The Evolution of Biodiversity: A Simulation Approach

by

Carlo C. Maley

B.A., Oberlin College (1991)
M.Sc., University of Oxford (1993)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1998

Author . . . .
Department of Electrical Engineering and Computer Science
May 1, 1998

Certified by . . . .
Rodney Brooks
Professor of Computer Science and Engineering, MIT
Thesis Supervisor

Certified by . . . .
Michael Donoghue
Professor of Biology, Harvard University
Thesis Supervisor

Accepted by . . . .
Arthur C. Smith
Chairman, Departmental Committee on Graduate Students

The Evolution of Biodiversity:

A Simulation Approach

by

Carlo C. Maley

Submitted to the Department of Electrical Engineering and Computer Science
on May 1, 1998, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

What have been the most important factors in the diversification of life? A configuration (individual-based) model (MoD) was constructed to examine the origins and maintenance of species diversity through time. The model represents a species as a connected component in a graph of the potential mating relationships between organisms. This allows us to detect speciation events and track species diversity over time, and so test many "untestable hypotheses" for the causes of diversification. The results suggest that much of the emphasis placed on the evolutionary innovations, in resource utilization and predation interactions, in order to explain the diversification of a group, is flawed at best. Furthermore, habitat heterogeneity had little impact on species diversity. Instead, the model points to the importance of geographical isolation as a primary causal factor for diversification, along with the evolution of specialization and sexual selection, in the form of positive and negative assortative mating. The influence of assortative mating is particularly interesting because it is opaque to the methods of paleobiology. Finally, both theoretical and experimental tests are sketched that may help to falsify the predictions of the model.

Thesis Supervisor: Rodney Brooks
Title: Professor of Computer Science and Engineering, MIT

Thesis Supervisor: Michael Donoghue
Title: Professor of Biology, Harvard University

Acknowledgments

The years of my graduate student experience have been almost embarrassingly happy ones. During that time I have incurred an intellectual debt to many friends, colleagues and mentors. It is a pleasure now to acknowledge their contributions. First let me thank my mentors, Rhys Price Jones, Hal Caswell, William D. Hamilton, Rodney Brooks and Michael Donoghue. They have been, without exception, generous in their support, encouragement, and license. Bob Berwick, Dave Cliff, Richard Lewontin, Dan Hartl, Charles Leiserson, Carl Pabo, Stephanie Forest, and Marvin Minsky have provided excellent instruction and advice. I have received technical and academic tutoring from many peers and friends, in particular much of this work, and my entire research career for that matter, has been made possible through frequent tutoring from Miller Maley. In addition Michael Wessler, Loren King, Eric Johnson, Lynne Butler, Una-May O'Reilly, and Belinda Chang have provided both desperately needed help and useful suggestions. I am grateful for both their contributions and their friendship. Thank you. I should note that both Michael Wessler and Una-May O'Reilly put in long hours pouring over drafts of this thesis. I am further indebted to Hyojoo Kang for assistance in editing this monster as well as all the myriad ways in which she makes life good. Thanks also to my loving parents. They deserve most of the recognition (blame?) for making me who I am today. I should thank Terry Jones, who unwittingly, via Melanie Mitchell, via Annie Wu contributed his pseudo-random number generator C code. I owe an emotional debt to Pico and Femto for their unwavering support. Oh yes, and lest I forget, most of this work was completed with the near-unconditional love of the United States Army, NDSEG grant DAAH04-95-1-0557 as well as the MIT Artificial Intelligence lab. Finally, while the camera is still on me, let me just say, "Hi Mom!"

Contents

List of Figures

List of Tables

Prologue

This dissertation is an attempt to understand the fundamental processes of biology through the construction of a computational model. It is thus an interdisciplinary work spanning one of the chasms between biology and computer science. I am committed to communicating to audiences in both fields, as well as the general public. I have striven to make what mathematical analysis I include both clear and accessible to the non-mathematically inclined. On the other hand, I have also tried to avoid excessive use of jargon from the oversized lexicon of biology. Where I do use technical terms, I usually tag them with an explanatory phrase or follow-up sentence. I can only ask the analytically inclined reader's forgiveness if she feels that I am insulting her mathematical talents. Similarly, I must ask the indulgence of my biologist colleagues for my preference for common English over the concise terms of their trade.

Perhaps one further note on style, or apology, should be made. I am worried by the cultural division between academia and the general public. I have taken a cue from an article a concerned relative once clipped for me. It was written by a historian who was worried about the same problem. One of the contributing phenomena to this chasm is the dry language in which most science, and for that matter, academic writing is couched. A century and a half ago, Darwin stunned the world with his arguments rich in metaphor and imagery (Darwin 1859). Of course, it took him twenty years to finish it. Not only must we bore our readers to tears in an attempt to sound objective, but we also seem to prefer impenetrable prose, as if to say, "if you don't understand me, it is probably because I am smarter than you." It is a nauseating kind of school-yard attitude. So I hope you will appreciate, or at least understand, why I have chosen to write with a relatively loose style and even a sprinkling of colloquialisms. I do not think I have sacrificed any rigor or precision. Although, I am sure that I have padded the text. With any luck, you may even …*gasp* …enjoy reading this. Well, I may be going a little too far there, but at least that is my ambition. Rod Brooks always says to set your sights high.

Chapter 1

Introduction

There are two dramatic aspects of life as we know it; adaptations and diversity. We are immersed in a natural world of intricate variation. The functional designs of life far surpass our best engineering efforts and its diversity has resisted centuries of effort to catalogue it. Darwin explained the origin of adaptations, and his ideas are now being used in engineering with greater or lesser success. However, the origin of biological diversity remains a muddle. Why has life diversified so dramatically over the last four billion years?

A large class of outstanding problems in evolutionary biology concerns the gap between our understanding of macroevolutionary dynamics like speciation, and microevolutionary dynamics like natural selection, migration, and genetic drift. Darwin (1859) and Mendel gave us the basis for a solid understanding of microevolutionary dynamics, but so far, macroevolution lacks the analogous grounding. Computational simulations offer the hope of examining the macroevolutionary dynamics that emerge from an implementation of the microevolutionary dynamics. The following pages document an attempt to bridge the gap between microevolution and macroevolution.

The primary contribution of this work is three-fold. First of all, I introduce a rigorous definition of species that allows the efficient identification of species within a collection of simulated individuals and can easily be implemented in any model that includes sexual reproduction. This definition is the rock upon which the entire edifice of results stands. A commitment to a different species definition would produce different results. Second, I develop a model that is arguably a reasonable test-bed for the comparison of the many theories for the diversification of life. The model provides a sort of level playing field upon which the various theories may compete and their relative strengths measured. Third, the results of the model point out a number of blind spots, reject some of the most favored hypotheses, and emphasize the importance of other hypotheses that have yet to enjoy significant attention in our pursuit of an understanding of diversification. The remainder of this chapter tackles the description of the problem of diversity, the kinds of answers that have been proposed, my approach to the problem, and previous related efforts by other researchers.

1.1 The Problem of Species Diversity

If we look back through the dark lens of fossil evidence on the state of the Earth 3.5 billion years ago, the world was not a very interesting place... at least, not from the perspective of biology. Life at that time was a collection of small, single celled organisms living in the oceans. The continents were barren. To a human eye, it probably would have looked like a big dead rock, nearly unrecognizable compared to the world we know today. Over those 3.5 billion years, life proliferated and diversified. What accounts for the change? Why has life diversified? There are three major obstacles to understanding the diversification of life; (1) the metric upon which diversity will be measured, (2) the poor quality of data in the fossil record, and (3) the impracticality of implementing controlled experiments.

Diversity might be measured either by morphological differences between organisms or by the number of distinct living (extant) species. Unfortunately, morphological complexity and variation is difficult to quantify (Roy & Foote 1997). Instead, I have chosen to focus on species diversity. That is, a direct count of the number of living, or extant, species. As we will see in Section 1.2, there is nothing "simple" about this count.

The fossil record is notorious for its poor quality of data (Darwin 1859). From our perspective, the problems in the data fall into two categories; poor resolution and poor sampling. The resolution of fossil data is such that paleontologists regularly work at the taxonomic level of the family (Erwin et al. 1987; Van Valen 1984; Raup & Sepkoski 1982; Flessa & Jablonski 1985; Hallam 1990a). It is often impossible to distinguish between fossilized species. Because families are relatively arbitrary groupings of species (Sepkoski & Kendrick 1993), there is some distortion between observational data and the underlying dynamics of species diversity. We can only look through the lens darkly. Furthermore, organisms are fossilized only under rare conditions, mostly where rapid sedimentation is occuring. This means that for any given time period, the fossil record only shows us a small fragment of what was going on in the world at that time. For example, in the entire North American continent, the Mississippi delta is the only location conducive to fossilization. Those fossils are unlikely to reflect the species distributions of other areas on the continent. The problems in the fossil record mean that it is extremely difficult to definitively answer questions about diversification. Paleontology does, however, shape and constrain the relevant open questions about the history of life.

Finally, the temporal and spatial scales of evolution preclude most controlled experiments. To date, the predominant response to this problem has been to use observational studies of phylogenetic trees (Mooers & Heard 1997; Sanderson & Donoghue 1996; 1994; Slowinski & Guyer 1990; 1991), the fossil record (Benton 1990; Raup & Boyajian 1988; Jablonski & Bottjer 1990), and limited experimental studies to resolve issues of the evolution of diversity (Rice 1985; Ehrman 1965; de Olivera & Codeiro 1980). A more controlled approach can be found in computational simulations. If a reasonable model of the biological dynamics can be constructed, experiments can be carried out on the model that go far beyond the control and time scales of laboratory or field experiments. However, these experiments cannot replace field and laboratory experiments. There will always be a conceptual leap between the data from the model and the real world (Caswell 1976). In this case, I am proposing to develop a model of the evolution of organisms in order to study the processes of diversification.

The problem of the origin and maintenance of species diversity can be sharpened to the question of what are the most important factors that have influenced speciation and extinction in the history of life?

1.2 What is a Species?

Investigations into the dynamics of species diversity depend fundamentally upon the concept of a species. Unfortunately, there is a vast diversity of species concepts and none of them unambiguously partitions organisms into mutually exclusive groups. (de Queiroz 1998; Mayr 1942; 1963; 1982a; Simpson 1961; Van Valen 1976; Wiley 1978; Cracraft 1983; Paterson 1985; Templeton 1989; Baum & Donoghue 1995; Mishler & Theriot 1998). If we were to plot organisms as points in some high-dimensional space where each dimension represented a measurable characteristic, organisms would form recognizable species clusters (Templeton 1989). This is not to say that we do always pick out the right set of characters[1] to identify species. The *Anopholes gambiae* mosquito, which transmits the *Plasmodium falciparum* amoeba that causes malaria, posed something of a mystery. Some of the mosquitos from West Africa were free of the amoeba and so posed no threat to the humans in the area, while other mosquitos in the same area were filled to the brim with the parasite. Further investigation revealed that what we thought was a single *Anopholes gambiae* mosquito species was actually a complex of six species, all nearly identical to the human eye (Mathiopoulos & Lanzaro 1995). However, only two of the six species served as vectors for transmitting malaria. Similar discoveries have been made all across the globe in Anopheline mosquito populations (Subbarao & Sharma 1995; Foley et al. 1996).

The predominant species definition, the "biological species concept," or more appropriately the "reproductive species concept," defines two organisms as belonging to the same species if they can potentially mate and produce fertile offspring (Mayr 1963; 1982a). More specifically,

> Species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups. (Mayr 1942, p.120)

Mayr later revised the definition by adding the requirement that a reproductively isolated group must acquire adaptations that allow it to co-exist with potential competitors. That is, the new group must find an unoccupied niche[2].

> A species is a reproductive community of populations (reproductively isolated from others) that occupies a specific niche in nature. (Mayr 1982b, p.273)

The requirement that a reproductively isolated community successfully compete with other species confuses speciation with the success of the new species. While there is no doubt that the availability of resources constrain the diversity we see, this constraint is mainly manifest through competitive exclusion. That is, through one species driving the other

[1] A "character" is biology jargon for a trait of an organism.

[2] A niche is essentially a set of resources in the environment that can support a population. See Section 6.1 for a discussion of this concept.

to extinction. There is no need to elaborate the reproductive species concept with the qualification that the population must occupy a niche. A reproductively isolated community shares an evolutionary fate, whether that fate is rapid decline or eventual success.

The clustering in multidimensional character space is probably due in part to reproductive barriers between gene pools that prevent the mixing of genes from different species (Mayr 1942). This forms the intuition behind the reproductive species concept. However, the definition obviously suffers problems when applied to asexual or extinct organisms. But even among living sexual organisms, there remain problems[3]. Species gene pools are sometimes mixed through hybridization of similar species. Biologists must live and work with the fact that the concept of a species, like many of the concepts in biology, does not have clean boundaries.

Predictably, the shortcomings of the reproductive species concept has sparked numerous attempts to improve upon it (de Queiroz 1998; Paterson 1985; Templeton 1989; Simpson 1961; Wiley 1978; Van Valen 1976; Cracraft 1983; Mishler & Theriot 1998; Baum & Donoghue 1995). The profusion of concepts and proposals remain today. Most biological experiments have not required us to operationalize the concept of a species in the abstract, so there has been little pressure to weed the field. However, any attempt to model species as collections of organisms must face the issue square on. For the purpose of modeling, I will briefly consider and, perhaps unfairly, reject all the common alternatives to the reproductive species concept.

The recognition concept (Paterson 1985) emphasizes a common fertilization and specific mate recognition system. But if such systems vary across organisms and evolve, how would we identify commonality? Patterson opts for the most inclusive population that shares a common fertilization system. In the implementation of a model, this would probably be equivalent to the reproductive species concept.

[3]Before we go any further, we should take a moment to consume the proverbial grain of salt. Life is filled with such complexity and variation that any general statement will be violated by some exceptions. The gene pools that we like to think of as species are not so inviolate as they might seem. We know from ancient times that after the formation of the eukaryotic cell, with its nucleus and mitochondria, DNA slopped across species barriers. Current accepted theory has it that the proto-eukaryotic cell, lacking mitochondria, engulfed what was at the time a free living bacterium. The two organisms became mutualists, relying on one another for survival. And over the eons, DNA from the mitochondria has been integrated into the eukaryotic nucleus (Margulis & Sagan 1986). Perhaps more dramatically, the entire class of retroviruses introduce foreign DNA into their hosts as part of their life cycle. While that DNA is generally "designed" to be excised when the virus leaves its dormant stage, bits are sometimes left behind and integrated into the host's DNA.

The evolutionary species concept (Wiley 1978) requires long time periods to identify populations evolving as a unit. This is neither well defined nor applicable to a snapshot or instantaneous description of a set of organisms. The problem of identifying a species from an instantaneous description of a set of organisms also shows up in most of the phylogenetic species concepts (Baum & Donoghue 1995; Cracraft 1983). At any given time in a model, we would like to be able to determine the number of species present. Reference to the history of the organisms would require us to maintain detailed records of ancestry.

The ecological species concept is based on a population occupying a niche or "adaptive zone" (Van Valen 1976). To implement this would require the quantification of adaptive zones as well as a method to unambiguously cluster organisms within that zone. However, the attraction of modeling microevolutionary dynamics and ecological interactions is that higher level phenomena such as adaptive zones, emerge out of the microlevel phenomena. In addition, the ecological species concept would elide extinctions due to competitive exclusion. Two groups of organisms that are reproductively isolated (species under the reproductive species concept) may share the same set of resources. However, the expectation is that they cannot share those resources indefinitely. One population will out compete the other and drive it to extinction. The ecological species concept would miss this dynamic.

The conventional wisdom for why organisms cluster into the groups in character space has been that survival poses a number of constraints on what character combinations are viable. In other words, organisms evolve to cluster around niches. However, Bernstein et al. (1985) counter this reasoning with the observation that sexual reproduction imposes a penalty for low population densities. If an organism can only mate with similar organisms (organisms near to them in character space), then an organism with rare characters will have difficulty finding a mate and so its lineage will tend to die out. The theory thus predicts that sexual organisms will tend to cluster in character space and asexual organisms will tend to spread out over character space. These two perspectives are not mutually exclusive. Both niches and sexual reproduction probably have important effects on the evolution of characters.

The problem of unambiguously clustering organisms applies to any attempt to classify organisms by similarity of phenotypic or genotypic characters. This approach shows up in the genotypic cluster definition, the morphospecies (Mayr 1942),or phenetic species concept and the diagnostic approach to the phylogenetic species concept (de Queiroz 1998, provides an excellent review of species concepts). The phenotypic species concepts are misled by interbreeding populations that exhibit multiple forms, undifferentiated but reproductively isolated sister taxa, as well as convergent evolution in reproductively isolated populations (Baum & Donoghue 1995).

The cohesion species concept (Templeton 1989) states that a "species is the most inclusive population of individuals having the potential for phenotypic cohesion through intrinsic cohesion mechanisms." A cohesion mechanism is anything that tends to promote phenotypic similarity in a population. For example, the potential to exchange genes through sexual reproduction is a cohesion mechanism. But so is the fact that many potentially viable mutations will not prove successful in the current ecological context of the population. The cohesion species concept has the nice property that it attempts to acknowledge the many pressures that maintain the boundaries of populations. However, because it subsumes aspects of the ecological species concept, it suffers from the same problems in implementation.

For all its drawbacks, the reproductive species concept, or the "gene pool species concept" in de Queiroz's (1998) terminology, has a certain algorithmic elegance that facilitates its application to a computational model. Because a model must specify the conditions under which two organisms can successfully mate, the concept can be applied unambiguously. Furthermore, by keeping track of the entire network of who can mate with whom, speciation events can be detected instantaneously. Section 2.2.3 will describe in detail how this is done. It is essential to note that, in choosing reproductive isolation as our criterion for identifying species, we have made a fundamental commitment to the separation between speciation and morphological divergence (Cracraft 1989). The results of the model will speak to a particular view of speciation and species diversity, but not to any observable differences across species. That is a problem for a different model and a different time. Having settled on an operational measurement of diversity, we may move on to consider the testable hypotheses for the causes of diversification.

1.3 Three Levels of Answers to the Problem of Diversification

Research into diversification typically focuses on physical and biological factors that might increase the rates of speciation or decrease the rates of extinction (Benton 1990; Marzluff & Dial 1991; Sanderson & Donoghue 1996; Mooers & Heard 1997). This is the level at which most of the experiments on the model will be conducted. However, in focusing on these factors, researchers often overlook two fundamental and still open questions; how do the basic processes of diversification work? That is, if the processes of speciation are such that speciation rates are inherently higher than extinction rates, increasing diversification is a predictable outcome. Yet, speciation rates and extinction rates may not be independent. This leads us to the second open question, is there some form of regulation on the numbers of extant species? Is diversification itself diversity dependent? Only if this last question is answered positively, such that there is a diversity equilibrium, need we speculate about what factors might have raised the ceiling on diversity over time.

1.3.1 The Basic Processes of Diversification

The processes of speciation and extinction depend on fundamentally different characteristics of a population; mating interactions and individual death. Thus, it is reasonable to consider them as independent until it is shown otherwise. But if speciation and extinction are independent, we only need evidence of higher speciation rates relative to extinction rates in order to explain the increase in species diversity over time.

