

## Feasibility Study

### *Speciation Dynamics of an Agent-based Evolutionary Model in Phenotype Space*

Adam David Scott

11 June 2012

#### Part 1:

Clustering and Phase Transitions on a Neutral Landscape

#### Part 2:

Development of a Simple Mean-Field Equation to Predict Criticality as a Function of Coalescence Radii for Asexual Fission on a Neutral Landscape

#### Part 3:

Scaling Behavior & Fractal Structures of Lineages and Clusters on a Neutral Landscape

Despite consensus that evolution drives the diversity of life, the mechanisms which control it are still debated [1]. Computational evolution models have provided great insights into the minimal requirements necessary for evolution in a variety of situations [2–4]. Many of these models use a mutation parameter to optimize, or simply determine, populations, species, diversity, etc. However, to my knowledge, no current models show that a mutation parameter causes phase-transition behavior. Dynamically, a model with this behavior may help guide predictions in evolutionary theory such as the possibility of multiple levels of selection and the rarity of sympatric speciation in nature [5–7].

My research proposal involves extending an original model developed by Dees and Bahar, in which the mutation parameter has been shown to tune population size and number of clusters (interpreted as species)[8]. The model interactions are simple, and any large scale behavior is a result of the local agent-based interactions. This lends the model to analysis using techniques from statistical physics and non-linear dynamics, which will provide the major tools I will use throughout. Due to the model's simplicity, major evolutionary events such as speciation occur in such a way that they can be considered sympatric, meaning that physical space is not a factor for speciation. Sympatric speciation is accepted as theoretically possible, but has been rarely observed in nature and therefore it is not recognized as a highly prevalent process [9–11].

Mathematically, the evolutionary model considered here constitutes a branching and coalescing random walk. Furthermore, given that there are many organisms which may be considered random walkers, our model may be considered to undergo super-Brownian motion [12].

I plan to divide my dissertation into three chapters. The first chapter will focus on simulation results demonstrating clustering and phase transition behavior without natural selection. I intend to

develop a simple mean-field approximation for the case of asexual fission without natural selection for chapter two. Finally, in the last chapter I will measure potential fractal properties of lineages and clusters which will address the evolutionary problem of group selection.

The first section explores the special conditions of our model in which natural selection is removed and allows organisms to evolve by random phenotypic drift. I have shown that the mutation parameter, mutability, acts as a control parameter by tuning the population from the absorbing state of extinction to survival. Over the same range of mutability values, the number of species (clusters) also increases. However, a peak is reached which, as I showed using the Clark and Evans nearest neighbor index, corresponds to the cluster quality changing from aggregated to uniform spacing of the organisms[13]. Furthermore, I showed that the cluster sizes follow power law behavior for mutabilities nearest to criticality. This suggests scale-free fluctuations in cluster sizes and is characteristic of a continuous phase transition. The conclusions from this work lay the ground work for the remaining chapters. This work has been submitted for publication in the *Journal of Theoretical Biology*.

In chapter two I will attempt to describe our model with a mean-field description in the case of asexual fission and neutral fitness. The results from chapter one suggest that the most likely phase transition of our model is of the directed percolation universality class. Spatially, there is no directedness for the diffusive movement of the organisms on a neutral landscape, so the populations may be considered isotropic percolation. The transition from extinction to survival suggests that the population percolates through time by surviving above critical mutability.

A simple model of directed percolation is typically described by a 1+1 dimensional model in which particles undergo a reaction-diffusion process. Active sites arise if neighboring sites are available in the next time step. The processes are: **Death**) If no neighboring sites (determined by the percolation probability parameter) are available in the next time step, then the particle's percolation terminates,  $A \rightarrow 0$ . **Diffusion**) If one neighboring site is available, then the particle percolates into the available site. **Production**) If two neighboring sites are available, then the particle generates an offspring to fill both sites,  $A \rightarrow 2A$ . **Coalescence**) If two particles percolate to the same site, then only one particle survives in a coalescence process,  $2A \rightarrow A$  [14].

Our model can be cast in a corresponding way. The death process,  $A \rightarrow 0$ , may be considered as random death in our model. Generation of offspring,  $A \rightarrow 2A$ , may be considered as asexual fission, assortative mating, or random mating. For coalescence  $2A \rightarrow A$ , this can be represented by the competition among organisms encroaching on some local phenotype space. Diffusion in our model is effectively achieved by coalescence of sibling offspring. With these analogies, our model may be a continuous space, discrete time (2+1 dimensions) variation of the more basic 1+1 lattice model described earlier. The organisms in our model drift and collide blindly.

