. This is absolutely the case here. What is questionable is which clusters are surviving. This cannot be answered with the current data presented. However, it may be inferred that there must be some survival of clusters

which yield more clusters, since those that produce more have a greater chance of survival. This is similar to the notion of Goodnight (2013), who suggested that organisms which are able to "cast a broader net" to have their descendants in more clusters are able to survive longer. The same could be said for the clustering in this model.

The average cluster fitness (Figure 4.3b) is quite reminiscent of Figure 2.5b and 2.6b, where average number of clusters was observed to rise steeply to a peak. For $\mu < \mu_c$, $\langle F \rangle$ was nearest to zero, grew sharply about $\mu = \mu_c$ until reaching a peak near $\mu = \mu_p$ and then declined again. Perhaps, this is not too unsuspected, since an increase in average fitness should lead to increased populations.

Figure 4.4 gives insights to the BCRW behavior of the clusters as they interact in the space. The distributions in Figure 4.4a show a decline in the leftmost peak as $\mu$ increases. This is likely due to the appearance of larger clusters granting greater step-sizes more often than the existence of smaller clusters interacting at shorter scales. This is likely the cause of the probability compression on the short-scale end and expansion on the longer-scale end. Bimodality in Figure 4.4a may be dependent upon three factors. At small $\mu$, there are far less large clusters than what exists for larger $\mu$. The larger clusters typically have greater spatial coverage (as could be measured by gyration radius), and therefore can offer longer-range jumps for any nearby clusters or through splitting itself. For large clusters, they tend not only to cover more space, but they are also more likely to contain many more organisms. Therefore, the cluster-size distribution (Figure 2.11) gives a cluster more chances to split apart. Indeed, this is supported by the increase in cluster fitness distribution seen in Figure 4.3a. There is likely a build-up in the longer-scale probabilities due to the finite-size of the landscape. Large clusters may not be able to sufficiently jump in all directions with the absorbing boundaries. Smaller hops may be lost at the edges of the landscape and contribute to a loss in shorter centroid hops. The boundaries may in general reduce the support for mid-range jumps, but deeper investigation is needed to confirm these predictions. The average step-size shows remarkable linearity with increasing $\mu$ (Figure 4.4b).

The spatial measures may be set into perspective with the NN index $R$ on the cluster centroids (Figure 4.5). For all active states of the system ($\mu > \mu_c$), cluster centroids are uniformly distributed. Only at criticality are cluster centroids in a purely

random distribution. Comparing this to the NN index on the organisms, Even when organisms were aggregated and surviving (recall $\mu_c < \mu < \mu_p$), their clusters appear to be distributed uniformly. The DP behaviors then manifest when organisms cluster and clusters are distributed purely randomly.

Ratios of the average times to MRCA are shown in Figure 4.6. The ratios $\langle T_3 \rangle / \langle T_2 \rangle$ and $\langle T_4 \rangle / \langle T_2 \rangle$ had values near 1.3 and 1.45 near criticality, respectively. Both decayed to just under 1.1 at $\mu = 0.45$. Interestingly, Kingman's coalescent produces ratios of $\langle T_3 \rangle / \langle T_2 \rangle = 4/3$ and $\langle T_4 \rangle / \langle T_2 \rangle = 3/2$ (Brunet, Derrida, & Simon 2008; Brunet & Derrida 2013). Whether there is true significance to the comparison of the ratio results is unclear. Cluster phylogenies structures are predicted to correspond to Ξ-coalescents with fragmentation, but at the DP critical point, the coalescent rates which determine average times to MRCA might still find correspondence with Kingman's results. No other known ratios could be found that fall within a relatively small range about the results. Furthermore, $\mu$ is shown to drive the CFP measures, forcing smaller ratios when the populations become uniformly distributed and forming large, indistinguishable clusters (Figure 2.2). This is a similar result seen by Brunet and Derrida (2012), in which a parameter tied to increased evolution rates continuously drove their system from having ratios described by Kingman to ones by Bolthausen-Sznitman. In much the same way, this effect is observed here. However, although the ratios appear to coincide with Kingman's at criticality, as $\mu$ increases, the ratios drop much lower than seen in the model by Brunet and Derrida (2012). This may suggest the influence of spatial constraints on the CFP structure and rates which are avoided in the generalized Λ-coalescent available in the model by Brunet and Derrida (2012).

The application of coalescents here is uncommon, since coalescents are generally applied to problems arising in population genetics (Teh, Blundell, & Elliott 2011; Zähle, Cox, & Durrett 2005; Li & Durbin 2011). Having shown this system belongs to the DP universality class, the upper critical dimension is four and greater than the dimension of the phenotype space studied here. Below the upper critical dimension, following cluster lineages in a phenotype space is suggested to be messy to the point that clusters are not well-defined entities (Lawson & Jensen 2008), but data suggests that, near the DP critical point, this system produces long lines of descent (Figures 4.2 and 4.6). This is despite the

fragmentation process which allows multiple parent clusters for some offspring clusters combined with the Ξ-coalescent events which corresponds to variability in cluster fecundity. However, well into the supercritical DP range, when populations better fill the phenotype space and organisms have greater branching mobility (due to greater $\mu$), clusters can grow much more massive. This allows for giant clusters to accumulate nearby clusters easier, and with the large noise in the system from organism coalescence death and random death, allows a once massive cluster to fracture into smaller clusters or join other nearby clusters. Greater space-filling populations experience greater amounts of trading of organisms between neighboring clusters which is the action that leads to greater mixing among cluster lineages. For populations with supercritical $\mu$, clusters do indeed become less well-defined as suggested by Lawson and Jensen (2008) and supported by Figure 2.10.

A long running debate between evolutionary biologists John Maynard Smith on one hand, and Elliott Sober and E. O. Wilson on the other, regarding group selection was addressed by Okasha (2006). It is generally well established that selection among individuals occurs, and is demonstrated by Okasha as natural selection acting upon the speed of antelopes. In the traditional view of group selection theory, one needs to show that a property of the organisms is inherited due to the benefit of belonging to a particular group. This would mean that a group of antelopes that run faster on average than a separate, slower group should have the evolutionary advantage to avoid fast predators. This, however, creates a hierarchical asymmetry in the process of natural selection at different levels, since the average group speed comes from a property of the organisms, not from the property of the group itself. The differences are subtle, but they can be parsed into better organized classifications having different theoretical constructions.

Okasha (2009) refers to different levels of selection based on organisms and their groups. When discussing the fitness of organisms in the models addressed in this work, the neutral landscape dictates that all organisms produce two offspring. This is the MLS1 modeling discussed by Okasha (2009). Typically, in more physical models, cluster masses are determined by the number of agents within them. Describing the coagulation-fragmentation processes of such a cluster is then based upon models such as Smoluchowski coagulation, Becker-Döring, or Marcus-Lushnikov models and their

variants (Aldous 1999). However, these processes are a mix of organizational levels since agent actions are built into the description of their cluster actions. Okasha (2009) refers to a more precise description of cluster behaviors, MLS2. At the MLS2 level, clusters become their own entities without regard to the agents from which they are composed. One can then talk about branching behaviors and/or fitness of clusters. The difference between MLS1 and MLS2 can be thought of as a sort of coarse-graining procedure like that in renormalization group methods and macroscopic modeling in statistical physics. Just as one does not need to know all of the microscopic details in order to know temperature, one does not need to know all of the organism behaviors in order to know the rate of cluster diffusion.

# 5. FUTURE DIRECTIONS OF RESEARCH

The BCRW processes exhibited by this model may be useful to further understand temporal behaviors in similar systems. In particular, advanced systems with a population of agents existing in a finite space with absorbing boundaries that contains moving, size-fluctuating targets are related to the cluster lineage dynamics. This is a stack of specifics, but they point to the need for a more generalized model of BCRWs. Several types of problems may be addressed by the cluster lineage dynamics. Applicable classes of models could be of epidemic dynamics of a localized aggressive contagion, or the genealogy of genetic mutations of migrating people. Additional insights could be gained through alternative model narratives such as in Pie and Weitz (2005), where in another 2+1 dimensions phenotype space, agents modeled lineages of species instead of organisms.