Speciation

There are a number of theories describing the process of speciation (Mayr 1982b; Rice & Hostert 1993). It has been hard to support any one over the others because it is difficult to observe the process as it unfolds. Most of our evidence for speciation events is post hoc. Speciation theories vary along two parameters: the importance of geographical separation of the incipient species and the importance of disruptive selection forcing populations to diverge. Geographical separation of populations in speciation models varies from an extreme of full separation (allopatry), to adjacent but separate (parapatry), to the budding of a small population on the periphery of a larger one (peripatry), to gradients across which populations vary continuously (a cline), to no separation whatsoever (sympatry). In most cases, there is an assumption of differential selection on the two populations, causing them to diverge. However, some theories of speciation have also emphasized genetic drift through sampling effects in small populations (founder effects or bottlenecks), or simply through the build up of selectively neutral mutations. An additional proposal has suggested that a partial divergence between populations should be reinforced if the hybrids are less fit than either parental population. Any allele[4] in the parental population that biases an organism to only mate with other organisms from its population should spread because its offspring will be more fit than the hybrids. So we might expect the two populations to diverge further. This is called speciation by reinforcement and it can be added as a variation to any theory that includes some degree of contact between diverging populations (Rice & Hostert 1993).

The best theory we have for the process of speciation focuses on geographical separation of populations (Mayr 1982b). Allopatric[5] speciation is thought to occur when some members of a population are physically isolated from the main body of the population. This might happen by being blown onto an island, or by migrating over some geographical barrier like a mountain range. Since the relatively small gene pool of the colonizers is unlikely to be representative of the main population, and the selective pressures in the new colony may differ from the pressures on the main population, the two populations are likely to diverge genetically over time. If the separation is maintained for long enough, the populations may diverge to the point that organisms cannot successfully interbreed between populations and thus the gene pools become separated. Needless to say, this is difficult to observe and test (Rice & Hostert 1993, review the attempts). However, the observation of diversity of similar species across island chains tends to support the allopatric speciation model (Futuyma 1986, p. 224).

Until recently (Feder et al. 1997; Shoemaker & Ross 1996), there has been little evidence to support sympatric speciation (Futuyma 1986; Futuyma & Peterson 1985). The most

[4]An allele is a particular value for a gene that has multiple different possible values. So "blue eyes" is an allele of an eye color gene.

[5]Technically, allopatric speciation envisions a population divided into two remote, possibly equal sized populations with no migration between them. There are a number of variations on allopatric speciation. Parapatric speciation posits a divergence between adjacent but non-overlapping populations. Peripatric speciation, or budding, is similar to allopatric speciation except that it specifies that a small number of colonizers are separated from the main population and establish a new population in a geographically semi-isolated locale. These are all essentially elaborations of the basic dynamic of separation and divergence. For simplicity I group them under the term allopatric speciation and refer to them as such.

extreme version of the theory requires a mutation that at once sets up a reproductive barrier and establishes a viable subpopulation. This is especially problematic for sexual species that must find a viable mate. A reproductive barrier is necessary to prevent the homogenizing force of gene pool mixing. A more digestible model of sympatric speciation relies upon disruptive selection. That is, selection for more than one extreme of a character. Disruptive selection will tend to split a population into subpopulations, clustered around the extremes of the selected character. If the population develops assortative mating, that is, organisms tend to choose similar organisms for mates, then a reproductive barrier might form (Mayr 1942).

In their review of the laboratory experiments on the different forms of speciation, Rice and Hostert (1993) argue that it is divergent selection, not geographic isolation that has shown the most promise as an agent of speciation. They find little support for speciation by reinforcement. Nor do they find support for speciation by bottlenecks where populations are repeatedly reduced to a single gravid (pregnant) female and then allowed to expand again to 1000 individuals. Such a treatment tends to only result in mild pre-zygotic and post-zygotic isolation, not nearly enough to cause speciation (Rice & Hostert 1993).

Extinction

There are two contrary forces in diversification, speciation and extinction. Diversification can just as easily be stimulated by a drop in extinction rates as it can by an increase in speciation rates. Extinction is thought to occur mainly through stochastic fluctuations in small populations, competition, predation, or habitat change (Raup 1991; Flessa et al. 1986). These fluctuations may be enhanced by environmental fluctuations or simply through the stochastic variations of prey, predators, or other species in the food web. Most of the research into extinctions has focused on "mass extinctions." This has included the attempt to distinguish mass extinctions from "background extinctions." It has also focused on the attempt to distinguish the causes of various mass extinctions (Raup 1991; Alvarez et al. 1980; Erwin 1994; Hallam 1990b; Brenchley 1989; McGhee 1989), and the parallel question of what characteristics tend to make species resistant to mass extinctions (Jablonski 1989). There is even some intriguing evidence that mass extinctions clear out adaptive space in such a way that the biota recovers to diversity levels that exceed the numbers of species previous to the mass extinction (Sheehan 1991; Sepkoski 1984). Why this happens still remains an open question.

Our interest in diversity dictates a primary focus on "background extinctions." Are there factors that link extinction rates to speciation rates? Some have suggested that as "adaptive space" fills up, competition for limited resources gets fiercer, and there is a concomitant rise in extinction rates (Ricklefs & Schluter 1993; Cornell 1993). Most of the theories for the diversification of life focus on speciation rather than extinction. This is probably, in part, a simple semantic bias in our perception of the phenomenon of diversification. But, more fundamentally, it may be a reflection of the underlying complexity of the phenomena. Extinction is just the sum of many individual deaths. Speciation, on the other hand, is the result of behavioral interactions (mating), genetic changes, as well as the death of the select group of individuals that are able to mate with both of the incipient species. There seems to be greater room for complexity, and ignorance, in the dynamics of speciation[6].

1.3.2 Regulation of Diversity

If it turns out, as some have suggested (Hoffman 1986), that the likelihood of speciation is greater than the likelihood of extinction, we should expect that diversity should increase exponentially. This might be considered the null model. However, the standard view, accepted by most biologists, is that there are some dynamics that regulate diversity so as to turn the exponential growth into logistic[7] growth (Valentine 1980; Walker & Valentine 1984). This is based on the idea of competitive exclusion (Walker & Valentine 1984; Huston 1994). When two species attempt to utilize the same resources, one is likely to drive the other to extinction. However, recent work by Benton (1995) finds that if one looks across all organisms, family diversity[8] has been increasing exponentially over time. The competing hypotheses are represented by the exponential and logistic curves shown in Figure 1-1.

It is possible that both sides are right. One the one hand, for any given non-universal taxon, or grouping of species, species diversity within that taxon approaches an equilibrium. On the other hand, across all taxa, species diversity may continuously increase. In general,

[6]Though I may be wrong. The fundamental mechanisms of extinction probably deserve more attention than they have been granted.

[7]The logistic curve begins like an exponentially increasing function but then starts to flatten out as it approaches some asymptote, or carrying capacity.

[8]Families are somewhat arbitrary groupings of genera, which in turn are somewhat arbitrary groupings of species. Counting families as a proxy for species diversity introduces a bias in our perception of diversity patterns over time (Maley et al. 1997)

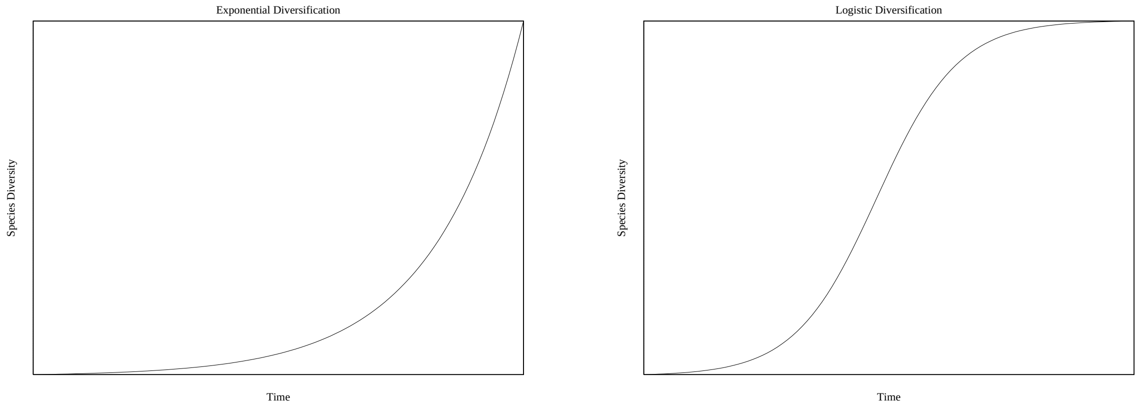


Figure 1-1: An exponential (left) and a logistic curve (right). Both have been proposed as a description of the pattern of species diversity over time. The answer to which is a more accurate description depends on whether life on earth has neared some form of maximum diversity as depicted by the logistic curve, or alternatively, life has yet to encounter any checks on its diversification.

every new species provides a new form of resource for other species. In this way the number of potential niches may grow over time, perhaps exponentially. However, if a study is restricted to a small taxon, it is unlikely, in general, that members of the same taxon could evolve to exploit new species of that taxon. To belabor the example, it is much more likely that insects, not mammals, would parasitize a new mammalian species. Thus an analysis of diversity levels of mammals would miss the increase in total diversity.

If diversification proceeds logistically, we are left with the question, what factors regulate diversity? If diversification proceeds exponentially, we must at least ask, why is speciation more likely than extinction? But if both are correct, then at least two questions must follow. What factors regulate diversity in non-universal taxa? And, what factors lift the ceiling on diversity?

1.3.3 Factors Influencing Diversity Equilibria

While there is clearly a significant amount of work that needs to be done to address these initial questions, most work in the field has focused on trying to answer the last question, what factors might drive diversity up beyond its previous levels?(Benton 1990, for a review)

Benton identifies 14 suggested explanations for the diversification of life and then attempts through argument and fossil data to distinguish the plausible from the implausible (Benton 1990). These explanations are broadly classified into four categories, (1) diversification patterns may simply be artifacts in the data, (2) diversification may be in some sense inevitable, (3) diversification may be driven by physical factors in the environment, and finally, (4) diversification may be driven by biological factors in the environment.

Diversification as an Artifact

Perhaps our perception of an the increasing diversity of life is an artefact. This supposition is a denial of the existence of an increasing diversification trend in biological history. Raup has argued for the possibility of four phenomena that might give the false impression of a diversification over time (Raup 1972). Perhaps we find more fossils for recent organisms because the more recent rocks are nearer to the surface and so more easily sampled by paleontologists. Similarly, as sedimentary rock is metamorphosed slowly over time, older fossils are lost. Paleontologists themselves might also be to blame for a misperception of diversification. Perhaps we tend to study recent organisms more than older organisms and so have identified a greater diversity in recent samples. There is a commonly acknowledged phenomenon called the "pull of the recent" in which our knowledge of living species informs our understanding of their close relatives in the recent past. This might also give us a better resolution on the more recent fossil samples and so give the impression of higher diversity in recent samples as compared to older samples.

While there is little doubt that these are all sources of error, the community has reached a consensus that the observation of diversification is real and not an artifact. This conclusion was reached through a variety of methods. Sepkoski et al. (1981) compared the diversity curves generated by five relatively independent data sets for marine fossils and found that all five qualitatively agreed. Signor (1982) attempted to correct for these biases in the data and still found strong evidence for the diversification of life.

One must also be careful when drawing conclusions about lineage diversity from family diversity data. The two are not well correlated (Valentine & Walker 1986). However, it is unlikely that an exponential curve in family diversity would hide a non-exponential curve in lineage diversity. If anything, there is probably a tendency to underestimate lineage diversity with family diversities (Valentine & Walker 1986).

The Inevitability of Diversification

Benton points out that a certain amount of diversification is inevitable. Because all clades begin with one species, and the absence of all species is an absorbing boundary, the fossils we will see are very likely to come from phylogenies that diversified over time. Benton calls this "cladistic inevitability." Of course, this does not mean that these clades will continue to diversify. It only argues for an initial diversification.

Most hypotheses for the inevitability of diversification reject the assumption that diversification follows a logistic curve and instead propose that life is diversifying exponentially. This opens up simpler explanations based on the hypothesis that speciation is more likely than extinction rather than requiring some driving force for diversification. Hoffman assumes that speciation rates and extinction rates are independent. He then argues that the available data cannot reject this null hypothesis. This means that "there is no demonstrated need for macroevolutionary laws of biotic diversification (Hoffman 1986)." In fact, Benton himself, argues that the record of all life, at the family level, shows an exponential growth (Benton 1995).

Diversification due to Physical Factors

A "physical" factor here is assumed to be some non-biological property of the environment. The two propositions for physical factors driving diversity focus on changes in climate and topography over either space or time. The hypothesis that fragmentation of climate and topography might stimulate speciation is based on an allopatric model of speciation. It may be construed on the large scale, as in the break up of the Pangaea continental land mass (Valentine & Moores 1972), or on a smaller scale, as in the fragmentation of a forest habitat (Benton 1990).

Cracraft (1985) has proposed that change in time, as opposed to space, may drive diversification. He is explicit in arguing that it is not the topographical complexity that drives speciation but the *change* in topography that is the important factor (Cracraft 1985, p. 799). Cracraft adds to this the hypothesis that "the probability of extinction is a function of the change in the value of environmental harshness to which a species is exposed (Cracraft 1985, p. 808)." The underlying assumption that supports the emphasis on change is that at any given time, the vast majority of extant species can survive in their environment, regardless of its absolute harshness. Thus, it requires a change in the harshness to push a species beyond its capacity to adapt and thus send it into decline and eventual extinction. Cracraft's hypotheses can be distinguished from Benton's proposal by the directionality of change. In Benton's simpler idea, species diversification is driven by topographical and climatic diversification. In contrast, Cracraft does not require the environment to become more complex, but only to change.

Benton (1990) incorrectly categorizes Vermeij's concept of "escalation" as a theory of diversification through biological pressures[9]. Vermeij's concept of "escalation" invokes competition and predation as key engines of evolutionary change (Vermeij 1987). While environmental pressures can lead to "escalation," perhaps the most interesting examples are the arms races between species. While this may explain rapid evolution of new characters, it does not explain an increase in speciation. In fact, Vermeij argues that the factors that stimulate competition and escalation, with the invasion of new species, also tend to stimulate speciation. What are those factors? Changes in climate and topography (Vermeij 1987, p. 377). An example of this might be the way in which fluctuations in temperature can repeatedly isolate and release populations in mountain habitats. This might approximate the "hotbeds of evolution" observed in archipelagoes.

Diversification due to Biological Factors

Biologists, being biologists, seem to favor biological explanations for diversification. These often rest on the concept of a "key innovation." This refers to the evolutionary "discovery" in a lineage of some novel character that in turn stimulates the diversification of that lineage. For example, Benton considers, as a sort of straw man, that perhaps key innovations that reduce gene flow between populations may have raised speciation rates. These "innovations" may have been any changes in the chromosomes, physiology or behavior of the organisms that result in a reduction of the population of potential mates that might yield fertile offspring. However, it is hard to see how such innovations could consistently arise, in preference to innovations that enlarged the pool of potential mates, over the last half billion years (Benton 1990; Cracraft 1985).

In contrast, a number of people have suggested that extinction has gradually selected for species with a resistance to extinction. This is based on Raup and Sepkoski's observation that family level extinctions appear to have declined in frequency over time (Raup & Sepkoski 1982). This would lead to an increase in diversity levels. The fact that this is a

[9]Benton also mistakenly includes Van Valen's "Red Queen's Hypothesis" (Van Valen 1973) as an argument for diversification. In fact, it is an argument to explain his perception of a constancy of extinction rates in lineages. Van Valen proposes nothing in the way of speciation probabilities, and so cannot be said to explain diversification with his hypothesis.

"group" (in this case, family) selection argument does not invalidate it. Selection certainly operates on such higher levels, although at a much slower pace.

Van Valen argues that there have actually been two distinct periods of decreasing family extinction rates: one before the Permo-Triassic mass extinction (the "mother of all mass extinctions"), and one after it (Van Valen 1984). After the Permo-Triassic mass extinction, family extinction rates were "reset" to a high rate and then decreased again, this time more quickly than the decrease before the mass extinction. He argues that the two distinct periods of decreasing extinction rates would not support a general "selection for resistance to extinction" hypothesis. Instead, he suggests that the decreasing rates of family extinction might come from a decrease in competitive pressure by the reduction of niche overlap. In other words, increasing specialization or expanding into unoccupied adaptive space.

Boyajian (1986) points out that younger families are more likely to suffer background extinction than older families. The fact that more recent families contain more species than families earlier in the fossil record is emphasized by Flessa and Jablonski (1985). Thus, the reduction of extinction rates in families may only be the result of the increase in family size. However, the connection between decreasing family extinction and species diversity levels is unclear. If species diversity really is increasing, and the pattern of family diversity not due to families evolving a resistance to extinction[10], then why is it happening?

Similar to Van Valen (1984), Benton favors arguments about adaptive space. Perhaps key innovations have continually arisen that have opened up new niches (Benton 1990). For example, the colonization of dry land by plants opened up a cornucopia of niches, while creating more for the insects and reptiles to follow. In other words, diversification may be driven by the occasional burst of expansion into new adaptive space. Diversity might also rise through the increasing subdivision of niches by specialization. That is, adaptive space might become more densely packed with species.

Sepkoski identifies three distinctive "evolutionary faunas" in marine animals, the "Cambrian, Paleozoic, and Modern fauna." Each fauna dominated the fossil record for a time but was replaced, after a mass extinction, by a successor. Importantly, the successor faunas, the Paleozoic and the Modern faunas, reached higher diversity levels than their predecessors.

[10]Flessa and Jablonski point out that an increase in speciation rate can be seen as a family level adaptation that increases resistance to family extinction. This is equivalent to our first biological hypothesis that perhaps there has been family level selection for traits that reduce or fragment gene pools.

Thus, the increase in diversity can be framed in terms of competition between these faunal groups. However, we have no explanation of why the fauna with the higher diversity capacity replaces its predecessor after a mass extinction (Sepkoski 1984). So this remains more of an observation than a hypothesis for what drives diversification. Knoll et al. (1996) argue that the late Permian mass extinction selected for physiological traits in the survivors. Yet, the connection between physiology and the subsequent diversification is still unclear.

In Summary

Benton argues for the importance of key innovations that have continually opened up new niches in adaptive space. He also supports the hypothesis of an increasing specialization in the biota, with some minor effects of climatic and topographical change fragmenting populations. He arrives at these conclusions by a process of citing contradicting evidence, dismissing hypotheses as untestable (e.g., Cracraft's (1985) suggestion that climatic fluctuations may have stimulated diversification), and using the fossil record as a sort of observational experiment to test the remaining hypotheses.

Many of the hypotheses that Benton dismisses as untestable can be addressed within a model of diversification. Furthermore, the use of the fossil record as an "experiment" generally limits the conclusions to mere correlations rather than causal effects. While a model of diversification may loose predictive power in its abstraction, it has the important advantage of allowing experiments that include the manipulation of parameters to determine causal relationships within the model. This is the challenge that lies before us.