In developing a theoretical basis of the reproduction, death, and coalescence processes, only the reproduction and death parts are trivial, given the parameters set in the model simulations. The greatest problem rests with determining probabilities of coalescence. I have already begun this work, and I have reduced the problem into two parts. The first part determines a constant probability that

only depends on the mutability of a parent organism. This describes the coalescence probability of sibling offspring ("sibling rivalry") and may act as a correction to the reproduction rate. The second part of the problem is to determine the coalescence probability as it depends on the distance between any neighboring organisms. For assortative and random mating, this includes offspring generated by either parent of any mating pair – coalescence of half-siblings – as well as coalescence of unrelated organisms. I will refer to any coalescence of this type as "neighbor rivalry". My focus thus far has been on the constant probability of sibling rivalry. I plan to extract an estimate for the probability of neighbor rivalry from simulation data, and I will be able to check my analytical solution for the probability of sibling rivalry with complementary data.

Determining the coalescence probabilities and using the fitness of organisms and the random death process in our model will provide a simple mean-field equation of the population density which has the form of a logistic equation. From this mean-field equation, I should be able to estimate critical mutability for a given competition radius as well as critical exponents for the scaling of population density, correlation length, and correlation time. I intend to show that if the mean-field equation works for the present implementation of our model, then I predict that by varying the competition radius, the parameters of the mean-field equation will guide the prediction for critical mutability. I then plan to measure the critical exponents from simulation data to determine if our model is of the directed percolation universality class. Already, with the control parameter determined from the mean-field equation, the predicted critical mutability is within 10% of the simulated critical mutability.

For the final research chapter, I will explore and quantify scaling behavior for the same model used in the first two chapters. I will consider scaling from the organism level to the lineage and cluster level. I define an organism lineage as the set of descendant offspring from an organism of the original population. Since the model appears to have directed percolation behavior in the temporal dimension, the lineages and clusters may provide measurable behavior to strengthen arguments for scale-invariance. In particular I predict that near criticality:

- 1) Branching and coalescing events at the cluster level exhibit scale-free properties and change according to the reproduction scheme.
- 2) The distribution of times to convergence of all surviving, coalescing lineages follows a power-law, thus exhibiting scale-free properties.
- 3) The distribution of times to extinction of populations scale-free for mutabilities below criticality.
- 4) The "centers of mass" (or centroids) of clusters, and possibly lineages, undergo Lévy walks.
- 5) The aggregation of clusters produces super-clusters whose sizes follow a scale-free (power law) distribution.

The branching and coalescing of clusters is similar to the way organisms branch and coalesce. The primary difference is that in a single time step, whereas organisms are allowed to produce a maximum of two offspring, a cluster may produce many unique clusters. Many clusters may also coalesce into a unique cluster, thus the binary coalescence process at the organism level may be modified to include more than two constituents to give rise to just one at the cluster level.

These mechanisms of clustering should differ according to the reproduction scheme. Clusters in assortative mating should be more “sink-driven”, that is, as organisms mate, their offspring may be more likely to spawn at phenotype locations between parents, and in turn cluster due to converging mate pairs. This should cause a decrease in the convergence of clusters because the offspring should, on average, be nearest to siblings and half-siblings which are most likely to be included in the same cluster. The decrease in convergence of clusters may be analogous with lineages which will be measured more thoroughly with times to -convergence (described below). Measuring nearest neighbor distances and tracking the inclusion of immediate relatives, i.e., siblings and half-siblings, among clusters may provide evidence for the validity of these claims.

Alternatively, clusters in the asexual fission case should be more “source-driven”. Clusters are more likely to form from the growth and breakup of clusters rather than from the coalescence of neighboring clusters. Again, tracking immediate relatives within clusters should provide a measure for this claim. However, to compare the reaction-diffusion events of clusters between asexual fission and assortative mating, I need to develop a measure which accurately parses the predicted differences. One potential measure may be to simply take the ratio of each process rate and the total number of events (i.e., number of coalescing events divided by the total number of branching and coalescing events).

The behavior of lineages is fundamentally different than that of clusters, because there is no branching process of lineages. Furthermore, lines of descent only converge for the assortative mating case. This process causes the number of unique lines of descent to decay monotonically. The times to convergence of unique lineages may follow a power-law in accordance with other scale-free measures at criticality. Simulations with larger initial populations of perhaps several thousand organisms will be needed to accurately test this. The results may provide an indicator of why sympatric speciation is so rare in nature. The probability of maintaining multiple unique lineages is probably vanishingly small, albeit not exactly zero, for all mutabilities. From preliminary data for critical mutability, only a few lineages from the initial 300 organisms survive the entirety of the simulations, suggesting further a notion of dominant lines of descent which are likely well-mixed.

Distributions of times to extinction of populations have been observed to change from normal to power-law with increasing mutability approaching criticality. Furthermore, the times to extinction grow longer both in the minimum and maximum of the distributions. I will characterize the distributions as they change by comparing them with normal, Poissonian, exponential, and power-law distributions. The progression of the distributions as mutability is increased shows an apparent critical slowing down phenomenon. Upon criticality, the times to extinction become infinite which is consistent with the directed percolation universality class. Computationally, an infinite time to extinction may be characterized by allowing simulations to run for comparatively very many generations on the order of  $10^7$ . Due to the large amount of noise in the system “infinite” times to extinction may not always be the outcome for mutabilities considered to be greater than and equal to criticality; however, these events should become vanishingly rare for mutabilities well above criticality.