Physical models of CFP have been used to model processes on seemingly all scales from polymerization of molecules to star and galaxy formations in the universe (Aldous 1999). There has been increased interest in the past decade of lineage processes regarding coalescents. Although the central measures in CFP are the rates of each event, there does not seem to have been a study on CFP of clusters in a system exhibiting DP. Equilibrium and non-equilibrium studies have been investigated in the context of cluster formation and dissolution in the equilibrium case or coalescence overpowering fragmentation to cause a snowball effect of particles into a giant cluster in the non-equilibrium context (Aldous 1999).

Several studies on FPT problems have taken into account the initial starting distances between an object and its target (Bénichou & Voituriez 2014; Condamin, Bénichou, & Moreau 2007; Condamin et al. 2007). Future studies on this model could look for potential relationships between the starting points between n-lineages to determine if the initial configurations influence their coalescence time to a MRCA. The centroid step-size measures shown here, along with a measure of cluster gyration radius, could provide necessary insights for this problem.

Cluster measures and resulting interactions may be dependent upon the mass density functions of each cluster. The typical use of a mass density distribution is used to

describe the total mass, center of mass, moment of inertia, etc. In the case of clusters, the cluster-sizes are the zeroth order measure, the cluster mass (assuming each organism provides unit mass to the cluster mass). Cluster centroids represent the first order measure of the mass density distribution. Second order and higher measures may also be determined in the usual manner of mass density distributions. Further investigation is needed to determine how much information of the branching-coalescing process of clusters can be derived from cluster mass density distributions. However, the information may be limited since the mass density distribution is a static measure, so knowing the time evolution of such distributions, similar to how master equations describe time evolving probability distributions, is likely needed.

Comparisons between this model and biological data should be top priority. There have been studies on the fossil record where morphological traits are the best information about how ancient organisms lived and evolved. Querying databases specializing in phenotypes, such as PhenomicDB, or more generalized biological public data sets, such as Dryad, may provide an initial step toward decent checks for the predictive power of this model (Groth, Kalev, Kirov, Traikov, Leser, & Weiss 2010; Abe & Lieberman 2012). However, the model will likely need expansion in dimensionality of the phenotype space to meet the level of detail used by biologists. This should also resolve issues with low dimensional phenotype clustering that was addressed in Lawson and Jensen (2008). Furthermore, comparisons of the universal tree structures as determined by the ratios of average times to MRCA could provide the best comparisons.

Finally, it has been observed throughout this dissertation that populations of many BCRW organisms in a phenotype space have exhibited clustering and phase transition behavior dependent upon a noise amplitude measure of their maximum mutation size. This occurs for both assortative mating and asexual fission because of an asymmetry between birth and death processes. Random mating was shown to destroy both clustering and phase transition behavior. Furthermore, in the case of asexually reproducing organisms, the model was shown to belong to the DP universality class. The lineage behavior of organism lineage lifetimes at criticality appeared to have a scale-free distribution tail that scaled according to the correlation time exponent of DP. The wide range of cluster fitness indicates the possibility for multilevel selection as does the

average fitness of the clusters. The universal coalescent time ratios of cluster lineages appears to sweep from a Kingman's n-coalescent to some other coalescent not yet described by the mathematical literature. These features indicate that there are rich dynamics that are yet to be explored.

# APPENDIX A.
# MAXIMUM POPULATIONS

The maximum population size depends on the linear size of the landscape L and the competition limit κ. In the model, organisms are allowed to exist on the boundaries, and are only absorbed if they travel beyond the bounds. In a plane, the maximum number of sites on a hexagonal lattice (Steinhaus 1999). Along the one side of the landscape, the organisms can be placed in a row with spacing κ, and the maximum number of possible organisms (sites) along this row is W. Assuming this for the horizontal or vertical side does not matter, since the landscape is a square

$$W = floor\left(\frac{L}{\kappa}\right) + 1. \tag{A1}$$

The floor function rounds down to the nearest integer and one is added to account for the site in the corner of two boundaries. The next row must be shifted by κ/2 to fit the hexagonal lattice. This gives a row width w

$$w = W - 1. \tag{A2}$$

The spacing between rows can be calculated by finding the height of a right triangle with width κ/2 and hypotenuse κ. The spacing is then

$$dy = \kappa\frac{\sqrt{3}}{2}. \tag{A3}$$

The maximum number of rows is then

$$H = floor\left(\frac{L}{dy}\right) + 1. \tag{A4}$$

The addition of unity on the right hand side of (A4) accounts for the row along the first boundary. The maximum number of rows with width W

$$M = Floor\left(\frac{H}{2}\right), \quad \text{(A5)}$$

And the maximum number of rows with width w

$$m = ceil\left(\frac{H}{2}\right), \quad \text{(A5)}$$

where the ceil function rounds up to the nearest integer. The number of sites from rows with width W

$$S_W = WM, \quad \text{(A5)}$$

and number of sites from rows with width w

$$S_w = (W - 1)m. \quad \text{(A5)}$$

Therefore, the maximum number of sites in a landscape is

$$N_{max} = floor(S_W + S_w). \quad \text{(A5)}$$

The maximum population sizes are listed, for each landscape used throughout this dissertation, in Table A.1.

| Landscape | 21x21 | 29x29 | 37x37 | 45x45 | 77x77 |
|---|---|---|---|---|---|
| $N_{max}$ | 8196 | 15611 | 25393 | 37544 | 109826 |

Table A.1. Maximum Populations for Each Landscape.

# APPENDIX B.

# SCOTT ET AL. 2013

A Letters Journal Exploring the Frontiers of Physics

June 2013

EPL, **102** (2013) 68003 www.epljournal.org
doi: 10.1209/0295-5075/102/68003

# Clustering and phase transitions on a neutral landscape

Adam D. Scott[1], Dawn M. King[1], Nevena Marić[2] and Sonya Bahar[1(a)]

[1] *Department of Physics and Astronomy and Center for Neurodynamics, University of Missouri at St. Louis St. Louis MO, 63121,USA*
[2] *Department of Mathematics and Computer Science, University of Missouri at St. Louis - St. Louis MO, 63121, USA*

received 2 April 2013; accepted in final form 6 June 2013
published online 9 July 2013

PACS 87.23.Kg – Dynamics of evolution
PACS 87.15.Zg – Phase transitions
PACS 82.39.Rt – Reactions in complex biological systems

**Abstract** – Recent computational studies have shown that speciation can occur under neutral conditions, *i.e.*, when the simulated organisms all have identical fitness. These works bear comparison with mathematical studies of clustering on neutral landscapes in the context of branching and coalescing random walks. Here, we show that sympatric clustering/speciation can occur on a neutral landscape whose dimensions specify only the simulated organisms' phenotypes. We demonstrate that clustering occurs not only in the case of assortative mating, but also in the case of asexual fission; it is not observed in the control case of random mating. We find that the population size and the number of clusters undergo a second-order non-equilibrium phase transition as the maximum mutation size is varied.