1.4 The Role of Computational Models

Computational models give us a vision of how the micro-level phenomena interact to produce the macro-level. They are the architectural sketches that show how the building blocks together form the visible structures. The power of reductionism is undisputed in modern science. But once we develop an understanding of how the constituent pieces work, there remains the problem of their interaction. How do the basic units of the system combine to produce the macro level phenomena?

In theoretical biology, the missing gap between micro and macro-level phenomena can be seen from both sides of the chasm. In paleobiology, we regularly study the dynamics of whole clades, and only if we are lucky, species dynamics. We lack the data to infer the micro-scale processes that lead to these changes. Across the void dwell the population geneticists. Population geneticists study the dynamics of a few genes, at the most, in populations of organisms. Occasionally we are able to say something about species and speciation (Orr 1995, for example), but the time scales are generally constrained to the sampling of organisms during a researcher's career and so can say little about the vast evolutionary changes that paleontologists document over hundreds of millions of years. Furthermore, the complexity of biological systems must be dramatically simplified in order for it to be represented as a tractable mathematical problem. Many locus systems and spatial structure in particular have proven difficult to address analytically.

Computational models facilitate the implementation of the micro-level dynamics and the simulation of those dynamics over long time scales so as to measure the macro-level phenomena (Huston 1994; Huston et al. 1988). Thus, computational models help us project the micro-level into the macro. This gives us a way to test our hypotheses for the mechanisms that underlie the phenomena that we observe.

Efforts to understand species diversity dynamics have been hindered by four problems. First, the lack of a clear species definition. Second, the related operational problem of identifying species given a definition. Third, the time and spatial scales of evolution and ecology are generally too large to manipulate experimentally. Fourth, spatial structure and complex ecological interactions are important parts of the processes of speciation. These are difficult to represent and solve as systems of differential equations. The following work addresses all of these problems. I have settled on an unambiguous species definition that, given all of the information available in a model[11], can be efficiently applied to the simulated populations of organisms. Current computational constraints make the modeling of individuals on a global scale impossible. The experiments reported in the following chapters operate on simulated ecosystems at a regional level, over evolutionary, though perhaps not geological, time scales. They represent an intermediate step between the scales of population genetics and paleobiology, in a theoretical milieu.

[11] Unfortunately, the lack of complete information for real biological systems makes the reproductive species concept less appealing in practice.

1.4.1 Configuration Models

There are two basic approaches to modeling a population. One can either stratify the population into mutually exclusive classes and describe how the distributions of individuals across those classes changes over time, or one can explicitly represent each individual and the configuration of their interactions that determines their future states. Most modeling efforts in biology have fallen into the first category. Because these models are based on the distribution of individuals across a set of classes, I will follow Caswell and John (1992) in calling them "distribution models." A simple example is the Lotka-Volterra equations that describe the population dynamics over two classes, predators and prey. More sophisticated models may break down a population by age, size, sex, region, or any other parameter of significance to the dynamics in question (Caswell 1989). However, distribution models assume that all the individuals in a class experience essentially the same environment. If this is not the case, if the future state of an individual depends on the particulars of the individuals with which it interacts, then the organisms cannot be classified into homogenous groups. Each individual must be explicitly represented. Such models are typically called "individual-based models" (Judson 1997). However, this is misleading because distribution models are also directly based on individuals. Again, following Caswell and John (1992), I will call these configuration models, because their dynamics depend on the configuration of interactions between the individuals. Durrett and Levin (1994) have shown that configuration models can lead to qualitatively different predictions as compared to distribution models for the same system. Even in cases where configuration models are not strictly required by the importance of local interactions, an unreasonably large set of equivalence classes (e.g., genotypes) or the subtleties of stochastic effects on small populations may make configuration models a desirable alternative to distribution models (Caswell & John 1992).

1.4.2 Methodology

Progress is made by examining hypotheses in biology and determining whether or not they hold true in a synthetic system. The methodology follows in five steps.

1. Isolate a question.

2. Identify the minimal requirements.
3. Build and run a model from those requirements.
4. Attempt to invalidate the model
5. Either:
   (a) Elaborate the question, or
   (b) Accept the results predicated on the assumptions and validation.

First a question must be isolated and well defined, so that it can be translated into a model. In order to make this translation, the minimal components necessary for testing the hypothesis must be specified. I am using "minimal" here to mean both necessary and sufficient. However, there will be a certain looseness in the term when the researcher comes to specific decisions on how to implement the requirements. Many biological systems are constrained by multiple predictions. For example, a reasonable model of a predator-prey ecosystem should include the dynamics of predator extinction in the absence of prey, and constrained growth of the prey population in the absence of the predator. That is, predictions other than the one being explored may be used to to test the model. Chapter 3 will extensively elaborate on this approach to testing models.

Once the model has been used to collect data, the results can be treated in one of two ways. First, it is quite possible that the minimal components necessary to test the hypothesis will lead to such a simplified model, that the results will be obviously biologically implausible. This is a second form invalidation. If the results are biologically implausible, we have learned something about the hypothesis. The hypothesis must rest on further assumptions or elaborations. Once those have been identified, we can start the modeling process over again. When the model reflects a sufficiently complete hypothesis, the results of the model can be used to either reject or support the hypothesis. Thus the model acts as a test of the internal consistency of the hypothesis. It tells us when the conclusions of the hypothesis do not actually follow from the assumptions. In this way, a model may weaken a theory.

This methodology has the benefit of providing useful results regardless of their biological plausibility. We are able to refine our biological hypotheses through the process of modeling. On the one hand, the model may elucidate unspoken assumptions behind the hypothesis. The very process of constructing a model forces us to spell many of the relationships between organisms and the environment. Frequently, this endeavor illuminates large gaps in our knowledge of the system. While we may be forced to make some minimal assumptions to fill in those gaps, field and laboratory biologists can then test those assumptions. In this way, we complete the cycle between theory and experiment.

1.4.3 Caveats

There is an important caveat to be made here. While computational models solve some of the methodological problems of other approaches, they introduce their own set of difficulties and limitations. As was already noted, the construction of a configuration model often illuminates our ignorance about how aspects of a biological system work. In order to proceed we are forced to make somewhat arbitrary decisions on how to model these processes. Unfortunately, even in the cases when we have some understanding of the processes, the details of how they are implemented can influence the macro-level dynamics. Changing the data-structure that supports the system can change the results. It is thus always informative to replicate a model and look for similar results. A model may be usefully "replicated" either by developing a different implementation of the description of the model, or by developing a different abstraction of the requirements for examining the biological system.

The results of the models are not transparent, in contrast to most analytical models. The model does not produce an answer but rather a slew of data that must be analyzed statistically. In this sense, configuration models resemble experimental methodologies. This makes their results relatively difficult to generalize. We have yet to develop the mathematical tools for analysis of these models to match the tools available for distribution models (Caswell & John 1992). Typically, a hopelessly large parameter space should be explored. In the end, the results of a configuration model experiment must be predicated on the assumptions built into the model, the implementation of those assumptions, and the extent to which the dynamics have been explored. Only if you accept the assumptions as a reasonable abstraction of the biological system can you have confidence that the results bear upon the dynamics of biology.

But the methodological issues are not all gloom and doom. Configuration models represent abstractions, surely, but they generally are not as divorced from reality as traditional differential equation models. We can generally construct a direct mapping between the concepts with which experimental biologists work (e.g., organisms, genes, resources, competition, mating, etc.) and the components of the model. This both helps to ground the model in reality and make it easier to communicate, and so be criticized by, experimental biologists. This is an important advantage in a world where theoretical and experimental work increasingly diverge.

Perhaps the most important caveat is that the goal of computation modeling is not to provide Truth, but to produce good hypotheses for the experimentalists to test. This is one of the essential steps in the cycle between theory and experiment. In the experiments to follow I will seek to thin and refine the forest of theory surrounding the diversification of life. In the process I will challenge the conventional wisdom that has been shaped by the dominant methodologies and I will explore the insights generated by a configuration model.

1.5 Previous Work

The literature on microevolutionary approaches to species diversification is still nascent, and originates primarily in the field of Artificial Life, which is dominated by computer science. This is probably an indication of the fact that such models have only recently been made possible by advances in computational power as well as the fact that they tend to be relatively complex. Theoretical work in biology has generally focused on models in which species, as opposed to organisms, are the basic units. Recent work in computer science has begun to explore the dynamics of individuals interacting in microevolutionary models.

1.5.1 Microevolutionary Models

Early work on modeling diversity has focused on genotype diversity, rather than species diversity. This is a reasonable response to the difficulty of defining a species. Bedau et al. considered the effects of selection and mutation rate on genotypic diversity (Bedau et al. 1992). They found that the absence of selection and high mutation rates both led to higher genotypic diversity relative to the effects of selection and low mutation rates.

Saruwatari et al. (1994) also set up an artificial ecosystem. Each organism had an 8-bit reproduction gene. An organism was allowed to mate with any other organism whose reproduction gene was no more than 2 bits different. However, they did not use these reproductive barriers to define species[12]. Instead, they examined genotypic diversity in these

[12]If they had, they would have found the reproductive genotype space severely constraining. There are

reproductive genes. They found that this genotypic diversity remained relatively constant over time. There was no spatial structure in their model. The implementation of predation is interesting. In occasional, biologically implausible, events, a new predator was created by fusing two genomes into one. Almost like polyploidy in plants, this instantaneously doubled the size of the organism's genome. Though they did not state it, we must assume that mating still depended on only 8 bits, otherwise the new mutant would have had no mates. Along with the reproduction gene, every organism had a prey gene. Predation depended on the match of the predator's prey gene and a prey organism's reproductive gene. Organisms with more than one prey gene, due to past fusion event, thus had multiple prey genes and so could consume a greater variety of prey. The longer the genome, the greater the generalism of the predator. Furthermore, organisms were constrained to only eat other organisms that had a shorter genome. A predator deducted energy from a prey in an amount equal to the proportion of bits matched in the best match between the prey's reproductive gene and one of the predator's prey genes. So a predator with many prey genes was more likely to find a good match against any given prey's reproductive gene relative to a predator with few prey genes. However, there was a trade-off between generalism and specialism. The longer the genome, the greater the energy tax per time step. While an ecosystem emerged from these dynamics, in every case it crashed. The predators drove the prey to extinction.

Herráiz et al. (1997) defined species by dietary similarity. This is essentially a niche definition of species. Reproduction of the simulated organisms in their model is asexual, so their diversity measure is fundamentally a measurement of genotypic diversity at the dietary loci. They found that the abundances of the different species in their model matched natural systems in displaying a log-normal distribution.

Recently, John Holland's Echo model (Holland 1992; 1993) (designed to "echo emergence") has generated considerable interest. This model inspired some of the ideas in my approach. Specifically, the ecological interactions of Echo are based on the matching of pattern genes between organisms. Sexual reproduction is also based on matching pattern genes, although all organisms may also reproduce asexually in Echo. Echo includes addi-

only 256 possible reproductive genotypes. Every genotype can mate with 36 others. As Section 4.3 will explain, this means that a mere 8 different random genotypes is sufficient to fill the space, such that all the organisms will form a single gene pool.

tional dynamics for trading resources, which can be used to model simplified economies and a form of mutualism. It has a variety of resources[13]. The original model did not include the transformation of resources from one type into another type. However, derivative models have added this complexity (Schmitz & Booth 1997). Without rules for the transformation of resources there was no loss of energy in the process of changing prey matter into predator matter. This allows an arbitrarily large number of trophic levels in the ecosystem.

Echo is not a model specifically designed for investigating speciation. Hraber and Milne examined diversity dynamics in the model, but they argued that the frequency of asexual reproduction rendered the traditional reproductive isolation criterion for species impractical (Hraber & Milne 1997). Yet, the presence of asexual reproduction in a sexual population does not change the fact that the barriers to sexual reproduction define a gene pool from which all descendants of that "species" will be drawn. Thus, the reproductive species concept could also be applied to Echo. Exactly how to do this in a model of an ecosystem will be described in Section 2.2.3. In any case, Hraber and Milne look at genotype diversity under the presence or absence of selection and varying mutation rates. Mirroring Bedau et al.'s (1992) results, they found that genotypic diversity was greatest under high mutation rates or in the absence of selection on the genome. More interestingly, they found that pairs of genotypes tended to evolve under selection in which each genotype would mate and trade resources only with itself. Furthermore, neither genotype would prey upon the other. This led to higher population sizes ("productivity") than the populations in the absence of selection. One weakness of this study was the fact that the organisms were constrained to 7-bit genomes, and so genome space was extremely constricted.

Jones et al. (1997) also examined genome diversity. They found that the Echo model did indeed demonstrate the power law of the species area curve (see Section 4.2) but that the exponent (z) was unrealistically large. This could be due to any number of abstractions in the model, but I suspect it can be explained in part by the attempt to substitute genome diversity for species diversity.

Gecko (Booth 1997), while originally derived from Echo has diverged dramatically, shifting to a continuous spatial structure. Gecko is an ecological, not an evolutionary model. It does not include a dynamic of mutation, and so cannot be used to examine speciation. How-

[13] This is in contrast to the simple energy or "meal" units that I use, described in Chapter 2.

ever, it has passed the "trophic cascade" test[14] as an ecological model and shows promise for future research. For a review of ecological configuration models of plant community dynamics, see Huston (1994).

Lindgren and Nordahl (1994) also use the matching of tags to establish an artificial ecosystem, except that the outcome of an interaction depends on the two organisms playing an iterated prisoner's dilemma (IPD) game[15] The tags are only used to select an opponent. They found that productivity was concentrated in the lowest levels of the emergent food web. This matches well with reality. They also found a correspondence with reality in that complexity in the resource base led to complexity in the structure of the food webs.

Liou and Price (1993) examined an issue closely related to diversity, speciation. In this case, they looked at the probability of speciation due to reinforcement of mating preferences by natural selection. They chose a configuration model approach because the theory of speciation by reinforcement calls for a complexity that makes a distribution model intractable. Specifically, they wanted to include sexual selection, natural selection against hybrids, density dependent population regulation and varying levels of gene flow between populations. They represented individuals with 10 diploid loci, coding for mating preferences and the preferred traits, along with tags to indicate an organism's population. Hybrids were randomly assigned to one of the two populations. The question was, under what conditions would the mating preferences of two populations diverge? If both populations survived and became fixed[16] for different mating preferences along with their preferred traits, then

[14] A model of an ecosystem should demonstrate a trophic cascade. A trophic cascade is a change in plant or autotroph biomass in response to predation. In the absence of herbivores, plant biomass ought to be relatively high. In contrast, the introduction of herbivores tends to dramatically reduce plant biomass. However, the further introduction of carnivores to prey upon the herbivores tends to restore plant biomass to some intermediate level. Trophic cascades are examined in Section 3.2.

[15] The prisoner's dilemma game is a two player game in which the players have a choice of either "cooperating" or "defecting." It is based on the idea of two criminals individually being offered a pardon if they testify against their partner. If one betrays the other, or "defects," and testifies against her partner, she will go free and her partner will get a stiff sentence. If they both defect, they will be thrown in prison but their sentences may be reduced in recognition of their attempt to assist the prosecution. However, if both keep their silence, if they "cooperate" with each other, then the prosecution may not be able to prove its case and so may be forced to plea bargain for a short jail term. This is translated into a game by scoring the players by the combination of their actions. In the prisoner's dilemma game, it is "rational" for each player to defect. Regardless of what their opponent does, they will be better off defecting. The paradox is that mutual defection scores lower than mutual cooperation. When the game is iterated, in the IPD game, it becomes "rational" to cooperate, as long as there is a strong possibility of facing the same opponent again in the future. Axelrod (1984) has written a fascinating discussion of this subject and its implications for all manner of interactions.

[16] Fixation at a locus means that an entire population has the same allele at that locus. Barring mutation and migration, this means an end to evolution at that locus.

the populations were said to have speciated. Liou and Price found that populations would only diverge if they were initially quite different in their mating preferences and if the hybrids suffered a significant selective disadvantage. Their model did not include mutation, so evolutionary change occured strictly through natural and sexual selection. They also avoided the effects of competition by imposing population regulation independently in the two populations. Natural selection was extrinsic in the model. It was not a function of the organism's genes. They even examined the effect of distributing the populations amongst patches, and found that this dampened the effects of reinforcement. Divergence was most intense when the two populations were sympatric.

In computer science, the field of genetic algorithms attempts to harness evolution to generate good solutions for a variety of problems. Starting from a random population of potential solutions, usually represented by bit strings, each solution is assigned a fitness value. Both survival and mating dynamics are based on those values. Two solutions "mate" by recombining portions of their bit strings to produce a new potential solution. Deb and Goldberg (1989) have added mating restrictions to a genetic algorithm to enhance its ability to maintain a diversity of solutions in the population. Only similar solutions were allowed to mate. This tended to maintain subpopulations, or "species"[17] on many different fitness peaks in the adaptive landscape.

1.5.2 Macroevolutionary Models

Most theoretical models of diversity developed by biologists work on species as individual components in the model. Thus, they cannot examine the mechanisms underlying speciation, but rather assume some probability of a speciation event, along with a probability that a species will go extinct during a time step. In this way a branching process is established which generates an artificial phylogeny.

Raup et al. began this line of research by examining the shapes of the curves describing the numbers of lineages in a clade over time (Raup et al. 1973). They called this the "clade shape." Their branching process produced a variety of clade shapes. They point out that paleontologists seeing such patterns in real fossil data would tend to assume that the clades

[17] Bull and Fogarty (1996) also use the metaphor of species in their genetic algorithm. However, they do not utilize mating barriers but instead "geographically" subdivide, mix, and re-subdivide the population. Thus, the species metaphor is something of a stretch.

were shaped by different characters that influence speciation and extinction so as to produce the shapes. In contrast, this model suggests that clade shape differences may be generated by random fluctuations in a single stochastic mechanism. Thus, the simulation provides an important null model for paleontology.

Future work focused on the relationship between species diversity and diversity at higher taxonomic levels (Valentine & Walker 1986; Sepkoski & Kendrick 1993). In this case, the species of the branching process are probabilistically classified into "families" and diversity dynamics at the family level are compared to diversity dynamics at the species level. In general, there is less of a correlation than had been expected. This highlights an important weakness in the typical paleontologic practice of counting families over time as a proxy for measuring the underlying species diversity. Sepkoski and Kendrick went further, to explore how different classification systems affect the accuracy of diversity measurements. Their conclusion that paraphyletic classifications[18] are just as good as monophyletic classifications[19] are widely cited and vehemently debated (Robeck et al. 1998).

Valentine and Walker extended the branching process model by constraining species to occupying locations on a two-dimensional grid (Valentine & Walker 1986). The grid was a metaphor for "niche space." The fixed probability for speciation was modified by the requirement that a randomly chosen location on the grid, within some radius of the ancestral species, must not already be occupied by a species. They could then implement "selective mass extinctions" which would wipe out the species in some contiguous block of niche space as well as "random mass extinctions" which would pepper the niche space with extinctions. They found that family diversity responded differently to the different forms of mass extinction, with the selective extinctions killing off a higher proportion of families. More importantly, they found that the plateaus in family diversity curves did not tend to represent a plateau on the underlying species diversities. This casts doubt on the standard interpretation of a similar plateau in marine animal family diversities during the Phanerozoic[20] as a form of equilibrium.