Centroids, the “center of mass” of clusters and lineages, may undergo scale-free random walks, Lévy walks. On the cluster level, centroids can be treated as particles undergoing a reaction-diffusion

process as described above. Looking at single lines of clusters, the step lengths may be distributed as a power-law. In considering the many lines of descent from clusters, care will be taken to determine an accurate sample of step lengths. One potential method of determining a single line of clusters can be done by starting with clusters from the final generation in a simulation. From a given cluster, tracing back in time through parent clusters and their centroids can provide positions in phenospace from which a total path length and the step lengths taken can be determined. If multiple parent clusters are available to track, then only one can be chosen at random. A possible pitfall of this method might result from some overlapping paths from different starting clusters. However, with the large amount of cluster mixing and the random parent cluster determination, these overlapping paths may be insignificant in the overall distribution of the step lengths within each path.

Lastly, clusters may correspond to a scaled version of organisms. This would indicate that a possible renormalization procedure for the model is to scale organisms to clusters. The sizes of these super-clusters would be expected to retain a power-law distribution near criticality.

This proposed dissertation focuses on the phase transition behavior of a relatively simple, agent-based evolution model in phenotype space on a neutral landscape. Each chapter investigates the phase transition behavior from a different perspective. First, I will demonstrate the occurrence of a phase transition, along with clustering behavior, on a neutral fitness landscape. Next, with the aid of a mean field model, the type of phase transition will be determined. Finally, a demonstration of scale-free behavior at criticality may be used to make predictions about possible evolutionary outcomes such as the difficulty of sympatric speciation in nature and the possibility for various levels of selection.

## Bibliography

- [1] J. Mallet, "A species definition for the modern synthesis," *Trends in Ecology & Evolution*, vol. 10, no. 7, pp. 294–299, Jul. 1995.
- [2] M. A. M. de Aguiar, M. Baranger, E. M. Baptestini, L. Kaufman, and Y. Bar-Yam, "Global patterns of speciation and diversity," *Nature*, vol. 460, no. 7253, pp. 384–387, Jul. 2009.
- [3] M. A. Bedau and N. H. Packard, "Evolution of evolvability via adaptation of mutation rates," *Biosystems*, vol. 69, no. 2–3, pp. 143–162, May 2003.
- [4] J. Clune, D. Misevic, C. Ofria, R. E. Lenski, S. F. Elena, and R. Sanjuán, "Natural Selection Fails to Optimize Mutation Rates for Long-Term Adaptation on Rugged Fitness Landscapes," *PLoS Comput Biol*, vol. 4, no. 9, p. e1000187, 2008.
- [5] V. C. Wynne-Edwards, "Intergroup Selection in the Evolution of Social Systems," *Nature*, vol. 200, no. 4907, pp. 623–626, Nov. 1963.
- [6] A. Traulsen and M. A. Nowak, "Evolution of cooperation by multilevel selection," *Proceedings of the National Academy of Sciences*, vol. 103, no. 29, pp. 10952–10955, Jul. 2006.
- [7] R. Barbuti, A. Maggiolo-Schettini, P. Milazzo, and A. Troina, "A Methodology for the Stochastic Modeling and Simulation of Sympatric Speciation by Sexual Selection," *Journal of Biological Systems*, vol. 17, no. 3, p. 349, 2009.
- [8] N. D. Dees and S. Bahar, "Mutation Size Optimizes Speciation in an Evolutionary Model," *PLoS ONE*, vol. 5, no. 8, p. e11952, 2010.
- [9] H. Cadillo-Quiroz, X. Didelot, N. L. Held, A. Herrera, A. Darling, M. L. Reno, D. J. Krause, and R. J. Whitaker, "Patterns of Gene Flow Define Species of Thermophilic Archaea," *PLoS Biol*, vol. 10, no. 2, p. e1001265, Feb. 2012.
- [10] M. A. Bees, P. H. Couillet, and E. A. Spiegel, "On the bifurcation of species," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 18, no. 4, p. 043114, 2008.
- [11] B. M. Fitzpatrick, J. A. Fordyce, and S. Gavrillets, "What, if anything, is sympatric speciation?," *Journal of Evolutionary Biology*, vol. 21, no. 6, pp. 1452–1459, 2008.
- [12] G. Slade, "Scaling limits and super-Brownian motion," *Notices AMS*, vol. 49, no. 9, pp. 1056–1067, 2002.
- [13] P. J. Clark and F. C. Evans, "Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations," *Ecology*, vol. 35, no. 4, pp. 445–453, Oct. 1954.
- [14] M. Henkel, H. Hinrichsen, and S. Lübeck, *Absorbing Phase Transitions*. Springer, 2008.