**Introduction.** – The possibility of speciation driven by genetic drift alone, in the absence of selective drive, was first suggested in the context of Kimura's neutral theory [1–3] and has recently received increased attention due to the controversial work of Hubbell [4] on neutral theory in ecology. Many experimental studies have been undertaken in recent years to investigate both the empirical basis of Hubbell's neutral theory, and its applicability for the distribution of animal species (see, for example, Ricklefs [5]), rather than the plant species which initially inspired it. A variety of computational studies have been performed as well, with an eye to understanding the possible mechanisms underlying neutral clustering. One such recent study was that of de Aguiar *et al.* [6], which demonstrated clustering in a hybrid neutral spatial/genotype landscape. In this model, organisms reproduced by assortative mating (selecting mates with similar genotype and spatial location). The authors suggested that assortative mating was necessary but not sufficient for speciation (*i.e.*, for the formation of reproductively isolated clusters of organisms in the spatial/genotype landscape). Furthermore, "neither spatial nor genetic restrictions alone" were found to lead to speciation [6]. The model also showed species abundance curves consistent with the predictions of Hubbell's neutral model, and was thus interpreted as providing strong support for neutral theory [7].

The de Aguiar *et al.* model bears a strong similarity to a well-studied class of clustering models derived from the mathematics of branching and coalescing random walks. In 1990, Zhang *et al.* investigated a population of random walkers which move through a space, reproduce by fission, and die [8]. They observed clustering over time on a neutral fitness landscape. Derrida and Peliti [9] studied a similar problem, using a genome of binary digits similar to that introduced by Kauffman [10].

Meyer *et al.* [11] investigated a modified version of the Zhang *et al.* model, noting that clustering resulted from a *spatial asymmetry between birth and death processes*: new organisms always appear near their parent organisms, while death picks off individuals regardless of their location in the theoretical space. The importance of spatial asymmetry between birth and death dynamics was also emphasized by Young *et al.* [12], who introduced a "Brownian bugs" model for the clustering of asexually-reproducing organisms in a neutral landscape.

We recently investigated the role of maximum mutation size in driving speciation (modeled as clustering in phenotype space) in an agent-based evolutionary model designed to capture the key features of Darwinian natural selection —variability, competition, and heritability [13]. In this model, mutability (maximum mutation size) serves as a

(a) E-mail: bahars@umsl.edu

control parameter; increased mutability can be considered analogous to adding noise to the system. Simulating assortative mating on various types of rugged phenotype landscapes, we observed the formation of reproductively isolated clusters, analogous to species. As the mutability was increased, phase-transition–like behavior was observed in various measures such as population size and number of clusters. The transition took the system from an absorbing state of extinction to a state of survival, with highly contingent behavior occurring for mutability values within the transition range [13].

In the present paper we demonstrate that clustering and phase transitions occur in this model *even on a neutral fitness landscape*, for both assortative mating and bacteria-like reproduction by fission. Clustering fails to occur only in the control case of purely random mating. Importantly, the landscape in the model presented here is simply a space of phenotypes: it contains no representation of physical space, and thus provides a model for sympatric speciation.

**Methods.** –

*Model.* The model consists of a population of organisms in a continuous two-dimensional *phenotype space* (or *morphospace*) whose coordinates represent independent, arbitrary phenotypes with finite, absorbing boundaries. For each generation, a population of offspring organisms was generated from the current (parent) population according to one of three reproduction schemes: assortative mating, asexual reproduction, or random mating. After reproduction, the parent population was eliminated. The new population then underwent a series of death processes in order to simulate random mortality due to competition, predation, or other causes. The surviving offspring then became the new parents for the following generation.

In each simulation, the mating scheme, the size of the morphospace, the minimum distance allowed between organisms (competition limit), and the mutability $\mu$ were held constant. Mutability serves as the primary control parameter in the results below; it represents the maximum possible distance of an offspring from its parent(s), and thus defines the maximum allowable mutation size in the morphospace.

The fitness landscape in all the simulations presented here is "flat", in the sense that every organism, regardless of location, produces the same number of offspring, and each offspring is subjected to the same set of randomized death processes. Specifically, in each case studied here, the individual fitness is two (each organism produces two offspring). The conditions thus correspond to neutral selection.

Simulations began with an initial population of 300 organisms uniformly randomly dispersed throughout the landscape, and were run for up to 2000 generations. (Preliminary results suggest that different initial population sizes have an effect only on the transient behavior of the model, but not on the fundamental characteristics of the phase transitions described in the following section.) Runs were terminated early if the population became "extinct", dropping below three organisms (the minimum needed for the definition of a cluster; see below). Simulations were performed using an original program written in MATLAB (The MathWorks), and run on PCs using a Windows 7 operating system.

*Reproduction schemes.* Assortative mating was implemented by the selection of nearest neighbors. During each generation, each organism selected as its mate the organism located at the shortest distance from it in the morphospace. A base area for offspring production was determined by the positions of the reference organism and its mate. The area was then increased on all sides by the mutability $\mu$. Two offspring were generated at uniformly distributed random coordinates within the defined area. More explicitly, an offspring's coordinates $(b_x, b_y)$ were determined by

$$b_x = m_x - \mu + \{(M_x + \mu) - (m_x - \mu)\} r_x, \qquad (1a)$$

$$b_y = m_y - \mu + \{(M_y + \mu) - (m_y - \mu)\} r_y, \qquad (1b)$$

where $r_x$ and $r_y$ are random numbers uniformly distributed between 0 and 1, $M_x$ ($M_y$) is the maximum of the $x$ ($y$) coordinates of the parent organisms, and $m_x$ ($m_y$) is the minimum of the $x$ ($y$) coordinates of the parents. During each generation, each organism generated two offspring with its nearest neighbor.

In the asexual reproduction model, the area in which new offspring were generated was defined only by the parent's location, extended on all sides by $\mu$: thus, a square of size $2\mu \times 2\mu$, with the parent organism at the center. As a "null condition", another set of simulations were performed identically to the assortative mating scheme, but with organisms mating with randomly selected partners.

*Competition, random death, and boundary conditions.* In order to simulate the death of offspring due to resource competition, unviable hybridization, predation, or other vagaries of fate, organisms were eliminated due to: 1) competitive constraints, 2) random death, and 3) boundary conditions. Competition death occurred if two organisms were found closer than a spacing of 0.25 units from each other in the morphospace; in this case, one of the two was randomly chosen for elimination. The number of organisms eliminated by random death in each generation was determined by choosing a percentage from a uniform random distribution, with a maximum of 70%, of the population. Finally, offspring which were found beyond the bounds of the morphospace were eliminated. Application of these removal methods were executed in the order described.

*Clustering.* Following the elimination of organisms due to competition, random death and boundary conditions, the remaining organisms were assigned to clusters, which we consider an analog of species, as

follows. A cluster "seed" was determined by a reference organism, its mate (nearest phenotypic neighbor), and its second nearest phenotypic neighbor (which may be considered as a plausible alternate mate in an assortative mating scheme). Each organism defined its own seed by acting as the reference organism. An iterative process was used to determine whether organisms within a cluster seed were contained in other cluster seeds. This led to the formation of groups of organisms composed of connected cluster seeds. Once a closed group of organisms was determined, those organisms were defined as a cluster; these clusters not only represented a reproductively isolated group in accordance with the "biological species" concept, but could be interpreted more generally as phenetic (phenotypically similar) species [14]. For the asexual reproduction model, clusters were determined in the same manner as with assortative mating. However, of course, the organisms in this version of the model do *not* mate. This process again found closed sets of related cluster seeds which contain a reference organism's nearest and second nearest neighbors. The process of identifying a closed set of organisms from cluster seeds thus identifies phenetic species, in a manner roughly analogous to the classification of bacteria into species (a notoriously difficult task in real biology).

Clustering in the random mating model was determined by first identifying cluster seeds, as in the assortative mating model, where a seed was composed of a reference organism along with its mate and alternate. The mate was randomly chosen as described above, as was the alternate, to ensure that the formation of reproductively isolated clusters was independent of the phenotypic constraint of assortative mating. Aside from this difference, the process of determining a closed group of reproductively isolated organisms from cluster seeds was performed identically to the assortative mating case.