Walker (1985) extended this work by considering the diversification of a monophyletic

[18] A paraphyletic group is one which includes an ancestral species but not all of its descendants

[19] A monophyletic group consists of a complete subtree of the phylogeny. That is, an ancestor and all of its descendants.

[20] The Phanerozoic eon is the name for period of time since the Cambrian explosion, about 545 million years ago

group in a continuous niche space. The assumption was that, while each species occupied an arbitrarily shaped niche, the taxon as a whole expands, roughly, as an n-dimensional sphere. Walker makes the further assumption that the center of the sphere would quickly fill, leaving only the species at the surface with the freedom to speciate. This would lead to a polynomial increase (with degree n) in diversity, in contrast to the exponential and logarithmic models. However, the assumption that the center of the sphere would be filled collapses if adaptive space is a high-dimensional space. Walker argues that the correlations amongst ecological variables reduces adaptive space to only a few dimensions (Walker 1985).

Flessa and Thomas (1985) explored a model of the geographical ranges of genera as a function of three dynamics: local extinctions in regions, expansions into new regions, and speciations. They found that the model only matched the data on geographical ranges in real genera if the probability of expanding into a new region increased with the number of regions a genus occupied. This positive feedback dynamic is necessary to produce the skewed distribution of many genera with small ranges and a few genera with very large regions.

Perhaps the most intriguing work is that of Gavrilets and Gravner (1997) because of their attempt to bridge the gap between genotypes and species. They simplify Kaufmann's (1993) rugged adaptive landscapes so that genotypes have either 1 (viable) or 0 (inviable) fitness. Gavrilets and Gravner call this a "holey adaptive landscape" (Gavrilets 1997; Gavrilets & Gravner 1997). In their n-loci model, each locus can have one of three states: the two homozygous states and the heterozygous state. The binary fitnesses are assigned randomly to each of the 3^n genotypes. They then use percolation theory (Grimmett 1989, see Section 4.3) to analyze the connected components of viable genotypes in the landscape. They define a species as a single entirely homozygous genotype, to make the analysis easier. Thus, many "species" can exist within a single connected component. This implies that a population can evolve through individually viable mutations from one state into a new state that is reproductively isolated from the first state.

1.6 Harbingers

Following the methodology of Section 1.4.2, we have now isolated an hypothesis (actually, a set of hypotheses) for the diversification of life. Our task is to test whether or not diversity is regulated, and if it is, to determine the relative importance of the proposed biological and non-biological factors for the generation and maintenance of diversity. The next chapter will implement the following two steps of the methodology by identifying the minimal requirements for testing those hypotheses and describing a model that fulfills those requirements. The model is described in sufficient detail that it should be independently implementable by other researchers. Chapter 3 attempts to invalidate the model through tests of its ecological and evolutionary dynamics. Specifically, I will show that the model demonstrates reasonable predatory-prey population dynamics over at least three trophic levels (plants, herbivores and carnivores), competitive exclusion, and adaptation. Chapter 4 explores the first two levels of answers to the question of diversification. It delves into the basic processes of speciation and extinction. The question of the regulation of diversity emerges in the context of an experiment on the species-area curve. Chapter 4 shows how the basic dynamics of island biogeography change in the context of evolutionary ecology when originations of species are based on speciation, not colonization events. I will also begin to develop the theory for understanding speciation dynamics under the reproductive definition of species.

The remaining three experimental chapters examine the third level of answer to the problem of diversification, factors that may change diversification rates and so alter the regulation of diversity. Chapter 5 compares the abiotic factors that have been proposed to affect diversity. These include geographical and habitat heterogeneity in both space and time, as well as factors that increase the productivity levels of organisms at the base of the food webs. Chapters 6 and 7 concern biotic factors in the diversification of life. Chapter 6 investigates the validity of changes in the utilization of adaptive space as causal explanations for changes in diversity levels. Chapter 7 examines a dynamic that is largely invisible to paleobiology, the effects of mating preferences on diversity. Finally, a comparison of all the factors can be found in the concluding chapter along with the significant results that have cropped up along the way. If you just want the answers, this is the chapter to read. You will be glad to note that most of the derivations of various details have been relegated to the appendices.

And now I shall get on with it.

Chapter 2

The Model

A model is a tool for theoretical exploration as well as a guide for experimental work (Wimsatt 1987). It is an active form of a theory about how nature works. I argued in Section 1.4.2 that the application of configuration models to biological questions requires us to first specify the components of a model that are required in order to examine a hypothesis. Then we may construct a model so as to include only those components that are necessary to implement the hypothesis. This is the principle of parsimony as applied to modeling. We may argue later about the importance of missing complexities in the model, in light of the results. And, depending on the outcome of those arguments, we may want to elaborate the model. However, for now, I will try to keep the model as simple as possible.

2.1 Requirements

The purpose of the model is to examine the macroevolutionary dynamics of diversification through the microevolutionary dynamics of organisms. What aspects of biology must be represented in the model?

2.1.1 Microevolution

Requirement 1 *Organisms.* The components of evolution on an organismic and genic level are well understood. For natural selection to work, you need something to select. In this case, we will focus on organisms.

Requirement 2 *Reproduction.* Natural selection is based on differential reproduction. Some organisms must reproduce more than others. This may be caused by differential survival, but it need not be. Natural selection may still occur if no organisms die, as long as some produce more offspring than others.

Requirement 3 *Heredity.* If there is no relationship between a parent's differential reproduction and its offspring's reproduction, if parents cannot pass on the secrets of their success to their children, then natural selection will not be able to *select* anything.

In biology, the correlation between a parent's success and its offspring's success is mediated through the gift of genes. These genes must have some influence over the differences between organisms' reproductive success. Organisms (or any unit of selection), differential reproduction, and heredity are all that are really needed for natural selection to kick in. It is kind of surprising that "death" has not yet made the list. Modern biologists tend to define evolution as the change in proportions of alleles, or different genes, in a population. Given a potentially infinite population, differential reproduction is sufficient to produce changes in the proportions of alleles in the population.

Requirement 4 *Mutation.* Besides altering the proportion of alleles in a population, mutation is the engine of change in evolution. It is the ultimate source of novelty, and as such, it is a necessary ingredient for testing hypotheses that focus on evolutionary innovation as the driving force behind diversification (Benton 1990).

All the previous requirements establish a framework simply to allow evolution to play out in the model. However, a number of the hypotheses for diversification make explicit reference to ecological interactions (Vermeij 1987; Van Valen 1984).

Requirement 5 *Competition.* Competition is the darling of most diversity arguments (Raup 1991; Flessa et al. 1986). It is thought to be the causal agent behind most extinctions at high diversities (Ricklefs & Schluter 1993; Cornell 1993). Competition also defines the boundaries of viable niches that colonizers may hope to fill. Competition is a sort of "horizontal" interaction within a single trophic level of a food web.

Requirement 6 *Predation.* Predation characterizes most interactions between different trophic levels of a food web. Predation is probably also important in shaping diversity trends (Vermeij 1987). Most discussions of diversification revolves around the concept of niche space (Benton 1990; Valentine & Walker 1986; Walker 1985, see Chapter 6 for a full discussion). If we adopt a simplified view of niches as the set of resources upon which a species depends, then the inclusion of predation, the consumption of resources, introduces niches to the model. However, the boundaries of a niche, the constraints on predation, are central to most of these discussions.

Requirement 7 *Specialization.* Some organisms specialize on a small number of food sources. Koala bears (*Phascolarctos cinerus*) are notorious for subsisting on only the leaves of Eucalyptus trees (genus *Eucalyptus*). Other organisms are generalists and can eat a wide variety of foods. Humans (*Homo sapiens sapiens*) are notorious for being willing and able to eat just about anything. After evolutionary innovations that open up new niches for colonization, Benton (1990) argues that specialization has been the second most important factor in diversification.

2.1.2 Macroevolution

Requirement 8 *Species.* We are fundamentally interested in macroevolutionary dynamics. The dynamics of species, speciation, and extinction. This requires that we be able to group organisms into species. If we are to use the reproductive concept of species, then two further requirements follow.

Requirement 9 *Sexual reproduction.* The reproductive species concept only applies to organisms that reproduce sexually.

Requirement 10 *Reproductive barriers.* Furthermore, the characteristics that prevent individuals from successfully mating must themselves evolve. These are called reproductive barriers. Putting this together with the microevolutionary dynamics, the evolution of reproductive barriers implies that there must be some genes that control aspects of those barriers.

Requirement 11 *Speciation.* We are not simply interested in species. We are interested in the creation of new species by the splitting of previous ones. So there must be some representation of speciation.

Requirement 12 *Death.* If we assume that parents and their offspring will be similar enough that they will not be separated by the reproductive barriers, then it is impossible for a gene pool to split apart without death.

It is also nice (as a model designer) to include death in a model because computers have difficulty representing infinite populations. Once a finite population capacity is filled, all evolutionary change will stop without death. This is not to say that death is adaptive, or evolved for the good of the population. It is simply to say that a model would not be very interesting if it froze once the population had hit the carrying capacity. These design issues aside, death is necessary to split apart gene pools and so allow speciation.

Requirement 13 *Geography.* The best supported theory for speciation suggests that geographic isolation is an essential factor in the process. We must therefore represent geography in order to examine allopatric speciation.

Requirement 14 *Migration.* Geography allows small populations to be physically isolated from other organisms. This isolation typically begins with the migration of a small number of colonists to a new location. Migration has the effect of connecting geographically separated gene pools and so plays an important role in speciation.

Requirement 15 *Habitat.* Most of the abiotic hypotheses for diversification are based on fragmentation of populations by the geography as well as differential selective pressures on the subpopulations due to their exposure to different climates or habitats (Vermeij 1987; Cracraft 1985; Benton 1990). The only missing ingredient left, then, is habitat.

2.2 An Instantiation

There are many potential instantiations of the requirements to model diversification. Furthermore, the details of an instantiation may have dramatic effects upon the results of a model. The results of a single model cannot be claimed to be the final judgment on the validity of a hypothesis. Rather, robust results from models must come through the agreement of parallel and independent efforts with alternative implementations of the critical details. The implementation that follows, dubbed “a model of diversification” or MoD, may not be the cleanest or the best. However, it draws on the common techniques in artificial life for representing individuals, their interactions, and evolution (Ray 1992; Holland 1992;

Maley 1997). All italicized terms below are parameters to the model. Their experimental settings are generally parenthesized.

Both time and space are discrete in the model. Time moves in "time steps." During each time step, all the organisms have a chance to eat, die, and possibly mate. The details follow.

2.2.1 Microevolution

Microevolution rests on the requirements of implementing organisms, reproduction, heredity and mutation. An organism is represented by: a set of 4 chromosomes, a tally of the number of other organisms it has managed to eat since the last time it reproduced, and a tally of the number of predators that have attempted to eat it during the current time step. All of these aspects have an ecological function which leads to the differential reproduction of the organisms, described in Section 2.2.2. Each "chromosome" is represented as a string of 32 bits, or "loci" in the language of genetics. The loci of a chromosome[1] represent the set of characters that influence a particular aspect of the life history of the organism. Rather than representing a number, these bits are just treated as a pattern of 1's and 0's. The organisms are haploid, in that they only have one copy of each chromosome, not two, like most animals.

Once an organism has managed to collect a number of meals equal to the *energy-conversion* (3) parameter, it garners a chance to reproduce[2]. Setting *energy-conversion* to 3 enforces a 3 to 1 conversion of prey matter into predator matter. After locating a mate[3] in the organism's patch, the chromosomes for their offspring are constructed by recombination from the chromosomes of the parents. I have used two-point crossover to implement recombination, as shown in Figure 2-1. For each pair of chromosomes from the two parents, copies of the bit patterns are lined up and then two points are randomly chosen along the sequence of bits. The subsequence between the two points is exchanged between the two parental chromosomes, and then one of the two new "mosaic" results is randomly chosen to be the chromosome for the offspring. This is repeated for each of the four chromosomes.

Choosing to cut the chromosomes at two points avoids a bias inherent in one-point

[1] Most artificial life models would call these chromosomes "genes."
[2] Eating another organism provides the diner with one meal or energy unit.
[3] The mate need not have collected 3 meals or energy units.

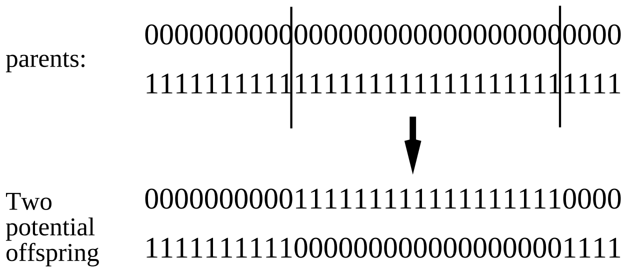


Figure 2-1: Two-point crossover.

crossover. Under one-point crossover, when the chromosomes are only cut once, the bits at either end of the chromosome are almost certainly going to be separated by recombination. However, two-point crossover, when implemented correctly, is equivalent to forming the bit strings into circles and then exchanging equivalent sections of the circles between parental chromosomes, as depicted in Figure 2-2. The chance that two bits will be separated by recombination is proportional to their distance from one another on the circle. All bit positions are thus treated equivalently.

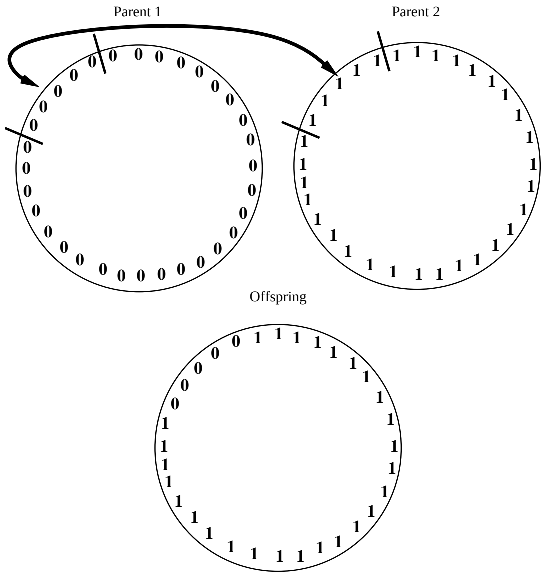


Figure 2-2: Two-point crossover on strings is equivalent to recombination between strings formed into circles. This makes it clear that there are no biases in the crossover based on the position of a bit in the string.

After we have determined the chromosomes of the new organism through recombination of the parental chromosomes, the new chromosomes are exposed to point mutations. A point mutation is simply the flipping of a single bit in a chromosome. The *mutation-rate* parameter (0.02) sets the average rate of bits flipped per chromosome per reproductive event. The number of bits to be flipped was determined by a Poisson process[4] Then a particular bit was chosen to be flipped by generating a random number[5]. between 1 and 32 with equal probability.

2.2.2 Ecology

The required ecological interactions include competition, predation, and the elaboration of predation by specialization or generalism. In other words, the interactions necessary to set up a basic food web.

The ecological interactions were implemented through 3 of the 4 "chromosomes." "Chromosomes" are metaphors here for phenotypic characters that are perfectly heritable. Thus, there is no distinction between the genotype and phenotype of an organism. This avoids the complexity, and our vast ignorance, of the process of development by which an organisms genes and the cell contents of the zygote create the full blown, multicellular organism. While this is a blatant simplification of the processes in evolution, the hypotheses for diversification do not make particular reference to the genotype-phenotype distinction. So, until the time when validation of the results may require us to include it, we are obligated to elide[6] the process of development. I will alternatively refer to a "chromosome" or a set of "characters" as equivalent aspects of the model.

Organisms have chromosomes for "predation resistance" characters, a preferred "prey template," and "generalism." These correspond to the three genes that Lindgren and Nordahl (1994) use to set up their ecosystem. An organism's prey template bit pattern encodes the bit pattern of the predation resistance chromosome in the prey organisms that it prefers to eat. On top of this, the 1's in an organism's generalism chromosome specify which bit

[4]Poisson processes are a form of stochastic processes that are typically used to model the random occurrence of discrete events over time. If the average rate of events is λ per time period, then the probability of observing k events over t time periods of a Poisson process is

$$p(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \tag{2.1}$$

We can generate a random number with a poisson distribution by first generating a random number (the probability) between 0 and 1 with uniform probability. Then we repeatedly subtract $p(k)$ for $k = 0, 1, 2, \ldots$ until the result is negative. The value of k when the result first becomes negative is the poisson random number.

[5]I have used a modified version of Knuth's (1981, pp. 171–172) subtractive method for generating pseudo-random numbers coded by Terry Jones.

[6]In computer science slang, elide = "punt," as in, "I will punt on the issue of biological development."

positions do not matter in its prey template chromosome when it tries to consume prey. In other words, the generalism chromosome creates "wild-card" positions in the organism's prey template chromosome. I will refer to the predator's prey template chromosome, modified by wild-cards from the predator's generalism chromosome, as a "search pattern." The predation resistance chromosome represents characters relevant to the sufferance of predation. This might include both behavioral and structural characteristics of the organism that help it to avoid predation. Figure 2-3 diagrams an 8-bit example (the real chromosomes are 32 bits) of a match between a predator and a prey organism as well as a mismatch that prevents the predator from capturing and consuming the prey.

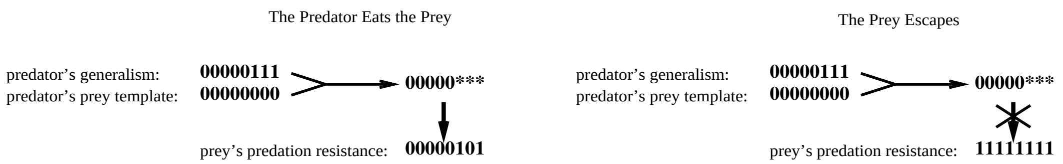


Figure 2-3: Determination of a predator-prey interaction. The predator's generalism and prey template chromosomes are combined to form a search pattern. This search pattern, with wild-card positions, is matched against prey organisms' predation resistance chromosomes. If all of the non-wild-card positions in the search pattern match the prey's resistance chromosome, then the prey may be consumed. Otherwise the prey may escape.

Notice that an organism with all 1's, all wild-cards, in its generalism chromosome, is an extreme generalist. It can eat any organism. At the other extreme, an organism with all 0's in it generalism chromosome is a Koala bear-style specialist. It can only eat prey that exactly match its prey template.

So far, I have described predation implying that only a perfect match (0 mismatching bits) between a predator's search pattern and a prey's predation resistance chromosome allows the predator to consume the prey. The reality of the algorithm is not so simple. It is not sufficient to model predation as a deterministic function of a perfect match between predator and prey. First of all, this "all or nothing" type of interaction creates sharp cliffs in fitness space, such that a single bit mutation is likely to change a well adapted lineage into a lineage with a search pattern that fails to match any of the potential prey organisms. When a fitness landscape has sharp drops, or is "rugged" in Kauffman's (1993) terminology, mutation with natural selection cannot easily herd a population up a fitness peak. Evolution is much more effective on smoother fitness landscapes. But more to the point, the model does not behave reasonably with such a simplistic predation function (see Sections 3.1 and

3.2).