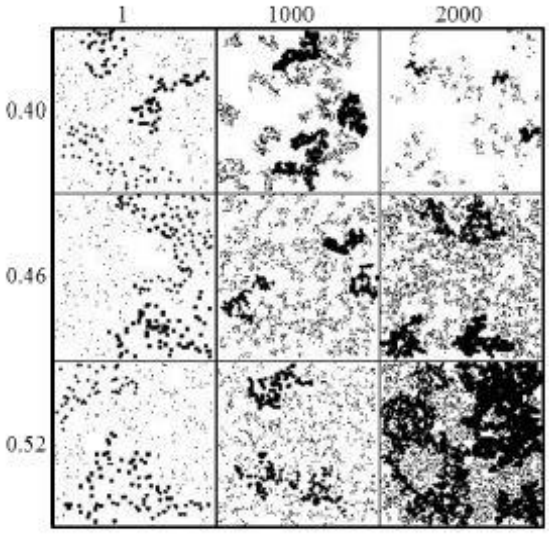


Fig. 1: Clustering in the assortative mating model: results are shown for different values of $\mu$ (from top to bottom: 0.40, 0.46, and 0.52) at different generations (from left to right: 1, 1000, and 2000) on a $45 \times 45$ landscape. The three largest clusters are emphasized in each panel by shapes, against the rest of the population (dots).

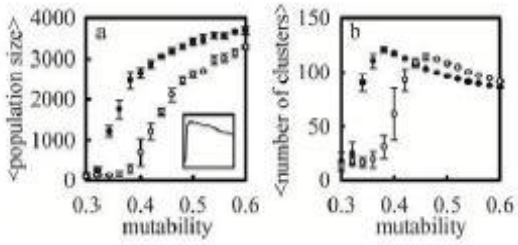


Fig. 2: Population size and number of clusters: (a) mean population and (b) mean number of clusters, as a function of $\mu$ for assortative mating (open circles) and asexual reproduction (filled circles), for a $45 \times 45$ landscape. Mean values are calculated over five different simulations (each run for 2000 generations, unless extinction occurred before this limit) at each value of $\mu$; error bars show standard deviation among these five runs. The inset in (a) shows the population over a larger range of $\mu$ for the assortative case; the horizontal axis shows $\mu$ from 0.3 to 8.0 (in steps of 0.02 from 0.3 to 1.5, and in steps of 1.0 from 2.0 to 8.0); the vertical axis shows the mean population on a scale from $-500$ to 4500).

**Results.** – Examples of typical simulation results are shown in fig. 1, illustrated by snapshots of the population at various generations for the assortative mating model. Figure 2(a) shows the mean population size, and fig. 2(b) shows the mean number of clusters, as a function of $\mu$ for assortative mating (filled circles) and asexual reproduction (open circles) for a $45 \times 45$ landscape. For small values of $\mu$, the populations quickly fall to extinction. Both mean population and number of clusters undergo a sharp rise for intermediate values of $\mu$; simulations on smaller ($21 \times 21$) and larger ($77 \times 77$) landscapes show that the steepness of this rise increases with system size. System size does not, however, affect the value of $\mu$ for which the number of clusters was maximal or the value of $\mu$ at which the populations survived up to 2000 generations in every trial. For much larger values of $\mu$, the mean population eventually decreases, due to a large number of organisms in each generation landing outside the boundary (fig. 2(a), inset). Note the similarity of this result to a stochastic resonance effect, in which an intermediate amount of noise optimizes a system's performance. These results are similar to those shown in Dees and Bahar [13], but, in sharp contrast to that work, we demonstrate clustering in this case *on an entirely flat (neutral) landscape*.

The transition from extinction to survival can also be visualized by constructing histograms of system lifetimes in simulations which are allowed to run until

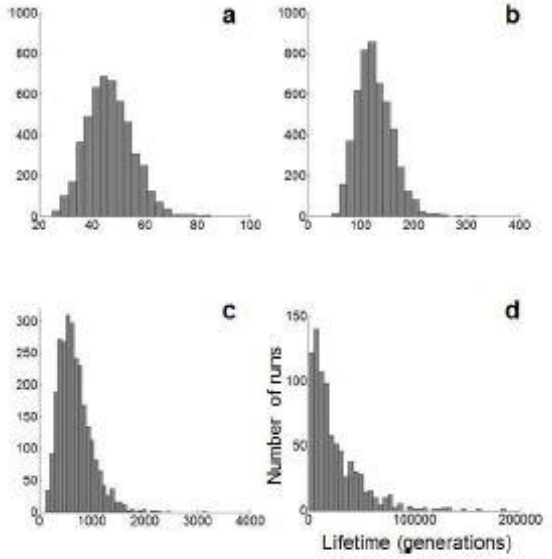


Fig. 3: Histograms of system lifetimes for the assortative mating model, on a $45 \times 45$ landscape: (a) $\mu = 0.20$, 5000 runs; (b) $\mu = 0.30$, 5000 runs; (c) $\mu = 0.37$, 2662 runs; (d) $\mu = 0.40$, 832 runs. Note the change in the scale of the horizontal axis as $\mu$ increases; axes (labeled in panel (d)) show the system lifetime (horizontal axis) and the number of runs with a given lifetime (vertical axis).

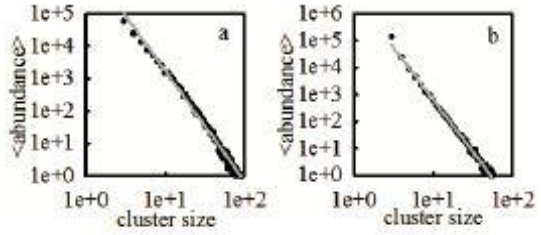


Fig. 4: Distributions of cluster sizes for the assortative mating model: (a) $\mu = 0.40$ and (b) $\mu = 0.42$, for the $45 \times 45$ landscape. Gray lines represent linear fits to the data.

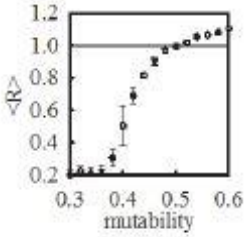


Fig. 5: Average nearest-neighbor index as a function of $\mu$: results are shown for assortative mating, for the $45 \times 45$ landscape. Values of $\langle R \rangle$ less than 1 indicate more aggregated clustering; $\langle R \rangle$ greater than 1 indicates more uniform clustering. Error bars show standard deviation over five runs.

extinction, rather than being cut off at 2000 generations. The histograms become increasingly "long-tailed" —and thus increasingly power-law–like— as $\mu$ moves into the transition range. This is illustrated in fig. 3 for the assortative case; similar results are obtained for the asexual reproduction scheme. At the rightmost edge of the critical range ($\mu = 0.42$ in the assortative case) the system survives indefinitely: runs were terminated after approximately 3 million generations, without showing any sign of approaching extinction.

Random mating, which acts as a control condition, dramatically changes the outcome of the simulation. First, the values of $\mu$ for which the population survives beyond extinction are much larger than those seen for the assortative and bacterial models. Moreover, the values of $\mu$ for which the population reaches a maximum in the random mating model exhibits a strong dependence on landscape size, and no phase transition is observed. Perhaps most significantly, the entire population almost always forms a single cluster.

In fig. 4, we show distributions of cluster sizes (defined as the number of organisms in a cluster) on a log-log scale for the assortative mating model, for values of $\mu$ over which the populations undergo a transition from extinction to survival (0.40–0.42). These plots show the average abundance of clusters of a given size (vertical axis) as a function of cluster size. Abundances were averaged for each cluster size from all generations over each of the five runs of the simulation. These plots can be considered as analogous to the well-known abundance curves (log-log Preston plots) of species distribution in Hubbell's neutral theory. For mutability less than 0.42, the plots are concave down, whereas at and above 0.42 they are concave up, indicating the existence of an exactly linear log-log plot (and thus of power-law behavior) for a value of mutability in this range. Similar results are obtained for the asexual mating simulations (data not shown).