Predation in the model is a probabilistic function. The chance that a predator captures and consumes a prey organism is a function of the number of bits that do not match between the predator's search pattern and the prey's predation resistance chromosome. It is also a function of the *prey-location* probability parameter (0.15), the number of times the prey organism has been pursued by predators during the current time step, and a *predation-distribution* parameter (2.0). The intuition and motivation for including these additional factors is discussed in Sections 3.1 and 3.2. In sum, the probability that a predator successfully captures a prey is

$$Pr[\text{prey capture}] = le^{-(m+nd)} \tag{2.2}$$

where l is the *prey-location* probability, m is the number of mismatching bits between the predator's search pattern and the prey's predation resistance chromosome, n is the number of times that prey organism has escaped predation this time step, and d is the *predation-distribution* parameter. The idea here is that the chance of a predator capturing the prey should decay with the increasing mismatch between the predator's search pattern and the prey's predator resistance chromosome. However, a probability of 1 when there is a perfect match is too big. So we scale it by l, the *prey-location* probability. Finally, as we will see in Section 3.2, competitive interference had to be added to MoD in order to elicit realistic predation dynamics. This is added through nd.

A random number between 0 and 1 is generated, and if it is below the probability of Equation 2.2 then the predator kills the prey organism and increments its record of the number of prey it has eaten. If this is the third prey it has eaten since its last mating attempt, the predator is granted a new mating attempt. However, even if it fails to reproduce, it will have to capture another three prey organisms before it is allowed to mate again. I have tied mating to survival for efficiency reasons. There is a general trade-off between modeling the complexity of an individual's life and modeling long time scales (with population size as a third constraint). By keeping the behavior of the organisms simple, we are able to run the model for more generations and so paint a better picture of evolution.

The introduction of predation would seem to establish a food web. However, as presented so far, the food web lacks grounding. What prey do the organisms on the lowest level of the food web eat? Every food web needs a foundation of organisms that can translate non-biological matter into biological matter. These are the autotrophs. The autotrophs with which we are most familiar are the plants. For simplicity I will refer to autotrophs as plants and the heterotrophs (organisms that eat other organisms) as animals. Every organism carries an additional flag indicating if it is a plant or an animal. These flags do not mutate. Only plants are allowed to mate as well as compete for space with other plants. The same is true for animals. Instead of matching their search pattern against a prey's predation resistance chromosome, a plant matches against a bit pattern in their environment that represents the habitat. The chance that a plant absorbs an energy unit, or "meal," from the environment is

$$Pr[\text{absorbing an energy unit}] = h \times \textit{energy input} \tag{2.3}$$

where h is the fraction of bits that the plant's search pattern matches with its habitat and *energy input* is a parameter (1.0) of the model so that I may pour more or less energy into the ecosystem. If this probability is greater than 1, the plant is granted energy units equal to the integer part of that value plus another unit with probability equal to the fraction. So if $h\times$ *energy input* $= 1.5$, then the plant would have a 50% chance of garnering 2 energy units and a 50% chance of absorbing only 1 energy unit.

Given this implementation of predation, one might expect all lineages to evolve towards greater generalism. All of the previous artificial ecosystems that implement generalism have used a trade-off between generalism and metabolism, or energy consumption (Lindgren & Nordahl 1994; Saruwatari et al. 1994; Herráiz et al. 1997). The fact that we see specialists in biology seems to imply that there is some trade-off such that specialists might be favored over generalists by natural selection under some conditions. In fact, there is very little evidence pointing to such a trade off. A competing hypothesis postulates that specialists evolve from generalists through random drift. Genes for utilizing resources that aren't currently present are not under selective pressure for maintenance and so may be destroyed by mutations. In other words, specialists may have no advantage over generalists. There is some evidence for this (Thompson 1994). The real story is bound to be a lot more complicated. Thompson (1994) takes a middle of the road approach to the problem arguing that any given lineage will have a collection of three different types of genes: (1) Genes that help utilize a resource but are independent of genes for utilizing other resources. (2) Genes that help to utilize more than one resource. (3) Genes that help utilize a resource at the expense of interfering with the utilization of other resources. Only the final form of gene results in a tradeoff between specialism and generalism. Since the jury is still out on this issue, I have made the simplest assumption, that there is no tradeoff between specialism and generalism.

The only remaining issue in implementing the microevolutionary requirements of MoD is the realization of competition. A plant with a search pattern that is a better match to the habitat will collect more energy than its competitors with poor matches to the habitat. It will thus reproduce more quickly, passing on those "genes" for the match to the habitat, and eventually come to dominate the population. Similarly, plants with predation resistance chromosomes that do not match any of the search templates of the predators, along with predators whose search templates match with many of the available prey, will be more likely to reproduce than the poor sods who succumb to predation or starvation.

2.2.3 Macroevolution

The purpose of this model is to examine macroevolutionary dynamics through the modeling of the microevolutionary interactions. It is thus appropriate that the heart of MoD has already been described in the previous section on microevolution. There are only two remaining phenomena that have not been explained: regional level issues, like geography, and species, with which we are primarily concerned.

Geography and Migration

As was mentioned before, space is discrete. However, MoD does not have the fine grained spatial scale of cellular automata. Rather, space is organized into "patches." There is little spatial structure within a patch. Each patch can hold up to 2K (2048) organisms[7]. Plants are limited to only fill half of a patch, otherwise they might take over the entire patch and so compete with animals for space. The 16 patches are organized into a 4 by 4 grid. The ecological interactions are constrained to organisms within a single patch. An organism may

[7]The organisms are stored in an array. In addition, free spaces in the array are wired up into a linked list stack so that finding a free space is an $O(1)$ time operation. This means that initially, new organisms are placed into the patch in consecutive positions of the array. However, as the pattern of deaths in the array becomes random, the placement of new organisms will also become random.

only mate[8] with and eat[9] other organisms within its patch. During each time step every organism in a patch is given a chance to eat[10]. Failing this, it starves to death. However, if it succeeds in finding food, and it has eaten thrice since its last mating attempt, it will try to find a viable mate. An organism may be eaten before it gets its chance to mate. However, new organisms produced in a time step are held in a separate data structure until the end of the time step. This prevents them from being born and acting or being eaten in the same time step. In this way, predation occurs before prey competition for space, and so reduces the competitive pressures amongst the prey (Crawley 1992).

Migration, or movement between patches only happens at the beginning of an organism's life. When a new organism is born, it takes a random walk across the environment to locate the patch in which it will spend the remainder of its life. The random walk is 0–2 steps long[11]. For each step of the walk a random number between 1 and 9 is generated. These numbers correspond to the patch where the organism is currently located along with the eight potentially neighboring patches. If the random number dictates a step off the edge of the 4 by 4 environment, the organism is placed in the border patch nearest to its intended destination. These random walks roughly model the dispersal of real organisms. The chance of finding an organism at a distance d from its point of origin is approximately d^{-2}.

The theory of allopatric speciation does not just require spatial structure, it requires "geography." This includes barriers to migration. The canonical example is a mountain range that separates two populations by making migration between them exceedingly rare. This is modeled by modifying the random walk algorithm. Every patch in the model is associated with a "migratory barrier" between 0 and 1. For every step of the random walk in an organism's migration, a random number between 0 and 1 is generated. If that number is below the migratory barrier of the patch that the organism is trying to enter, it fails to enter that patch and must remain in its current patch for that step of the random walk. If there are more steps to follow in the walk, it may end up trying again, and perhaps even

[8]Potential mates are found by starting at a random position in the array and scanning the entire population until either a compatible mate is found or the population in that patch has been exhausted. When the end of the array is hit, the search continues from the beginning of the array.

[9]Similar to mating, a prey organism is located by starting at a random position in the array and then scanning the next *search-radius* (512) positions. This continues until either a prey organism is consumed, or 512 positions in the array have been checked.

[10]The live organisms are also kept in a doubly linked list so as to not waste time processing empty positions in the array of organisms. The double linkage is useful since an organism in the middle of the list may die and have to splice itself out of the list. New organisms are inserted at the front of the list.

[11]The length of the random walk is determined randomly with uniform probability.

succeeding the next time. Thus, if all the patches have 0.9 migratory barriers, then there is only a 10% chance that an organism will be able to enter a neighboring patch on a step of its random walk. As was mentioned above, habitats are modeled by associating a bit pattern with each patch representing the relevant climate or habitat characteristics for the plants.

Species and Speciation

A species is a reproductively isolated gene pool. If two organisms can mate and produce viable offspring, then they are, by our definition, part of the same species. This is a non-transitive relationship. Organism A may be able to reproduce with organism B, and organism B may be able to reproduce with organism C, but it sometimes happens that organism A cannot reproduce with organism C. Yet all three are part of the same species. It is still reasonable to say that organisms A and C are part of the same species, despite their inability to mate, because their genes can still come together in a grandchild or later descendant. They are part of the same gene pool.

Reproductive barriers are implemented by giving the organisms a fourth chromosome, twice the length of the others, called the "reproduction" chromosome. The 64 bits (loci) of the reproduction chromosome[12] are assumed to represent the set of characters influencing reproduction such as mating season, reproductive morphology, courtship behavior, etc. A pair of organisms may only mate if their reproductive genotypes differ in at most one bit. That is, potential mates must have reproductive genotypes within Hamming distance one of their mate. As with earthworms, there are no male and female "sexes" per se. If there were only a few bits in the reproductive chromosome, then the probability of two lineages randomly evolving to within a single bit would be relatively high. Conventional wisdom in biology assumes that it is highly unusual for two species to coalesce into one.

Using a Hamming distance restriction on mating mimics the non-transitivity observed in nature. Sometimes individuals of a species may be well adapted to their environment across a gradient of resources in the environment. It occasionally happens that while individuals of the species may mate with neighboring individuals, individuals from more distant

[12]The chromosome is represented by two 32-bit integers. Mutation is applied to each of the 32-bit halves of the chromosome in turn, so the reproductive chromosome has an average of 0.04 bits flipped per new organism. The mutation rate was set so as to generate an expected total of one bit flip for every 10 new organisms (with 5, 32-bit chromosomes in each organism).

populations do not produce viable offspring (Oliver 1972; Fox 1951). As both Orr (1995) and Gavrilets (1997; 1997) point out, the accumulation of mating incompatibilities is necessarily a gradual process. If a dramatic mutation reproductively isolates and organism, it will likely have the same fate as Lonesome George[13]. With only one living member of an isolated gene pool, a species is doomed.

Orr (1995) has analyzed the dynamics of speciation due to genetic incompatibilities in the genomes of the parents. He found that speciation should progress at approximately the same rate regardless of whether the incompatibility is based on a pair of loci that have a dramatic effect or on many loci, each having a small effect on reproduction. There is thus reason to believe that the dynamics of the 1-bit restriction on mating should be similar to the dynamics of allowing mating with multiple differences in their reproductive chromosomes. However, allowing matings between organisms with multiple loci differences would probably require proportionally longer chromosomes to avoid the accidental coalescence of two species.

Notice that MoD includes both pre-mating and post-mating reproductive barriers. The reproductive genotypes encode the pre-mating reproductive barriers and any incompatibilities that prevent the production of viable offspring. However, a "viable" offspring is likely to have ecological chromosomes (the predation resistance, prey template and generalism chromosomes) that are different from both parents' ecological chromosomes, due to recombination. Depending on the result of this recombination, the offspring may acquire a prey search pattern that fails to match any of the organisms in its patch. It may also acquire a predation resistance chromosome that makes it extremely vulnerable to predation. In either case, even though the offspring is "viable" it will soon die because it is not well matched to the ecological context of its time. This sort of fitness reduction is not taken into consideration in our definition of species because of its fundamental dependence on context.

It may seem difficult to identify a gene pool amongst a large number of organisms with this non-transitive mating relationship. However, there is a surprisingly efficient algorithm ($O(n)$ where n is the number of different reproductive genotypes in the population) for identifying the reproductively isolated gene pools. Remember that, as long as there is an indirect connection between two organisms through which a series of matings might bring

[13]Lonesome George is a Galapagos Giant Tortoise who is the last survivor of his subspecies (*Galapagos elephantopus abingdoni*). All attempts to get him to mate with members of other subspecies have failed.

their genes together in a descendant, then those two organisms share the same gene pool. Consider a diagram, or "graph" in computer science lingo, of who can mate with whom. Each node in the graph represents a unique reproductive genotype in the population. Since many organisms can have the same reproductive chromosome, a node may represent many organisms. We draw a line, or "edge," between any two nodes whose organisms can mate. That is, we connect nodes that are only one bit different. If we imagine that the nodes are beads and the edges are threads connecting the beads, then a species is the dangling set of beads you would get if you picked up one bead. A group of beads connected by threads is called a "connected component" of the graph. A graph may have many connected components. An example of a reproductive graph with two species is shown in Figure 2-4.

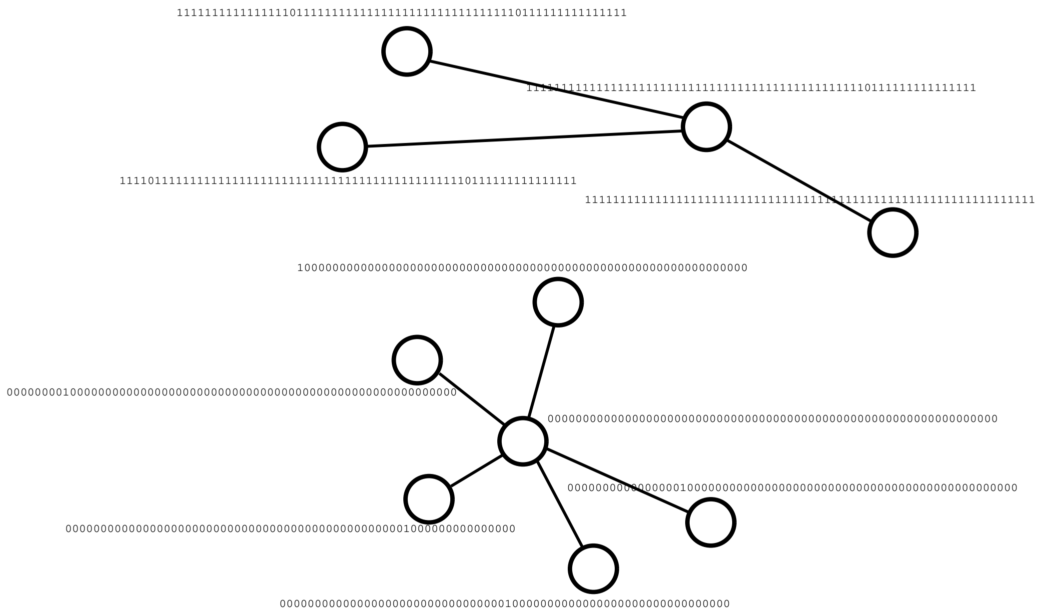


Figure 2-4: A graph representing the reproductive genotypes in a population and their relationships. The nodes represent the reproductive genotypes of the organisms currently alive in the model. An edge connects the nodes whose organisms are allowed to mate. These are the nodes that have only a single bit difference. There are two species in this graph, represented by the two connected components.

To identify a component we mark nodes with a depth-first search[14] of the graph, starting

[14]A depth-first search of a graph is a particular style of working through a graph. Starting at some arbitrary origin node, you mark it and then apply this same search recursively to each of its un-marked neighbors in turn, finishing with one neighbor before processing the next. If you think of the living members of a family tree as a graph, then this would correspond to starting with a grandmother, marking her, then moving to her first child, marking that child, then moving to that child's children, and so on. All the grandchildren of the first child will be marked before the second child of the grandmother is marked. This is why it is called "depth-first." You work as deeply as possible into the graph before popping back up levels

at any arbitrary node. All those nodes that get marked, and only those nodes that get marked are members of a single connected component, a single species. We repeat this until all the nodes have been marked and so all the species identified. Because this is an efficient algorithm, we can execute it every time step to track the species and identify exactly when a species splits apart.

A species splits when all of the organisms that carry a particular reproductive genotype die, and the loss of this node in the graph splits apart a connected component into more than one component. One strange byproduct of this implementation is that a species can split into more than two species. Notice that if either of the central nodes in Figure 2-4 disappear, many single node components will be created. We do not know if this happens in nature.

Death comes in three forms in MoD. An organism can be eaten by another organism. If an organism fails to find food in a time step, it will starve to death[15]. Finally, if a patch has been filled to its spatial capacity of 1K plants or 1K animals, then a new organism has a 50/50 chance of replacing an organism of its same kind (plant or animal). Thus, an organism can die in a sort of competition for space. Without this final form of death, evolution might grind to a halt. If there is no space left for new organisms, and no predators left to make space amongst the prey, then new mutations will never enter the population and nothing will change. As it turned out, plant population sizes were generally constrained by space, not herbivores. Because of the 3 : 1 transformation of prey matter into predator matter, the animal populations rarely rose above a third of the plant population and so never neared the carrying capacity of the space in a patch.

The implementation of a species as a connected component of a reproductive graph bears a certain similarity to the "holey adaptive landscape" model of Gavrilets and Gravner (1997; 1997). Both systems are constrained to high dimensional hypercubes. On closer consideration, there are a few important aspects that distinguish the two approaches. First of all, Gavrilets and Gravner consider only monomorphic species. That is, there is no genetic variation within a species, and so they can use a distribution model to represent

to finish off the upper levels of the graph. See Cormen, Leiserson, and Rivest (1990, p.478) for an $O(n)$ depth-first search of a graph.

[15] I played with allowing organisms to survive for longer without eating but it only decoupled the population dynamics of the predators and prey and so destabilized the model thus increasing the probability that one of the two populations would go extinct.

its dynamics rather than a configuration model. Second, the reproduction chromosome of our model explicitly codes pre-mating reproductive barriers, while the post-mating barriers of inviability emerges out of the starvation and predation of an organism in the ecosystem dynamics. In contrast, the analysis of holey adaptive landscapes has not yet been extended to coevolutionary dynamics or pre-mating reproductive barriers (Gavrilets & Gravner 1997).

2.2.4 Parallelism

The remaining aspects of MoD concern how it was implemented on a parallel computer. This is only of interest to programmers, and so other readers may want to skip to the next section.

The current model was written in CILK, a parallel extension of C, and run on the Xolas cluster of Sun Ultra symmetric multiprocessors at MIT[16]. CILK is simply C with the addition of two new key words, `spawn` and `sync`. By using `spawn` in front of a function call, a new thread is started and the function is executed in parallel with the rest of the code. A `sync` command will freeze the processing in the current thread until all the threads that it has spawned finish their work. CILK guarantees load balancing amongst the processors without programmer control. These two commands provide an easily understandable abstraction of parallelism at the cost of some restrictions. Most strikingly, there is no inter-thread communication[17]. The language is best suited to problems where the work can be split up into independent pieces.

The execution of the model can be nicely divided into the different patches. Almost all of the work occurs within a patch, independently of the others. The only exception is migration between the patches. So every time step, a separate thread was spawned to handle the interactions between organisms within each patch. At the end of the time step, these threads were synchronized and a single thread handled the distribution of organisms through migration as well as the collection of any statistics about the state of MoD.

To maximize the amount of work that was done in parallel, newborns were put in a special array in their parents' patch. The random walk of a newborn was implemented in parallel with newborns in other patches except that, instead of placing the newborn into its

[16]See `http://xolas.lcs.mit.edu/` for information on the processors and links to the CILK distribution pages.

[17]There are a few additional control abstractions for aborting threads and protecting memory from collisions between threads, but their use is officially discouraged and turned out to be unneeded for this model.

destination patch, the location of that destination patch was noted in the data structure that held the newborns. So, once the threads finished their work in all the patches, migration was a simple matter of reading the destination location associated with each newborn and stuffing the newborn into its specified patch.

2.3 Caveats

Meeting the requirements dictated by the hypotheses necessitates a non-trivial degree of complexity. However, there are many dynamics and characteristics of real biological systems that have been ignored. There is even good reason to believe that some of the missing pieces have significant impacts on the evolution of diversity. Whether or not they are necessary to address diversification is an open question, and so, for the nonce, they have been left out.

As was mentioned in Section 2.2.2, MoD makes no distinction between genotypes and phenotypes. The mapping from genotype to phenotype has implications for the constraints on variation across a population as well as the evolutionary trajectory of a species. Unfortunately, just what those implications are remains an open question.

I have consciously avoided the encoding of the size of organisms into MoD. This seemed an unnecessary elaboration. However, the size of an organism has dramatic consequences on its ecological behavior (Haldane 1927; Charnov 1993). In particular, size is important to the success of parasites. There are no parasites in MoD. Unlike predation in the model, a host can provide ample resources for many parasites as well as the parasite's offspring. Parasites are thought to have played an important part in the evolution of life (Hamilton et al. 1990; Paterson & Gray 1997; Anderson & May 1982; Hurst 1991), but none of the hypotheses for the diversification of life made explicit mention of them.

It is highly unlikely in MoD that a new mutant organism will establish a new species by the fact of its mutation. For a start, this would require two bits being flipped in its reproductive chromosome, which happens with probability 0.0002. Then, lest the mutant suffer the fate of Lonesome George, it must survive long enough for a viable mate to enter its patch. In contrast, it is relatively easy for a mutation in a real plant to establish a new species. This is simply because most real plants can fertilize themselves, and so need not depend on the arrival of a compatible mate.

The self-fertilization of plants brings up another important omission of the model. I have adopted the reproductive species concept and the following chapters will use the elegance of this definition as leverage for many of their results. However, the reproductive species concept only applies to organisms with some degree of bi-parental reproduction. If an organism was completely self-fertilizing or was asexual, then the reproductive species concept would not be applicable. Historically, most of life has been asexual. The phenotypic similarity within asexual "species" probably does not stem so much from sharing a gene pool as it does from the constraints of the fitness landscape. The innovation of sexual reproduction is interesting in part because it binds together a population of organisms into a sort of super-organism. An innovation in one lineage can eventually be shared by all the lineages in a sexual species.

There is very little behavior allowed to the organisms. Perhaps most strikingly, organisms only migrate once during their life time. The spatial scale of a patch is supposed to be such that almost all movement of individuals can be considered to take place within a patch. Furthermore, many ecological models focus on nutrient cycling through the food web and environment. Besides the fact that the hypotheses do not make reference to this level of detail, I have consciously chosen to work at a higher level of abstraction for reasons of both parsimony and computational resources.

Finally, perhaps the most stringent limitation on MoD is the fact that the speed of computation restricts the model to a regional, not global, level of representation. With 16 patches, a population that hovered around 20K organisms, and 5000 time steps, one run of MoD took about 20 minutes on an 8 processor Sun Ultra. One advantage of this model is that, if patch sizes remain constant, processing time scales linearly with the population/spatial scale. So, to model a world with twice as many patches and twice as many organisms in the same amount of time, we would only require a doubling in processor speed. But for the moment, all the results of MoD are restricted to evolutionary dynamics on a regional scale.

Is there any reason to believe that the model is an adequate representation of even the regional dynamics? This poses the problem of invalidating the model, to which we will now turn.

Chapter 3

Invalidation

Every model is an abstraction and a simplification. The goal of a model is to capture the essence of a system in the real world such that the behavior of the model matches the qualitative behavior of interest in the real system. Thus, for any model we may ask if it is a reasonable representation of the real system. Answering this question is the problem of validation. Traditionally we can try to invalidate the model by collecting data from the real system and comparing it to the behavior of the model. However, it should be noted that failure to invalidate can never prove the validity of a model (Caswell 1976).

In artificial life we rarely have the luxury of comparison to the real system. Artificial life models tend to be highly abstract and general because the field is striving to discover general properties of life. This makes experimental testing extremely difficult. The time scale of evolution tends to restrict experiments to observation of the fossil record (Benton 1990, for example) or manipulation of organisms with extremely short life-cycles in simplified environments (Lenski & Travisano 1994; Bull et al. 1997; Krukonis 1996, for example). Similarly, the complexity and size of ecosystems makes ecological experiments cumbersome and difficult to control.

An alternative form of testing can be pursued indirectly through reference to ecological and evolutionary theory. Instead of asking if the model matches the experimental data, we can ask if the model matches our understanding of the dynamics of ecology and evolution. Then, to the extent that the theories of ecology and evolution have been corroborated by experimental observations, we can invalidate the model when it fails to match those theories. These tests must be carried out at the appropriate level of abstraction. Because every model is a simplification, it will fail to match reality at some level of detail. For example, in the model described in Chapter 2, there is no distinction between phenotype and genotype. The fact that this is in blatant conflict with biological reality is peripheral to the issue of whether or not the model captures the essence of diversification. We are only interested in testing the dynamics represented in the model against their analogues in theoretical biology.

What follows is an example of this technique applied to MoD. While the purpose of this model is to explore new theoretical ground in biology, the ecological and evolutionary dynamics in the model have been tested against theories of predation, competition, and adaptation. As a result of these studies, the algorithms for implementing predation had to be elaborated, in accordance with the methodology I laid out in Section 1.4.2. Initially, a predator could automatically find and consume a prey organism in its patch, as long as the predator's search pattern matched the prey's predation resistance chromosome. As we will see below, such simplicity fails to capture the essence of a predator-prey interaction. The implementation of predator-prey interactions had to be fixed, and the final result is described both below and in Section 2.2.2.

The problems of speciation include both ecological and evolutionary dynamics of the model. Thus, it is important to test both the ecological dynamics and the evolutionary dynamics. Ecological theory provides qualitative predictions for predator-prey dynamics as well as competition. Evolutionary theory predicts changes in populations as they adapt to their environments. In the ecological experiments below, the dynamics within a single patch were examined when the mutation rate was 0. This prevented evolutionary change from interfering with the underlying ecological interactions.

3.1 Predator-Prey Oscillations

The Lotka-Volterra equations were an early abstraction of the dynamics of predator and prey populations. The details of the Lotka-Volterra equations are "hopelessly unrealistic" (Crawley 1992, p.68) because they lead to neutrally stable oscillations. Any perturbations to the system should generally result in either a stable fixed point or a stable limit cycle (May & Watts 1992). The introduction of a time lag between changes in the prey population and the response of the predator population tends to cause the two populations to oscillate stably. In our case the model requires three time steps to translate prey matter into predator matter. When predators are rare, prey populations will naturally expand, providing more food for the predators. Once the predator population expands in response to this abundance, they cull back the prey population. Eventually the prey become hard to find and predators begin to starve to death. And so the process repeats.

At a minimum, a model of an ecosystem should exhibit these oscillations. The model with a naive implementation of predation does not. As originally written, a predator could automatically find a prey organism in its patch. This meant that even when prey organisms were rare, the predators could still find them. The prey were driven to extinction and the predator population crashed shortly thereafter. This style of predation, where each predator captures a constant number of prey is called "fixed number predation," and, as we have seen, is unstable when prey are scarce (Crawley 1992).

A *prey-location* probability was introduced to the model to represent the chance that a particular predator might find a particular prey organism in its patch. However, a predator still gets one attempt for every potential prey organism in its patch. This results in stable herbivore-plant oscillations across a wide range of parameters. In particular, the *prey-location* probability can range from about[1] 0.15 to 0.3 and still produce stable oscillations. Higher than that and the herbivores overeat, killing off all the plants. Lower than that and the herbivores cannot find enough food to sustain life.

3.2 Trophic Cascades

While the introduction of the *prey-location* probability stabilized the two-trophic level dynamics, the difficulties multiply when we move to three trophic levels by the addition of a carnivore species. Unfortunately, there is only a knife-edge of parameter space that can sustain both a two and a three trophic level ecosystem. Since herbivores are rare, relative to plants, the *prey-location* probability must be high to allow the carnivores to find the herbivores. However, if it is too high, the carnivores will kill off all of the herbivores. All three trophic levels could only be sustained by lowering the *energy-conversion* parameter to unrealistic levels, requiring the consumption of only two prey organisms to produce a new predator organism.

[1]Tweaking other parameters like *energy-conversion* or the amount of energy given to the plants can increase this range of stability.

The problem comes when the predator population has expanded and culled back the prey population to low levels. For a prey organism to survive during a time step, it must evade location by *every* predator in the patch. Even a low *prey-location* probability cannot save a prey organism from a large predator population.

The spatial structure of the real world tends to soften intense predation. A prey organism is generally not directly threatened by all the predators in the patch, because some of the predators are busy stalking other prey or pursuing other interests. Predators sometimes patrol a fixed territory from which they exclude one another. Furthermore, there are many forms of refuge in which prey can find safety. Reasonable forms of refuge that might be captured in the modeled system include hiding places, microhabitat selection by the prey, and aggregation of the predators leaving some areas unpatrolled. These are but a small subset of the forms of refuges found in biology (Crawley 1992). However, the important characteristic of all of them is that there is variation across prey individuals in the probability they will be attacked by predators.

The model was elaborated with two mechanisms, effectively simulating the addition of interference between the predators, or possibly the presence of prey refuges. First of all a predator *search-radius* was added, restricting the amount of the patch that a predator can patrol in one time step. Each predator starts searching in a randomly chosen location in the patch's organism array, and then looks for prey in the *search-radius* consecutive following locations[2] in the array. Second, a prey organism that survives an encounter with a predator, because of the *prey-location* probability[3], becomes more difficult to find by other predators. Specifically, the *prey-location* probability is multiplied by $e^{-(\text{number of encounters} \times \textit{predation-distribution})}$. When the *predation-distribution* is 0, there is effectively no spatial structure, and the model behaves as it did before. However, when the *predation-distribution* is positive, a form of spatial structure is imposed on the patch, and a small prey population has a better chance of surviving intense predation. This introduces interference competition between the predators.

An ecological model with multiple trophic levels, in this case plants, herbivores, and carnivores, should demonstrate a trophic cascade in the plant density. A trophic cascade is a

[2] The array was formed into a "ring" so that the last location in the array is adjacent to the first location.

[3] If the predator search pattern mismatches the prey's predation resistance chromosome in m bits, then an additional e^{-m} is factored into the probability of the predator consuming the prey.

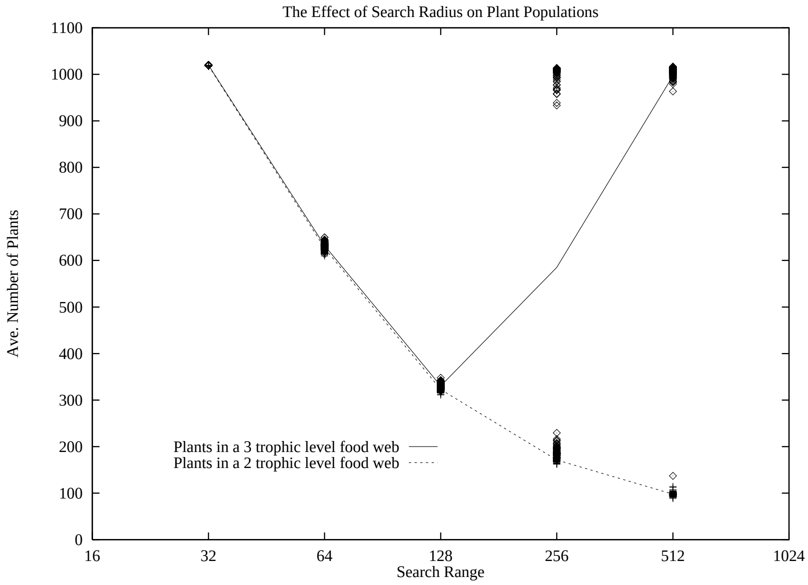
change in plant or autotroph biomass in response to predation. In the absence of herbivores, plant biomass ought to be relatively high. In contrast, the introduction of herbivores tends to dramatically reduce plant biomass. However, the further introduction of carnivores to prey upon the herbivores tends to restore plant biomass to some intermediate level. Does the model demonstrate a trophic cascade?

3.2.1 Results

The model was run for 500 time steps with a single patch under varying parameters. As long as the carnivores were able to survive, a trophic cascade was found. In most cases, the model was run at least 40 times for each parameter setting, although in some cases it was run as many as 100 times.

The addition of the *search-radius* had little effect. A small *search-radius* actually helps to support a larger herbivore population by preventing over predation in a two trophic level system. However, if we want carnivores to survive, they must find the herbivores scattered amongst the plants, and so require a large *search-radius* to survive. In a patch that could hold 2048 organisms (with a maximum of 1024 plants), carnivores required a *search-radius* of more than 128, and tended to be stable with a radius of 512 as shown in Figure 3-1. When the search range is small, the carnivores cannot find the herbivores scattered about amongst the plants. The carnivore population crashes and the herbivores are left unmolested in their consumption of the plants. However, when the search range is greater than 128, the carnivores sometimes manage to hang on, reducing the herbivore population and allowing the plants to fill their environment to the carrying capacity of 1024 plants. With a search radius of 256, the model settled into one of two states, represented by the two clusters of points at that setting. The curves, representing the averages of the data at each parameter setting, lie between the two clusters. Data for single trophic level ecosystems are not shown because the plants always fill the environment to its carrying capacity.

Each data point indicates an independent run of the model, and the curves connect the means for data points at each parameter setting. A trophic cascade can be seen when the plant biomass in a two trophic is significantly less than the plant biomass with three trophic levels. This is illustrated by the solid curve diverging from the dashed curve in the graph of plant populations.



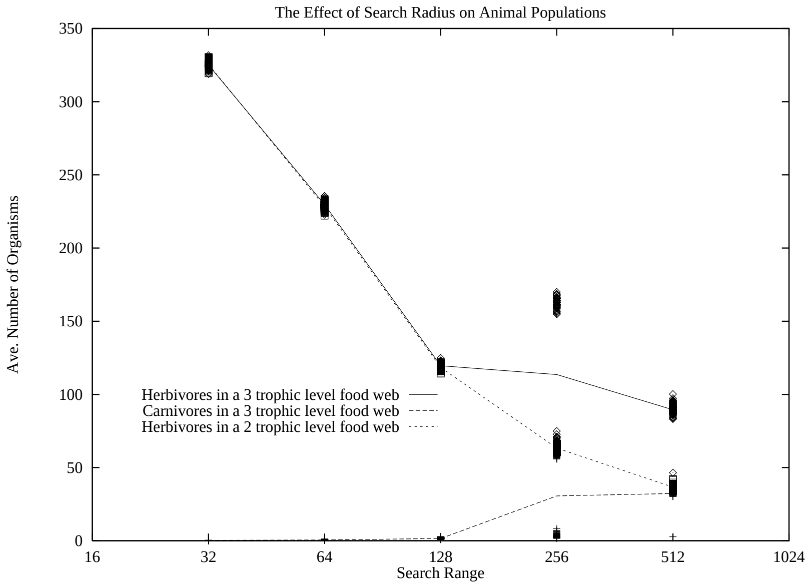


Figure 3-1: Each data point represents the average number of plants in a run, while the lines represent the average across the 100+ runs for each parameter setting. The "search-range" is the number of organisms that may potentially be encountered in a time step by a predator (both herbivores and carnivores). The upper graph shows the effects of the search range on the plant populations. The lower graph shows the effects of search range on the herbivores, in both a two and three trophic level food web, as well as the carnivores in the three trophic level food web. It is interesting to note that when the search range is small the herbivore population thrives relative to a moderate search range, even in the absence of carnivores.

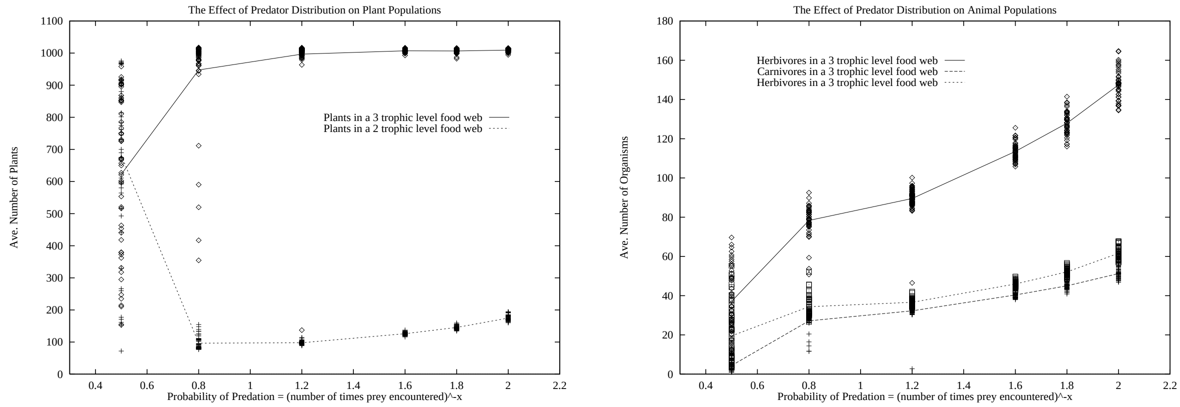


Figure 3-2: The effects of enforcing more spatial structure and so adding prey refuges with the *predation-distribution* parameter on the biomass of the three species. The left graph shows the plant biomass in the two and three trophic level ecosystems. The right graph shows biomass of the herbivores and carnivores in the two and three trophic level ecosystems.

The *predation-distribution* parameter had a much more dramatic effect. Both herbivore and carnivore populations were stable with a *predation-distribution* ≥ 0.8. In fact, the average biomass of both herbivores and carnivores increased with the *predation-distribution*, up to the maximum value that was examined (2.0), as seen in the right-hand graph of Figure 3-2. The predator populations expand as the *predation-distribution* parameter makes it harder for them to capture prey. A similar thing could be said of the herbivores in Figure 3-1. As the search radius decreases and the herbivores are restricted to feeding on fewer plants, herbivore biomass actually rises.