Finally, fig. 5 demonstrates the quality of clustering using the nearest-neighbor index $R$ [15] for assortative mating. This index provides a measure of the relative proximity of nearest neighbors (mates, in the case of assortative mating) as compared to the proximity they would have in a random spatial distribution ($R = 1$). Since cluster assignments in the assortative mating model are defined in part by the assignment of nearest neighbors, this index provides a relevant measure of the quality of clustering in our model.

The horizontal line in fig. 5 represents the transition between aggregated distributions ($R < 1$) and more uniformly spaced distributions ($R > 1$). For assortative mating and asexual reproduction, the mutabilities for which $R$ is closest to 1 are also the values which optimize the number of clusters. Landscape size has minimal effect on the behavior of $R$, except for slight differences over the range of mutabilities which result in extinctions.

**Discussion.** – We have demonstrated clustering and an extinction-survival phase transition on a neutral landscape *in a pure phenotype space, with no representation of spatial separation*, for both assortative mating and asexual reproduction. In contrast, only a single giant cluster is formed, and no phase transition is observed, in a random mating case. While the results above present a striking contrast to those of de Aguiar *et al.* [6], who demonstrated clustering in a hybrid spatial-genotype space, and only for assortative mating, our findings are in fact quite consistent with a body of literature characterizing clustering due to asexual reproduction on neutral landscapes [8,9,11,12,16–18]. (Though note that actual fitness landscapes can be traversed even in the absence of any genetic changes, as, for example, in a recent study of stochastic gene expression in yeast [19].)

Derrida and Peliti [9] investigated the lineages of a fixed population, defined by genomes of spin-like alleles, subject to a fixed mutation rate, undergoing bacterial fission on a neutral fitness landscape. The model was extended by Serva and Peliti [20] to include random mating, and modified by Higgs and Derrida [21], to require mating only within a defined region of genome overlap; under these last conditions, the formation of species was observed. Although similar in its demonstration of species formation due to restrictions on mating behavior, the present model is not restricted to a specific population size; this allows a far richer dynamics of emergent speciation and "biodiversity".

Our results also bear a strong similarity to studies of branching and coalescing random walks, such as that studied in a discrete space by Athreya and Swart [22]. Their particles perform independent random walks, and undergo binary splitting, coalescence and death. Our model, although it is constructed on a continuum, shares the main features of this Markov process. It should be noted that the death of an organism due to the competition limit corresponds to the coalescence of two particles. Additionally, our morphospace has finite boundaries, so the process may be seen as a branching and coalescing random walk with fixed barriers. Deaths due to the boundary conditions increase the likelihood of extinction, so it could theoretically be expected that the critical value of $\mu$ in our finite landscape would be significantly different in an infinite space model (Marić, unpublished), and thus that the critical value of $\mu$ might change as our landscape size is increased. However, the critical range of $\mu$ ($\sim 0.40$) is likely sufficiently close to the competition limit (0.25) that this effect is below the level of detection in our simulations. In fact, the majority of the deaths in our simulations are due to local interactions between organisms and to the random death process; death at the boundaries constitutes only a small fraction of the total deaths for the range of $\mu$ studied here. The lack of system size effects on the critical value of $\mu$ may be interpreted in light of the arguments of Privman [23], and Toral and Tessone [24], who show that critical parameters are effectively independent of system size when the correlation length is much smaller than the system's dimensions. System size effects do become prominent in our model for much larger values of $\mu$, for which a much larger fraction of the organisms land at the boundary. These effects should also be observable in our system for much smaller system sizes, which would also increase the fraction of organisms landing at the boundary.

As discussed by Young *et al.*, the key ingredients for clustering lie in the spatial asymmetry of birth and death processes: birth occurs near parent organisms, while death is distributed across the landscape [12]. This asymmetry produces clustering both in phenotype space and in the physical (spatial) clustering of organisms. The role of spatial asymmetry in our model is illustrated vividly by random mating: with spatial asymmetry eliminated, the system forms a single giant cluster. Similar problems were investigated by Fuentes *et al.* [25] and Law *et al.* [26].

The work of Fuentes *et al.* [25] and that of Young *et al.* [12] highlight the importance of the balance between local and global interactions. This suggests that the competition limit, in addition to local mating and reproduction requirements, may play a key role in determining the system behavior in our case. If the competition limit goes to zero, it is possible that the critical mutability might also approach zero, though this would also depend on the percentage of random death. In such a case, the phase transition should vanish entirely, yielding a result similar to the extreme local limit described by Fuentes *et al.*, in which all population structure is lost [25].

Our observations of clustering on a neutral landscape are quite consistent with the various studies reviewed above. The most novel aspect of the present study is the demonstration of a non-equilibrium critical phase transition and optimization of the number of clusters on a neutral landscape as a mutation parameter, mutability, is varied. As with our previous study on a rugged (non-neutral) fitness landscape [16], a sharp rise in both the population size and the number of clusters is observed for intermediate values of $\mu$ (fig. 2); a similar transition is found in the nearest-neighbor index $R$ (fig. 5). In contrast, this transition is gradual, rather than steep, in the control case of random mating. The system lifetime in the assortative and bacterial cases also undergoes a sharp transition in the same range of $\mu$ values, as the system shifts from the "absorbing state" of extinction to a state of survival.

Phase transition-like behavior has been previously suggested in evolutionary models (for example, [27]) but not, to our knowledge, as a function of mutation-related parameters. Critical, scale-free behavior in the transition range is a hallmark of second-order (also called continuous or critical) phase transitions, and the linearity of the log-log abundance curves shown in the critical range of $\mu$ (fig. 4), as well as the power-law–like distribution of system lifetimes (fig. 3) provide evidence for such a transition in the model studied here. Linearity in log-log plots corresponds to power-law scaling, and in fig. 4 this

corresponds to distributions of clusters of all sizes, with no characteristic scale. The change in concavity of the plots in fig. 4 strongly supports the existence of a critical value of $\mu$ for which the distribution of cluster sizes is truly scale free. This suggests that the phase transition observed here is second order; it is also a non-equilibrium transition, given the presence of the absorbing state of extinction, from which the system cannot escape.

The nearest-neighbor measure, $R$, shown in fig. 5, characterizes the filling of the landscape by the organisms. When $R<1$, the morphospace is not well filled, and the organisms which exist in the space are clumped together. As $R$ increases toward unity, clusters fill more of the landscape, as seen in fig. 1, and the spacing between clusters decreases as populations rise. As the population fills the morphospace more uniformly, the clusters become less well defined. Biologically, this would correspond to phenotypically less well-separated species, though the identity of species in the model, as in real biology, could still be parsed, given sufficient knowledge of their genealogical/genetic history. Note that for large values of $R$, when clusters are virtually impossible to separate by eye, the clustering algorithm used here can still separate the population into distinct clusters (see, for example, the bottom right panel in fig. 1). The peak in the number of clusters (fig. 3(b)) is observed for values of $\mu$ for which $R$ undergoes a transition through a value of 1 (random distribution) to $R>1$. This, in effect, means that the spacing between organisms in the landscape undergoes a transition from clumpy to uniformly spaced, and that the transition to uniformity is intimately linked to the decline in the number of clusters. This model may be extended —and made more biologically realistic— by including more dimensions to the phenotype space, thus characterizing each organism by a larger number of traits. However, it is a well-documented result of renormalization group methods that critical behavior is tied to the dimensionality of a system, which may result in different critical behavior and thus in modified clustering behavior [28]. The relation between the biodiversity produced in such higher-dimensional models, and actual biological data, will be a key target of future investigations.

* * *

The authors thank Dr NATHAN D. DEES for his encouragement, and for his work in developing the original version of this model. The authors gratefully acknowledge support from NSF grant DMS-1007823 (to NM), from a James S. McDonnell Foundation "Studying Complex Systems" grant (to SB), and from University of Missouri Research Board Grants (to both NM and SB).