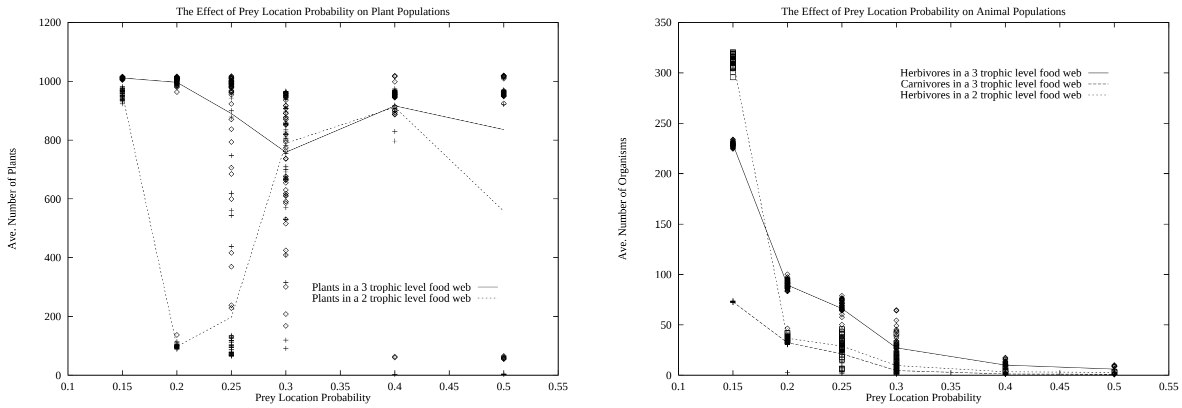


Figure 3-3: The effect of the prey location probability on the population sizes of plants (left) and animals (right).

A similarly counter-intuitive result can be seen if we look at the effect of the *prey-location* probability on the populations of herbivores and carnivores as shown in Figure 3-3.

The most dramatic trophic cascade appears when the *prey-location* probability is 0.2 (the setting used for future experiments). It is interesting to note that all three trophic levels thrive when the probability is at a minimum 0.15. Once again, as the parameter values make predation more difficult, the average predator populations expand. This is probably due to a more efficient allocation of resources to the predators. The average maximum and minimum population sizes of the predators are shown in Figure 3-4 along with the mean values. When predation is easy, the predator populations can expand rapidly, reaching higher values than the maximum population sizes attained when predation is hard. However, the success of the predators has a dramatic effect on the prey, and so the predator populations collapse, reaching lower minima as compared to the minima in the runs where predation is more difficult. Ease of predation tends to destabilize the populations, causing larger fluctuations and resulting in lower average predator populations over time. Crawley (1992) argues, from the distribution models, that prey refuges will stabilize predator-prey dynamics as long as the refuges support a large enough prey population to preserve the predator population in lean times.

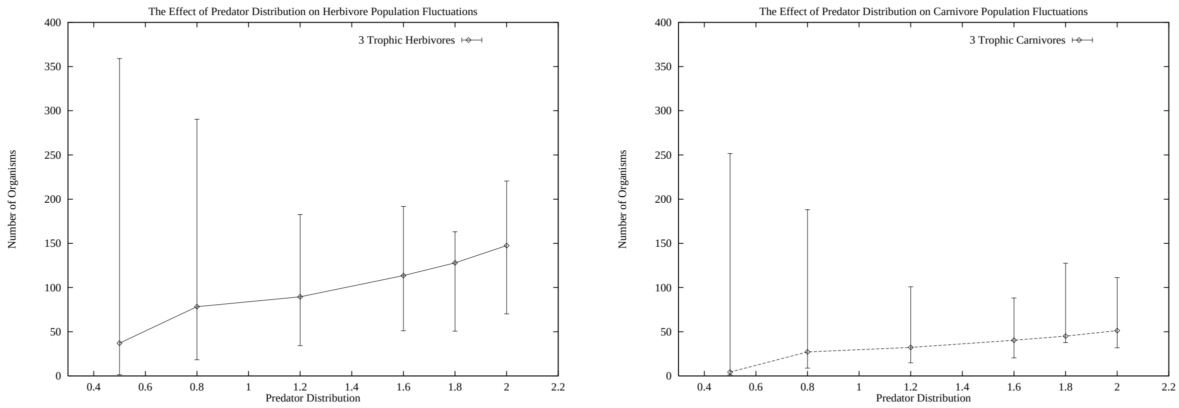


Figure 3-4: The average maximum and average minimum population sizes have been added to the herbivore and carnivore curves from the lower graph of Figure 3-2.

The effects of two further parameters were investigated, *energy-input*, as shown in Figure 3-5, and *energy-conversion*, as shown in Figure 3-6. When the *energy-input* is less than sufficient to give each plant a single energy unit in a time step, the plant populations do not replenish themselves fast enough to support many animals. Once the *energy-input* rises above 1, herbivores begin to thrive. Note the difference between the herbivores with no predation and the herbivores that suffer predation. The presence of carnivores seems to reduce competition amongst the herbivores, when the *energy-input* is 1, and consequently allows the herbivores to expand to greater population sizes than they could maintain without the carnivores. Yet, when the *energy-input* is greater than 1, the presence of the carnivores prevents the herbivores from responding to the greater plant abundance. A trophic cascade can be observed in the left graph of Figure 3-5 when the *energy-input* is 1 and 1.25.

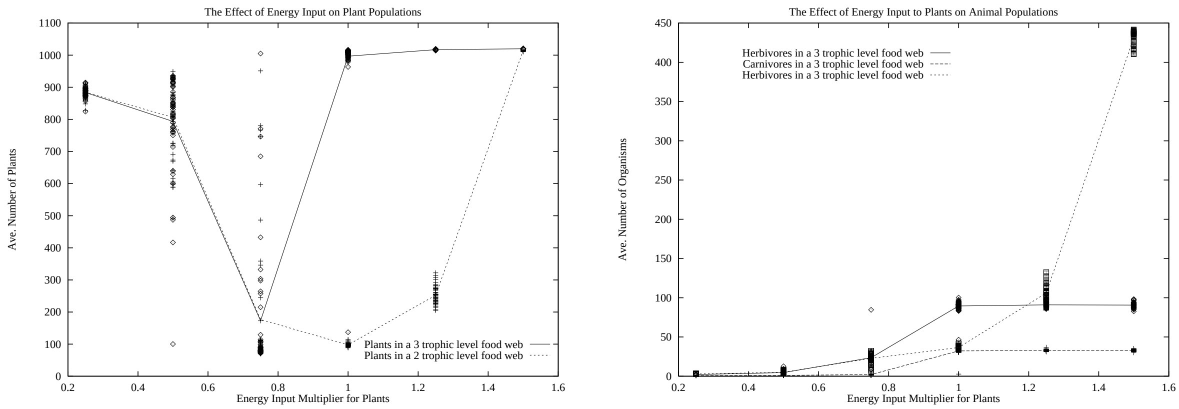


Figure 3-5: The effect of *energy-input* on population sizes. The *energy-input* is the number of energy units a plant can absorb in one time step if it is perfectly matched to its environment (as all the plants are in these experiments).

Finally, the right graph of Figure 3-6 shows that the animal populations decline as the *energy-conversion* parameter forces them to eat more prey organism before they can reproduce. When the *energy-conversion* parameter is high, the predators have difficulty gathering enough food to survive and their populations die off. This accounts for why the plant populations are high when the energy conversion is 8 : 1. Biologically realistic conversion values range from 3 : 1 (for plant to animal matter as in the above) to 10 : 1 (for abiotic to plant matter), though distribution models often use 1 : 1 (Schmitz & Booth 1997). In all other runs, the energy conservation parameter was set to 3. A trophic cascade can be seen in the left graph of Figure 3-6 when the *energy-conversion* parameter was set to 3 or 4.

3.3 Competitive Exclusion

Competition between two species in an ecosystem often results the competitive exclusion (extinction) of one species by the species better adapted to the environment. While relative "adaptation" is often difficult to quantify independently of competitive exclusion experiments, it is easily identifiable in the model. A series of competitive exclusion experiments

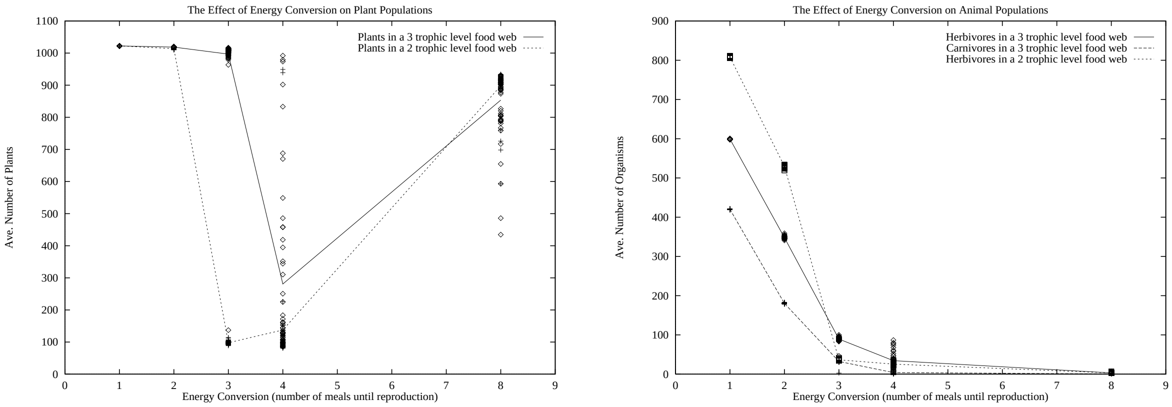


Figure 3-6: The effect of *energy-conversion* on population sizes. The *energy-conversion* parameter is the number of meals an organism must accumulate in order to produce a single offspring. This is the scaling factor by which prey matter is transformed into predator matter.

were designed wherein the environment was seeded with equal numbers of two species of plant. The "climate" of the environment had all its bits set to 0 and the prey template chromosomes of the plant species differed from each other by only one bit. That is, one of the two species was exactly one bit better adapted to the environment than the other. This should result in a slightly better energy absorption and a consequent slightly higher reproduction rate in the species with fewer bits set to 1. A species was considered to have excluded the other species when it had expanded to 90% of the carrying capacity of the environment. 100 trials were run with 1, 4, 8, 16, 24, and 31 bits set to 1 in the prey template chromosome of the inferior species. Mutation was turned off for these experiments. The superior species excluded the inferior in all the trials.

One might ask how long does it take for the superior species to take over 90% of the carrying capacity? If competitive exclusion tends to occur quickly, then we can infer that the slope of the fitness landscape in that region of the prey template chromosome is relatively steep. Figure 3-7 shows the inverse of the average number of time steps required for the exclusion of the inferior plant along with a linear sketch of the fitness landscape. When there is only 1 bit different from the optimum, the left graph indicates that the slope of the fitness landscape is very steep. The landscape has a single optimum surrounded by a steep slope that becomes more gentle by the time that the prey template chromosome is 4 bits away from the optimum. This is sometimes called a "Mount Fuji" fitness landscape because there is only one fitness peak in the landscape (there are no local optima besides the global

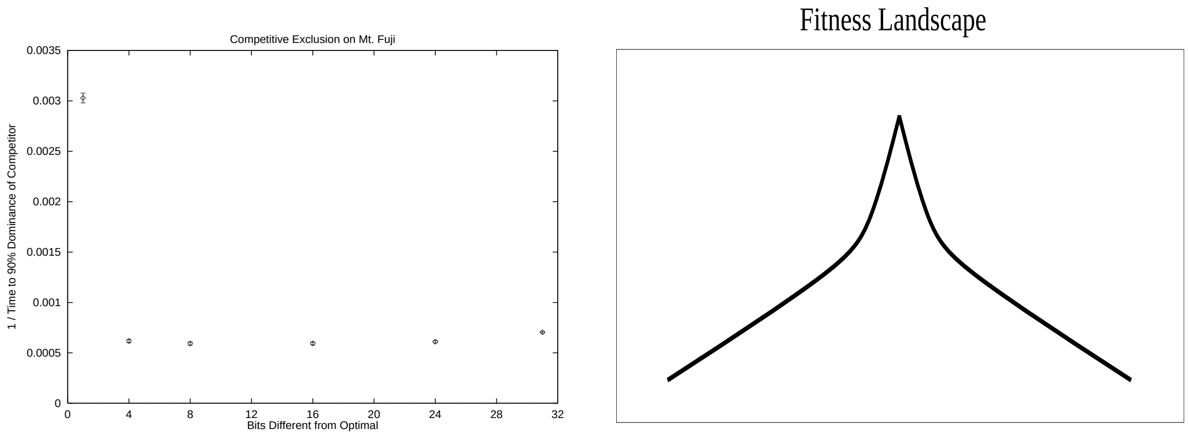


Figure 3-7: The fitness landscape for plants. The left graph shows the inverse of the average number of time steps, with standard error bars, required for the superior plant species to take over 90% of the carrying capacity in a patch. The intensity of selection is inversely proportional to the amount of time it takes for one species to exclude the other. Thus, the slope of the fitness landscape is inversely proportional to the amount of time before competitive exclusion occurs. From this, we can infer the slope of the fitness landscape at the various distances from the optimum, depicted in the right graph.

one) and all other points in genotype space rest on a slope that leads to that peak.

3.4 Adaptation

One of Darwin's (1859) key insights was that biological species are not immutable Platonic entities but rather they are dynamic conglomerations of lineages flowing over a fitness landscape. His theory of natural selection is in part a prediction that a species will adapt to its environment. That is, populations should tend to climb the peaks in the fitness landscape.

A model of evolution should at least demonstrate that, all other things being constant, organisms evolve to be better adapted to their environment. This can be tested in the model when mutation is introduced into the dynamics. Consider a plant species evolving in the conditions of the competitive exclusion experiments. The optimal genotype matches the climate bit pattern of all 0's. This is achieved by having 0's (or wild-cards) in all the loci of the organism's search pattern. If the prey template and generalism chromosomes were evolving neutrally, with no natural selection, then we would expect one in four bits in the plant's prey template to have a 1 that is not masked by the generalism chromosome. We can thus detect natural selection in any significant reduction in that proportion of unmasked

1's below 25%.

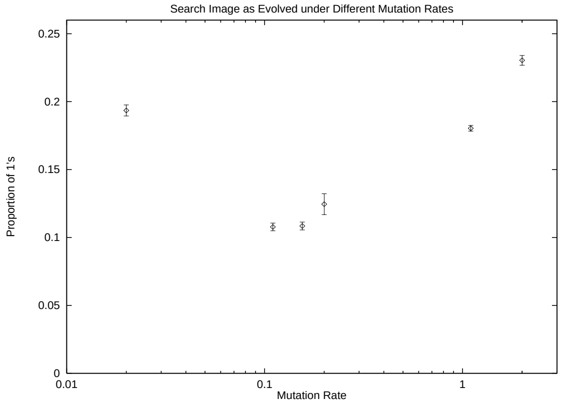


Figure 3-8: Adaptation in plants. The normalized distance (with standard errors) of plant populations from the optimal genotype under different mutation rates. In all cases the null hypothesis of neutral evolution can be rejected.

The model was initialized with a plant species with exactly 8 out of 32 positions in its prey template set to 1 and no wild-card positions in its generalism chromosome. After 5000 time steps the proportion of unmasked 1's was counted and averaged. The data for at least 32 runs in each setting of the *mutation-rate* parameter is shown in Figure 3-8. For all mutation rates, the resulting data is significantly below a mean of 25% unmasked 1's as tested by a one-sided t-test ($p < 0.001$ in all cases). At low mutation rates, the populations have not yet had time to reduce the number of 1's in their search patterns, and so the proportion of 1's is still high after 5000 time steps. At high mutation rates, mutation keeps adding new 1's to the search pattern and so the proportion of 1's stays high. At intermediate mutation rates, selection has enough time to weed out the deleterious mutations while preserving the beneficial ones. There can be little doubt that the plant species have adapted to their environment.

3.5 Conclusion

The initial failure of the model and subsequent elaboration of the predation algorithms illustrates an important benefit of attempts to invalidate a model. Failure helps to sharpen our understanding of the essential features of the system and so helps to guide the further development of the model. It is also important to note that even when there exists a paucity of theory that can be used to test the central results of a model, there generally exists a rich field of theory that impacts upon the dynamics of the model. In this case, a model of species diversification was designed for the very reason that there is a lack of established theory covering that topic. However, such a model must include both ecological and evolutionary dynamics and so I have tested it against both ecological and evolutionary theory. Because the dynamic of diversification intimately depends upon ecological and evolutionary interactions, these studies help to corroborate the end results of the model.

The fact that a model passes a series of tests against theory does not guarantee the quality of its results. The model described in Chapter 2 is only one instantiation of the requirements for modeling species diversity, and so stands as a single data point for theory. This perspective on models is particularly important to artificial life where seemingly trivial implementation details often manifest in artefactual results. In the end, the model does not provide Truth but rather a hypothesis that should be experimentally tested in the laboratory and field, similar to any other theoretical result. It must be granted that such experimental corroboration is difficult. Meanwhile, by testing our models against theory, we may at least make progress toward significant theoretical insight.

3.6 Parameters for the Following Chapters

Species	**Chromosome**	**Loci**
Plant	reproduction	0000000000000000000000000000000011111111111111111111111111111111
	predation resistance	00000000000000000000000000000000
	prey template	00000000000000000000000000000000
	generalism	00000000000000000000000000000000
Herbivore	reproduction	0000000000000000000000000000000000000000000000000000000000000000
	predation resistance	10101010101010101010101010101010
	prey template	00000000000000000000000000000000
	generalism	00000000000000000000000000000000
Carnivore	reproduction	1111111111111111111111111111111111111111111111111111111111111111
	predation resistance	01010101010101010101010101010101
	prey template	10101010101010101010101010101010
	generalism	00000000000000000000000000000000

Table 3.1: The genotypes of the three species used to seed the model. All the species are extreme specialists on their respective food sources. Furthermore, the reproductive genotypes of the three species are different from each other in at least 32 loci.

For the experiments in the following chapters, I chose to set the parameters that were not being examined in the experiments to maximize biomass in both the plants and animals.

Yet, I have kept to the parameter ranges that demonstrated realistic ecological behavior. Specifically, I used a *search-radius* of 512, a *predation-distribution* of 2.0, a *prey-location* of 0.15, an *energy-input* of 1.0, an *energy-conversion* of 3, and a *mutation-rate* of 0.02.

In every case, the model was initialized with 500 plants and 200 animals. The animals were roughly composed of 150 herbivores and 50 carnivores. The chromosomes for the three species are shown in Table 3.6. Initially, all members of a species had identical chromosomes.

And now for something completely different...

Chapter 4

Speciation

A common problem in all experimental science is that our theories and discussions tend to be guided by the phenomena that are easily observed. However, what is easily observed may not be a good sample of the important dynamics that shape our world. This is the old problem of looking for your keys under the lamp post. Thus the importance of new technologies for observation in the advancement of science is that they allow us to peer into the darkness and perhaps catch a glimpse of what is out there.