REFERENCES

[1] KIMURA M., *Nature*, **217** (1968) 624.
[2] KIMURA M., *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge) 1983.
[3] KIMURA M. and CROW J. F., *Genetics*, **49** (1964) 6191.
[4] HUBBELL S. P., *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, Princeton, NJ) 2001.
[5] RICKLEFS R. E., *Ecology*, **87** (2006) 1424.
[6] DE AGUIAR M. A. M., BARANGER M., BAPTESTINI M., KAUFMAN L. and BAR-YAM Y., *Nature*, **460** (2009) 384.
[7] BANAVAR J. R. and MARITAN A., *Nature*, **460** (2009) 334.
[8] ZHANG Y.-C., SERVA M. and POLIKARPOV M., *J. Stat. Phys.*, **58** (1990) 849.
[9] DERRIDA B. and PELITI L., *Bull. Math. Biol.*, **53** (1991) 355.
[10] KAUFFMAN S., *The Origins of Order: Self-organization and Selection in Evolution* (Oxford University Press, Oxford) 1993.
[11] MEYER M., HAVLIN S. and BUNDE A., *Phys. Rev. E*, **54** (1996) 5567.
[12] YOUNG W. R., ROBERTS A. J. and STUHNE G., *Nature*, **412** (2001) 328.
[13] DEES N. D. and BAHAR S., *PLOS ONE*, **5** (2010) e11952.
[14] RIDLEY M., *Evolution*, 3rd edition (Blackwell Science, Ltd., Malden, Mass.) 2004.
[15] CLARK P. J. and EVANS F. C., *Ecology*, **35** (1954) 445.
[16] HOUCHMANDZADEH B., *Phys. Rev. E*, **66** (2002) 052902.
[17] HOUCHMANDZADEH B. and VALLADE M., *Phys. Rev. Lett.*, **68** (2003) 061912.
[18] LAWSON D. J. and JENSEN H. J., *Phys. Rev. Lett.*, **98** (2002) 098102.
[19] NEVOZHAY D., ADAMS R. M., VAN ITALLIE E., BENNETT M. R. and BALÁZSI G., *PLOS Comput. Biol.*, **8** (2012) e1002480.
[20] SERVA M. and PELITI L., *J. Phys. A: Math. Gen.*, **24** (1991) L705.
[21] HIGGS P. G. and DERRIDA B., *J. Mol. Evol.*, **35** (1992) 454.
[22] ATHREYA S. R. and SWART J. M., *Probab. Theory Relat. Fields*, **131** (2005) 376.
[23] PRIVMAN V., in *Finite Size Scaling and Numerical Simulation of Statistical Systems*, edited by PRIVMAN V. (World Scientific, Singapore) 1990.
[24] TORAL R. and TESSONE C. J., *Commun. Comput. Phys.*, **2** (2007) 177.
[25] FUENTES M. A., KUPERMAN M. N. and KENKRE V. M., *Phys. Rev. Lett.*, **91** (2003) 158104.
[26] LAW R., MURRELL D. J. and DIECKMANN U., *Ecology*, **84** (2003) 252.
[27] LUZ-BURGOA K., MOSS DE OLIVEIRA S., SCHWÄMMLE V. and SÁ MARTINS J. S., *Phys. Rev. E*, **74** (2006) 021910.
[28] HINRICHSEN H., *Adv. Phys.*, **49** (2000) 815.

# BIBLIOGRAPHY


Abe, F. R., Liebermann, B. S., 2012. Quantifying morphological change during an evolutionary radiation of Devonian trilobites. *Paleobiology* **38**:292.

Aldous, D. J., 1999. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli* **5**(1):3.

Angel, Berestycki, Limic, 2012. Global divergence of spatial coalescents. *Probability Theory and Related Fields* **152**:625.

Athreya, S. R., Swart, J. M., 2005. Branching-coalescing particle systems. *Probability Theory Related Fields* **131**:376.

Bacher, A., 2013. Directed and multi-directed animals on the King's lattice. In *The Seventh European Conference on Combinatorics, Graph Theory and Applications* **16**:535. Scuola Normale Superiore.

Banavar, J. R., Maritan, A., 2009. Towards a neutral theory of biodiversity. *Nature* **460**:334.

Barato, A. C., Bonachela, J. A., Fiore, C. E., Hinrichsen, H., Muñoz, M. A., 2009. Simplest nonequilibrium phase transition into an absorbing state. *Physical Review E* **79**:041130.

Bénichou, O., Chevalier, C., Klafter, J., Meyer, B., Voituriez, R., 2010. Geometry-controlled kinetics. *Nature Chemistry* **2**:472.

Bénichou, O., Voituriez, R., 2014. From first-passage times of random walks in confinement to geometry-controlled kinetics. *Physics Reports* in press.

Berestycki, J., 2003. Multifractal spectra of fragmentation processes. *Journal of Statistical Physics* **113**(314):411.

Berestycki, J., 2004. Exchangeable fragmentation-coalescence processes and their equilibrium measures. *arXiv preprint math/0403154*.

Berestycki, N., 2009. Recent progress in coalescent theory. *Ensaios Matematicos*. **16**:1.

Bernal, J. D., Oparin, A. I., Mueller, G., Haldane, J. B. S., Synge, A., 1967. *The origin of life*. Weidenfeld and Nicolson, London.

Bertoin, J., 2006. *Random fragmentation and coagulation processes*. Vol. 102. Cambridge University Press, New York.

Bertoin, J., Goldschmidt, C., 2004. Dual random fragmentation and coagulation and an application to the genealogy of Yule processes. In *Mathematics and Computer Science III*:295. Birkhäuser Basel.

Bhattacharjee, S. M., Seno, F., 2001. A measure of data collapse for scaling. *Journal of Physics A: Mathematical and General* **34**:6375.

Bousquet-Mélou, M., 1996. A method for the enumeration of various classes of column-convex polygons. *Discrete Mathematics* **154**(1):1.

Brunet, É. Derrida, B., 2009. Statistics at the tip of a branching random walk and the delay of traveling waves. *Europhysics Letters* **87**:60010.

Brunet, É. Derrida, B., 2013. Genealogies in simple models of evolution. *Journal of Statistical Mechanics: Theory and Experiment* **2013**(01):P01006.

Brunet, É., Derrida, B. (2012). How genealogies are affected by the speed of evolution. *Philosophical Magazine* **92**:255.

Brunet, É. Derrida, B., Simon, D., 2008. Universal tree structures in directed polymers and models of evolving populations. *Physical Review E* **78**:061102.

Cadillo-Quiroz, H., Didelot, X., Held, N. L., Herrera, A., Darling, A., Reno, M. L., Krause, D. J., Whitaker, R. J., 2012. Patterns of gene flow define species of thermophilic archaea. *PLoS Biology* **10**(2):e1001265.

Chevalier, C., Bénichou, O., Meyer, B., Voituriez, R., 2011. First-passage quantities of Brownian motion in a bounded domain with multiple targets: a unified approach. *Journal of Physics A: Mathematical and Theoretical*, **44**(2):025002.

Clark, P. J., Evans, F. C., 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* **35**(4):445.

Condamin, S., Bénichou, O., Moreau, M., 2007. Random walks and Brownian motion: A method of computation for first-passage times and related quantities in confined geometries. *Physical Review E* **75**:021111.

Condamin, S., Bénichou, O., Tejedor, V., Voituriez, R., Klafter, J., 2007. First-passage times in complex scale-invariant media. *Nature* **450**:77.

Damore, J. A., Gore, J., 2012. Understanding microbial cooperation. *Journal of Theoretical Biology* **299**:31.

Damuth, J., Heisler, I. L., 1988. Alternative formulations of multilevel selection. *Biology and Philosophy* **3**(4):407.

Darwin, C., 1859. On the origins of species by means of natural selection. London: *Murray*.

De Aguiar, M. A. M., Baranger, M., Baptestini, E. M., Kaufman, L., Bar-Yam, Y., 2009. Global patterns of speciation and diversity. *Nature* **460**:384.

De Cara, M. A. R., Barton, N. H., Kirkpatrick, M., 2008. A model for the evolution of assortative mating. *The American Naturalist* **171**(5):580.

Dees, N. D., Bahar, S., 2010. Noise-optimized speciation in an evolutionary model. *PLoS ONE* **5**(8):e11952.

Derrida, B., Peliti, L., 1991. Evolution in a flat fitness landscape. *Bulletin Mathematical Biology* **53**(3):355.

Dieckmann, U., Doebeli, M.,1999. On the origin of species by sympatric speciation. *Nature* **400**(6742):354.

Diestel, R., 2012. *Graph Theory*. Springer, Heidelberg. From the series *Graduate Texts in Mathematics* **173**. (ELECTRONIC COPY ONLINE)

Dobramysl, U., Täuber, U. C., 2008. Spatial variability enhances species fitness in stochastic predator-prey interactions. *Physical Review Letters* **101**(25):258102.

Durrett, R., Limic, V., 2002. A surprising Poisson process arising from a species competition model. *Stochastic processes and their applications* **102**(2):301.

Dutta, C., Pandurangan, G., Rajaraman, R., Roche, S., 2013. Coalescing-branching random walks on graphs. In *Proceedings of the 25th ACM symposium on Parallelism in algorithms and architectures* 176. ACM.

Ehsasi, M., Matloch, M., Frank, O., Block, J. H., Christmann, K., Rys, F. S., Hirschwald, W., 1989. Steady and nonsteady rates of reaction in a heterogeneously catalyzed reaction: Oxidation of CO on platinum, experiments and simulations. *The Journal of Chemical Physics* **91**(8):4949.

Feder, J., 1988. *Fractals*. Plenum Press, New York.

Ferreira, C. P., Fontanari, J. F., 2002. Nonequilibrium phase transitions in a model for the origin of life. *Physical Review E*, ***65***(2):021902.

Fisher, M. E., 1998. Renormalization group theory: Its basis and formulation in statistical physics. *Reviews of Modern Physics* **70**(2):653.

Fisher, R. A., 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Foote, M., 1990. Nearest-neighbor analysis of trilobite morphospace. *Systematic Biology* **39**(4):371.

Fröjdh, P., Howard, M., Lauritsen, K. B., 2001. Directed percolation and other systems with absorbing states: Impact of boundaries. *International Journal of Modern Physics B* **15**(12):1761.

Fuentes, M. A., Kuperman, M. N., Kenkre, V. M., 2003. Nonlocal interaction effects on pattern formation in population dynamics. *Physical Review Letters* **91**(15):158104.

Garcia, A. L., 2000. *Numerical Methods for Physics* (2nd ed.). Prentice Hall Inc., Upper Saddle River, New Jersey.

Goodnight, C., 2013. On Multilevel Selection and Kin Selection: Contextual Analysis Meets Direct Fitness. *Evolution* **67**(6):1539.

Gould, S. J., 1990. *Wonderful Life: The Burgess Shale and the Nature of History*. W. W. Norton and Company, New York.

Grimmett, G., 2008. Space-time percolation. *Progress in Probability* **60**:305.

Groth, P., Kalev, I., Kirov, I., Traikov, B., Leser, U., Weiss, B., 2010. Phenoclustering: online mining of cross-species phenotypes. *Bioinformatics* **26**(15):1924.

Hamilton, W. D., 1964a. The genetical evolution of social behavior. I. *Journal of Theoretical Biology* **7**(1):1.

Hamilton, W. D., 1964b. The genetical behavior of social evolution. II., *Journal of Theoretical Biology* **7**(1):17.

Henkel, M., Hinrichsen, H., Lübeck, S., 2009. *Non-Equilibrium Phase Transitions: Volume 1: Absorbing Phase Transitions*. Springer, Dordrecht.

Henkel, M., Schütz, G., 1994. Boundary-induced phase transitions in equilibrium and non-equilibrium systems. *Physica A: Statistical Mechanics and its Applications* **206**(1):187.

Higgs, P. G., Derrida, B., 1992. Genetic distance and species formation in evolving populations. *Journal of Molecular Evolution* **35**(5):454.

Hinrichsen, H., 2000. Non-equilibrium critical phenomena and phase transitions into absorbing states. *Advances in Physics* 49(7):815.

Hinrichsen, H., 2003. Stochastic cellular automaton for the coagulation-fission process. *Physica A: Statistical Mechanics and its Applications* **320**:249.

Houchmandzadeh, B., 2002. Clustering of diffusing organisms. *Physical Review E* **66**:052902.

Houchmandzadeh, B., Vallade, M., 2003. Cluster in neutral ecology. *Physical Review Letters* **68**:061912.

Hubbell, S. P., 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton.

Hubbell, S. P., 2005. Neutral theory in community ecology and the hypothesis of function equivalence. *Functional Ecology* **19**:166.

Jensen, I., 1996. Temporally disordered bond percolation on the directed square lattice. *Physical Review Letters* **77**(25):4988.

Johannesson, K., Rolan-Alvarez, E., Ekendahl, A., 1995. Incipient reproductive isolation between two sympatric morphs of the intertidal snail Littorina saxatilis. *Evolution* 49(6):1180.

Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature* **217**:624.

Kimura, M., 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

Kimura, M., Crow, J. F., 1964, The number of alleles that can be maintained in a finite population. *Genetics* **49**:6191.

Kingman, J. F. C., 1982. The coalescent. *Stochastic Processes and Their Applications* **13**(3):235.

Knežević, M., Vannimenus, J., 2002. On directed interacting animals and directed percolation. *Journal of Physics A: Mathematical and General* **35**(12):2725.

Kondrashov, A. S., Shpak, M., 1998. On the origin of species by means of assortative mating. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**(1412):2273.

Kuhr, J. T., Leisner, M., Frey, E., 2011. Range expansion with mutation and selection: dynamical phase transition in a two-species Eden model. *New Journal of Physics* **13**:113013.

Lavrentovich, M. O., Korolev, K. S., Nelson, D. R., 2013. Radial Domany-Kinzel models with mutation and selection. *Physical Review E* **87**:012103.

Lawson, D. J., Jensen, H. J., 2002. Neutral evolution in a biological population as diffusion in phenotype space: reproduction with local mutation but without selection. *Physical Review Letters* **98**:098102.

Lawson, D. J., Jensen, H. J., 2008. Understanding clustering in type space using field theoretic techniques. *Bulletin of Mathematical Biology* **70**(4):1065.

Lesne, A., 1998. *Renormalization Methods: Critical Phenomena, Chaos, Fractal Structures*. John Wiley & Sons, Inc. New York.

Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. *Nature* **475**:493.

Limic, V., Strum, A., 2006. The spatial Λ-coalescent. *Electronic Journal of Probability* **11**(15):363.

Lipowski, A., Ferreira, A. L., Wendykier, J., 2012. Critical behavior of a tumor growth model: directed percolation with a mean-field flavor. *Physical Review E* **86**:041138.

Lipowski, A., Lopata, M., 1999. Model of biological evolution with threshold dynamics and infinitely many absorbing states. *Physical Review E* **60**:1516.

Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K., Rodriguez-Schettino, L., 1998. Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, **279**(5359):2115.

Lübeck, S., 2004. Universal scaling behavior of non-equilibrium phase transitions. *International Journal of Modern Physics B* **18**:3977.

Luz-Burgoa, K., Moss de Oliveira, S., Schwämmle, V., Sá Martins, J. S., 2006. Thermodynamic behavior of a phase transition in a model for sympatric speciation. *Physical Review E* **74**:021910.

Marckert, J. F., 2012. Directed animals, quadratic systems and rewriting systems. *The Electronic Journal of Combinatorics* **19**(3):P45.

MATLAB, Parallel Computing Toolbox, and Distributed Computing Server Release 2013a. The Mathworks, Inc. Natick, Massachusetts, United States.

Maynard Smith, J., Szathmáry, E., 1995. *The Major Transitions in Evolution*. Oxford University Press, New York.

Meyer, M., Havlin, S., Bunde, A., 1996. Clustering of independently diffusing individuals by birth and death processes. *Physical Review E* **54**(5):5567.

Möhle, M., Sagitov, S., 2001. A classification of coalescent processes for haploid exchangeable population models. *The Annals of Probability* **29**(4):1547.

Moon, F. C., 1992. *Chaotic and Fractal Dynamics: An Introduction for Applied Scientists and Engineers*. John Wiley & Sons, Inc., New York.

Norton, A., Tandy, B., 1999. Cantor sets, binary trees and Lipschitz circle homeomorphisms. *Michigan Mathematics Journal* **46**:29.

Ódor, G., 2004. Universality classes in nonequilibrium lattice systems. *Reviews of Modern Physics* **76**:664.

Oborny, B., Meszéna, G., Szabó, G., 2005. Dynamics of populations on the verge of extinction. *Oikos* **109**:291.

Okasha, S., 2005. Maynard Smith on the levels of selection question. *Biology and Philosophy* **20**(5):989.

Okasha, S., 2006. The levels of selection debate: philosophical issues. *Philosophy Compass* **1**(1):74.

Okasha, S., 2009. *Evolution and the Levels of Selection*. Oxford University Press, New York.

Otto, S. P., Servedio, M. R., Nuismer, S. L., 2008. Frequency-dependent selection and the evolution of assortative mating. *Genetics* **179**(4):2091.

Park, S. C., 2011. Critical study of the absorbing phase transition in a four-state predator–prey model in one dimension. *Journal of Statistical Mechanics: Theory and Experiment* **2011**(9):L09001.

Park, S. C., Park, H., 2008. Nonequilibrium phase transitions into absorbing states. *The European Physical Journal B* **64**:415.

Pie, M. R., Weitz, J. S., 2005. A null model of morphospace occupation. *The American Naturalist* **166**(1):E1.

Pitman, J., 1999. Coalescents with multiple collisions. *The Annals of Probability* **27**(4):1870.

Privman, V., 1990. Finite-size scaling theory. In *Finite Size Scaling and Numerical Simulation of Statistical Systems*, V. Privman, Ed. Singapore. World Scientific, Singapore.

Reinhardt, H., Böhm, F., Drossel, B., Hinrichsen, H., 2006. Nonequilibrium critical behavior of a species coexistence model. *The European Physical Journal B - Condensed Matter and Complex Systems* **51**(2):245.

Ricklefs, R. E., 2006. The unified neutral theory of biodiversity: do the numbers add up? *Ecology* **87**(6):1424.

Ridley, M., 2004. *Evolution* (3rd Ed.). Blackwell Science, Ltd., Malden, Massachusetts.

Schluter, D., 1994. Experimental evidence that competition promotes divergence in adaptive radiation. *Science* **266**(5186):798.

Schweinsberg, J. R., 2000. Coalescents with simultaneous multiple collisions. *Electronic Journal of Probability* **5**(12):1.

Scott, A. D., King, D., Marić, N., Bahar, S., 2013. Clustering and phase transitions on a neutral landscape. *Europhysics Letters* **102**:68003.

Serva, M., Peliti, L., 1991. A statistical model of an evolving population with sexual reproduction. *Journal of Physics A: Mathematical and General* **24**(13):L705.

Sipos, M., Goldenfeld, N., 2011. Directed percolation describes lifetime and growth of turbulent puffs and slugs. *Physical Review E* **84**(3):035304.

Solé, R. V., 2011. *Phase Transitions*. Princeton University Press, Princeton.

Stauffer, D., Aharony, A., 1992. *Introduction to Percolation Theory* (2nd Ed.). Taylor & Francis, Ltd., Bristol, Pennsylvania.

Steinhaus, H., 1999. *Mathematical Snapshots*. Dover Publications, Mineola.

Takeuchi, K. A., Kuroda, M., Chaté, H., Sano, M., 2007. Directed percolation criticality in turbulent liquid crystals. *Physical Review Letters* **99**(23):234503.

Takeuchi, K. A., Kuroda, M., Chaté, H., Sano, M., 2009. Experimental realization of directed percolation criticality in turbulent liquid crystals. *Physical Review E* **80**(5):051116.

Täuber, U. C., Howard, M., Vollmayr-Lee, B. P., 2005. Applications of field-theoretic renormalization group methods to reaction-diffusion problems. *Journal of Physics A: Mathematical and General* **38**(17):R79.

Traulsen, A., Nowak, M. A., 2006. Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences* **103**(29):10952.

Teh, Y. W., Blundell, C., Elliott, L. T., 2011. Modelling genetic variations with fragmentation-coagulation processes. *Advances in Neural Information Processing Systems*.

Tellier, A., Lemaire, C., 2014. Coalescence 2.0: a multiple branching of recent theoretical developments and their applications. *arXiv preprint arXiv:1401.5248*.

Timofeeva, Y., Coombes, S., 2004. Directed percolation in a two-dimensional stochastic fire-diffuse-fire model. *Physical Review E* **70**(6):062901.

Toral, R., Tessone, C. J., 2007. Finite size effects in the dynamics of opinion. *Communications in Computational Physics* **2**(2):177.

Tran, T. D., Hofrichter, J., Jost, J., 2013. An introduction to the mathematical structure of the Wright-Fisher model of population genetics. *Theory in Biosciences* **132**(2):73.

Traulsen, A., Nowak, M. A., 2006. Evolution of cooperation by multilevel selection. *PNAS* **103**(29):10952.

Tuckwell, H. C., Wan, F. Y., 1984. First-passage time of Markov process to moving barriers. *Journal of Applied Probability* **21**(4):695.

von Kiedrowski, G., 1996. Primordial soup or crêpes? *Nature* **381**:20.

Watson, H. W., Galton, F., 1875. On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland* **4**:138.

Young, W. R., Roberts, A. J., Stuhne, G., 2001. Reproductive pair correlations and the clustering of organisms. *Nature* **412**:328.

Zähle, I., Cox, J. T., Durrett, R., 2005. The stepping stone model. II: genealogies and the infinite sites model. *The Annals of Applied Probability* **15**(1B):671.

Zambelli, T., Wintterlin, J., Trost, J., Ertl, G., 1996. Identification of the "active sites" of a surface-catalyzed reaction. *Science* **273**(5282):1688.

Zhang, Y. -C., Serva, M., Polikarpov, M., 1990. Diffusion reproduction processes. *Journal of Statistical Physics* **58**:489.

# VITA

Adam David Scott was born in Osage Beach, MO to amazing, loving parents, Joe and Anne Scott. In nearby Camdenton, MO, Adam attended Camdenton High School where he received his high school diploma in 2003. For his undergraduate studies, he went to Drury University in Springfield, MO and in May 2007, received Bachelor of Arts degrees in both Physics and in Mathematics while minoring in Global Studies. After his first year of graduate studies in physics at the University of Arkansas – Fayetteville, Adam transferred to the Physics and Astronomy Department and the Center for Neurodynamics at the University of Missouri at St. Louis in June 2008. In May 2009 he received his Master of Science degree in Physics. He then entered officially into the cooperative doctoral program for physics at the University of Missouri of St. Louis and the Missouri University of Science and Technology in January 2010. Adam married Staci Erin Smith, a doctoral student in the Neuroscience of Behavior and Cognition program of the Psychology Department at the University of Missouri at St. Louis, in June 2013. He then completed the degree of Doctor of Philosophy in Physics in May 2014.