This problem plays out in evolutionary biology through the difficulty of defining, and consequently identifying, species and speciation events. Let us consider the adoption of the predominant species definition, the reproductive species concept. As I argued in Section 1.2, for modeling purposes, a species is best defined for the purposes of modeling to be a reproductively isolated gene pool. As we shall see, adopting the reproductive species concept has some surprising consequences and at the very least helps to resolve some common confusions. While we can most easily observe various aspects of organism's phenotypes, these characteristics are often not the ones that are most important in determining reproductive isolation. And for centuries now, the most striking aspect of an organism's phenotype is that it often seems to be adapted[1] to its environment and lifestyle. This has led to a common confusion between species and niches, or peaks on an adaptive landscape (Gavrilets 1997, discusses an aspect of this confusion). The concepts of clinal speciation (White 1978), and ring species (Oliver 1972;

[1]Gould (1979, for example) has written extensively on the bias in evolutionary biology to see everything as an adaptation. I accept his point as a valuable contribution to our view of the history of life. Nevertheless, the presence of adaptations in organisms is undisputed.

Fox 1951) are based on identifying changes in phenotype that correlate (not necessarily causally) with decreased viability of hybrids. The argument for focusing on non-reproductive phenotypic characters is strongest when considering post-zygotic reproductive barriers. It seems reasonable that if two organisms with divergent phenotypes produce an offspring, that offspring may inherit an intermediate phenotype that is untenable in any of the local habitats. Both parents may have been well adapted to their different environments but their offspring will not survive. However, undue attention to easily observable phenotypic characters is misleading. Pre-zygotic reproductive barriers and hybrid infertility may have little to do with the observed phenotypic characters and tend to depend more on the specific changes in characters influencing reproduction (see Futuyma (1986) for a review of this issue). These often unseen characters are the ones that determine the limits of a gene pool which in turn binds the lineages together into a common fate. Thus, by accepting the reproductive species concept, the dynamics of speciation shift into this largely unseen world of the evolution of reproductive characters.

With the exception of some specialists, it is probably fair to say that biologists are fundamentally more interested in phenotypic change than in species and speciation. In some sense, the attention paid to species has been due to a confusion of the two processes. Species are often viewed as proxies for phenotypic categories of individuals. It is certainly true that there is an intimate relationship between the phenotypic changes in a lineage and the evolution of species and speciation. However, this chapter will illuminate the dynamics of evolution when we ignore phenotype for the most part and just fix our eyes on reproductively isolated gene pools, that is, species.

The intuition behind Mayr's reproductive species concept is that an isolated gene pool constrains the evolution of phenotype. He even goes as far as to view speciation as an important engine of change in evolution (Mayr 1982b, p.296). Once a subpopulation is reproductively isolated from its "parent" population, traits that are fixed[2] in the subpopulation cannot be changed by contact with the parent population. Whether these traits have become fixed through random genetic drift in the, perhaps smaller, subpopulation, or through selection in a different environment, is immaterial. The reproductive isolation of a

[2]To say that a trait or an allele is "fixed" in a population means that all the individuals carry that allele, and only that allele, at the locus. The result is that the trait will not change over time through recombination or natural selection. However, mutation may destroy fixation.

speciation event acts as a ratchet in evolutionary change such that there is no going back[3]. Species thus form the foundation upon which phenotypic changes are built. But, under the isolated gene pool view of species, to mistake a phenotypically differentiated population for a species is like a frog in a well mistaking the well for the entire world[4]. We have been ignorant of all the species that have not become differentiated[5].

This chapter investigates the tempo and mode of diversification in reproductively isolated gene pools. It seeks to provide answers to the first two levels of questions in the diversification of life. Are speciation rates and extinction rates independent as Hoffman (1986) suggests, such that diversification is climbing an exponential curve? And if not, what is the nature of the regulation on the numbers of living species? The model suggests that the process of diversification is indeed regulated leading to a diversity equilibrium. We can then ask if that equilibrium scales with area in a similar fashion to the species-area curve of island biogeography. But more fundamentally, we may ask why is there such an equilibrium? In the process of answering these questions it will become apparent that a successful species must be producing small reproductively isolated populations, that is, new species, at a tremendous rate, like the eddies along the banks of a turbulent river. This may well be occuring in the absence of geographical isolation.

4.1 Diversity Over Time

Does species diversity rise exponentially in the model, or does it settle down to some diversity equilibrium? The model was run 50 times under the so called "vanilla" parameter settings. There were no migratory barriers between patches. The climate, or habitat, was the same in every patch, giving the plants a homogeneous, abiotic environment. And there was no assortative mating, mating was random within a patch with no mate preferences amongst the organisms. The model was seeded with three species, 500 plants, approximately 150 herbivores and 50 carnivores. After 5000 time steps the number of species remaining in the model, as well as the number of speciation events, were counted and stored. This provides the baseline species diversity and speciation rates against which the effects of the

[3]This irreversibility is perhaps overstated. The complexities of development generally determine the likelihood that a mutation might reverse a change that has become fixed in a population. We do not yet have a good grasp of when these complexities allow a reversal and when then tend to prevent it.

[4]An old Korean proverb.

[5]This conclusion is true for all species concepts that do not depend on morphological differentiation.

hypothesized manipulations will be measured in future chapters. The vanilla conditions resulted in an average of 2.5 extant, or living, species after 5000 time steps. It generally was not able to sustain all 3 species introduced into the model. However, over those 5000 time steps there was an average of 0.0425 speciation events per time step. The diversity dynamics, averaged over all 50 runs are shown in Figure 4-1.

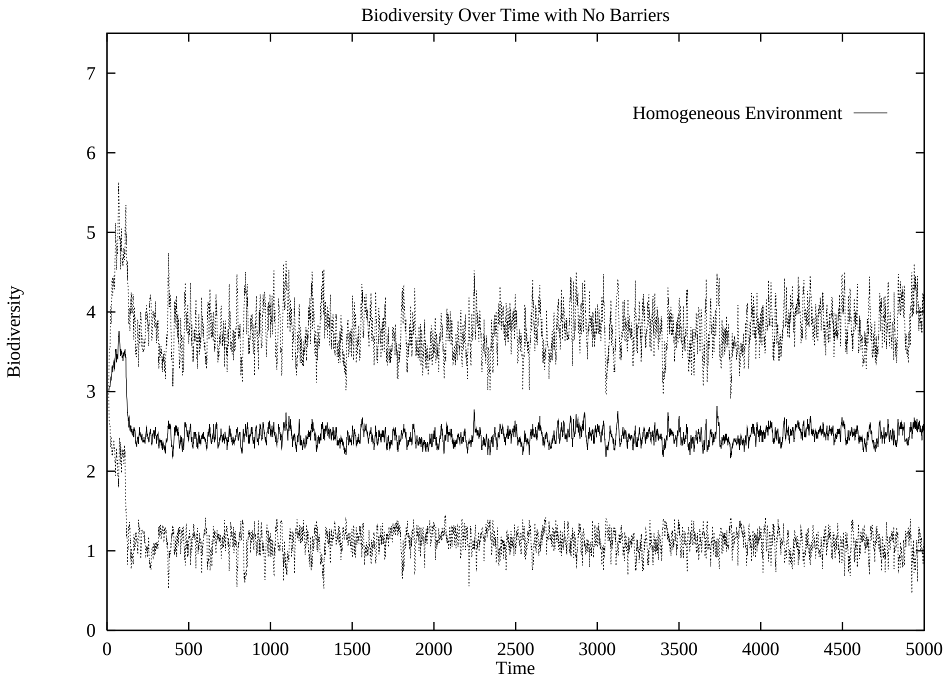


Figure 4-1: The average biodiversity (middle curve) for 50 runs of the model with no barriers under a homogeneous environment. The 95% confidence interval has also been plotted, surrounding the mean. The model was seeded with three species, a plant, an herbivore, and a carnivore, but is not able to sustain all three species.

The vanilla conditions are not adequate to sustain all three of the species initially introduced into the model. Of the 2.5 species remaining at the ends of the runs, 1.38 were plants and 1.12 were animals. The results are a little more interesting for 0.9 migratory barriers in every patch and a heterogeneous environment. Under the heterogeneous environment condition, the climate patterns of the patches differed from one another in an average of 10 bits, as described in Section B.1. This resulted in a biodiversity of 5.74 species and a speciation rate of 0.1538 species per time step. Out of the 5.74 species surviving in the heterogeneous environment, on average 2.00 of them were animals. The homogeneous environment with 0.9 migratory barriers produced a diversity of 5.04 species (1.58 of which were animals) and a speciation rate of 0.1392 species per time step. The diversity time series, averaged over

50 runs for both homogeneous and heterogeneous environments is shown in Figure 4-2. In addition, Figure 4-3 shows the mean of species diversities during 50 runs with parameters set to maximize animal species diversity, as determined from the experiments in Chapter 3.

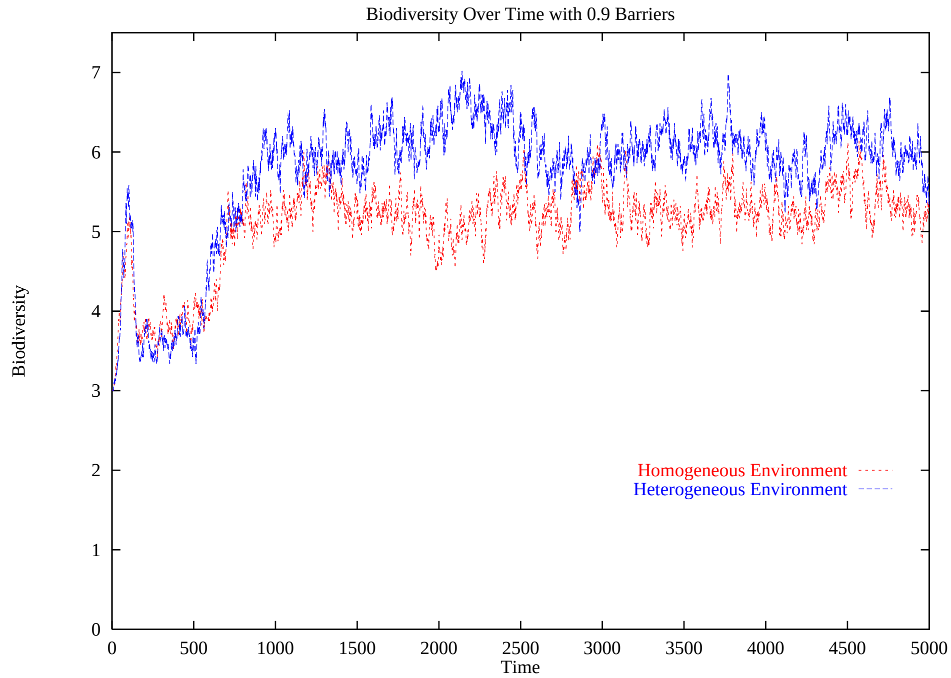


Figure 4-2: The average biodiversity for 50 runs of the model with 0.9 barriers under both a homogeneous and a heterogeneous environment. In the homogeneous environment case, there was no variation between patches in their habitats. The habitats in a heterogeneous environment differed in an average of 10 bits. The model was seeded with three species, a plant, an herbivore, and a carnivore.

There does appear to be a sort of carrying capacity for diversity in the model, as seen in Figures 4-1, 4-2, and 4-3. After an initial steep increase in the number of species, the curves level out. The transient dynamics with an initial spike in diversity at the start of the model is probably an artefact of the particular starting conditions (500 plants, 200 animals, and genetic homogeneity within each species). If we accept this apparent diversity equilibrium as realistic, it shifts the debate on exponential versus logistic diversification to the question of where the Earth is on the logistic curve. Have we yet to reach the natural diversity capacity of the Earth? Or are there factors that have been raising the ceiling of the diversity capacity over time?

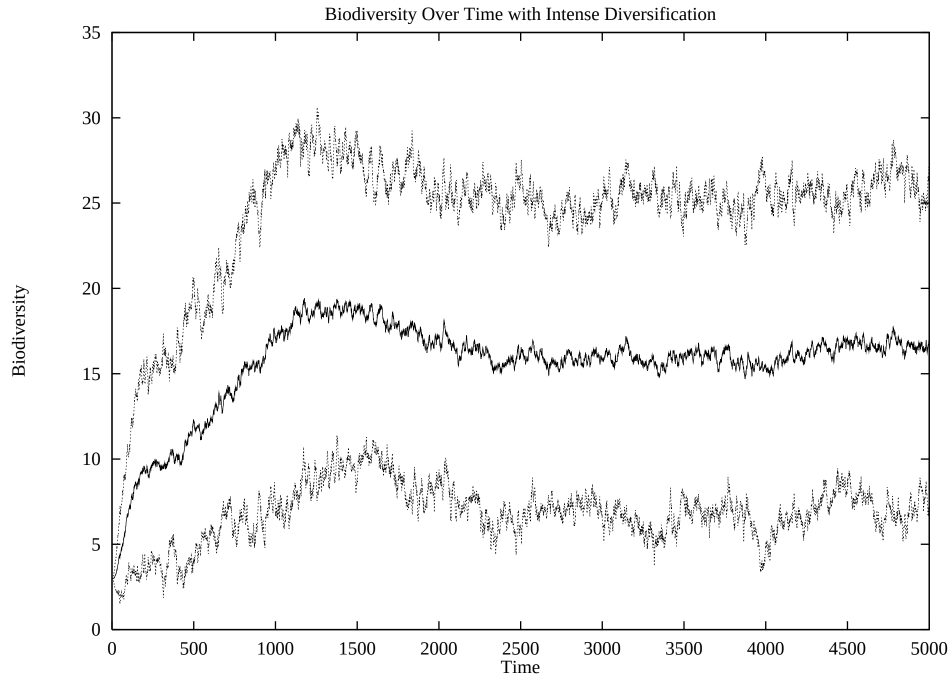


Figure 4-3: The average biodiversity (middle curve) for 50 runs of the model with parameters set so as to maximize animal species diversity. This means a heterogeneous environment, 0.95 migratory barriers, and positive assortative mating based on predation resistance chromosomes. The 95% confidence interval has also been plotted, surrounding the mean. The model was seeded with three species, a plant, a herbivore, and a carnivore.

4.2 The Species-Area Curve

In 1967, MacArthur and Wilson launched the field of island biogeography with the book "A Theory of Island Biogeography" (MacArthur & Wilson 1967). They had noticed that the number of species on an island remains relatively constant, despite the continuous influx of colonizers and frequent local extinctions on the island. They proposed the elegant explanation that the frequency of a new species colonizing an island is negatively correlated with the biodiversity on the island, while the frequency of extinctions is positively correlated with biodiversity.

The theory proposes that there is some source from which the colonizers arrive, generally the mainland or a large neighboring island. As more of these colonizers establish a foothold on the island, it becomes less likely that a new arrival is novel to the island. Furthermore, as the number of species increases on the island, the amount of free resources, or open niches, declines while competition increases. In other words, life becomes harder. And so, the frequency of extinctions increase. Thus, at low diversity levels, life is easy and

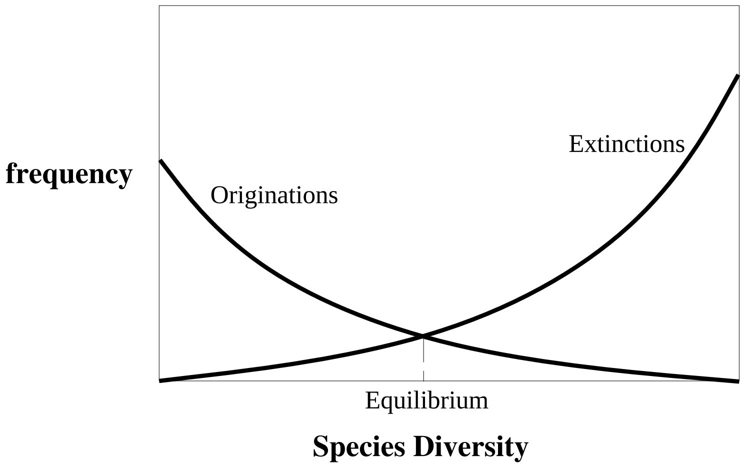


Figure 4-4: Species diversity equilibrium in island biogeography. The probability of a new species colonizing an island drops off with diversity because there is a higher chance that the species already exists on the island and competition makes it more difficult to establish a foothold. Similarly, extinctions become more likely with high species diversity because competition is more intense. At some point these two forces balance out in a species diversity equilibrium.

new arrivals have a relatively easy time establishing a colony. The result is that species diversity increases. However, at high diversity levels, the arrival of a species that is not already present becomes less likely and extinction rates increase, driving diversity levels back down. At some point these two opposing forces should balance out. This is the diversity equilibrium point shown in Figure 4-4. It should be noted that it is a dynamic form of equilibrium. Origination and extinction rates on islands were surprisingly high when ecologists first measured them. There is a lot of turnover or churning in the species composition of an island. However, the overall diversity levels tend to remain constant.

It is tempting to draw an analogy between the diversity capacity in the model and the diversity equilibria observed in island biogeography. The diversity equilibrium of Figure 4-4 derives from the balance between immigration and extinction on an island. The important difference in the evolution of biodiversity is that the "originations" derive from speciation events, not colonizers from some distant mainland. The theory remains the same in regards to extinction rates, suggesting that extinction rates should increase with diversity levels. However, it is not clear what the relationship is between speciation rates and diversity levels. For the phenomena to parallel the dynamics of island biogeography, the speciation rates would have to decline with diversity levels. Huston (1994) argues that the predictions of island biogeography are only likely to hold under fairly restrictive qualifications. Specifically, when diversity in a single taxon is examined and originations derive from immigration not speciation (Huston 1994). The following experiment violates both of those restrictions.

4.2.1 The Effects of Area on Diversity

The experiments of Chapter 3 were conducted within a single patch. To investigate the relationship of species diversity to area, we must move to multiple patch experiments. The central result of island biogeography is that the number of species (S) on an island is a power law of the area (A) of the island (MacArthur & Wilson 1967). This is called the species-area curve. Specifically, $S = cA^z$, where c is a constant and the exponent z is particular to the group of organisms being studied. Begon et al. (1990, p.778) have summarized observed values for z which range from a low of $z = 0.10$ for English flowering plants to a high of $z = 0.43$ for mountainous mammals in the USA, with an outlier of $z = 0.72$ for cave dwelling invertebrates.

The model was run in four spatial configurations, 1 by 2, 2 by 2, 2 by 4, and 4 by 4 patches. The habitats of the patches were made to differ by randomly flipping 2 of their 32 bits. There were topological barriers to migration of 0.9 between all patches. From the perspective of population genetics, where it only takes a few migrants per generation to prevent adjacent populations from diverging, this may seem like it would result in a high migration rate. However, the actual rate of migration between any two neighboring patches is approximately 0.005 (see Appendix B for the relevant derivations). This should allow for a moderate level of divergence between neighboring populations.

Species diversity data were collected after 5000 time steps for 50 runs of the model under each spatial configuration. A linear regression of the natural-log of the species diversity as determined by the natural-log of area (number of patches) gives $z = 0.44$ ($p < 0.001$, standard error $= 0.05$), as shown in Figure 4-5. While this is on the high end of the typical range found in nature, it is not unheard of. Most island biogeographical studies have focused on a single group of organisms, such as birds, ants, or land plants. They do not look at diversity across multiple trophic levels, and so miss coevolutionary diversification effects. The z of 0.44 compares well to the values (0.7–0.95) found by Jones et al. (1997) for genotype diversity in the Echo model.

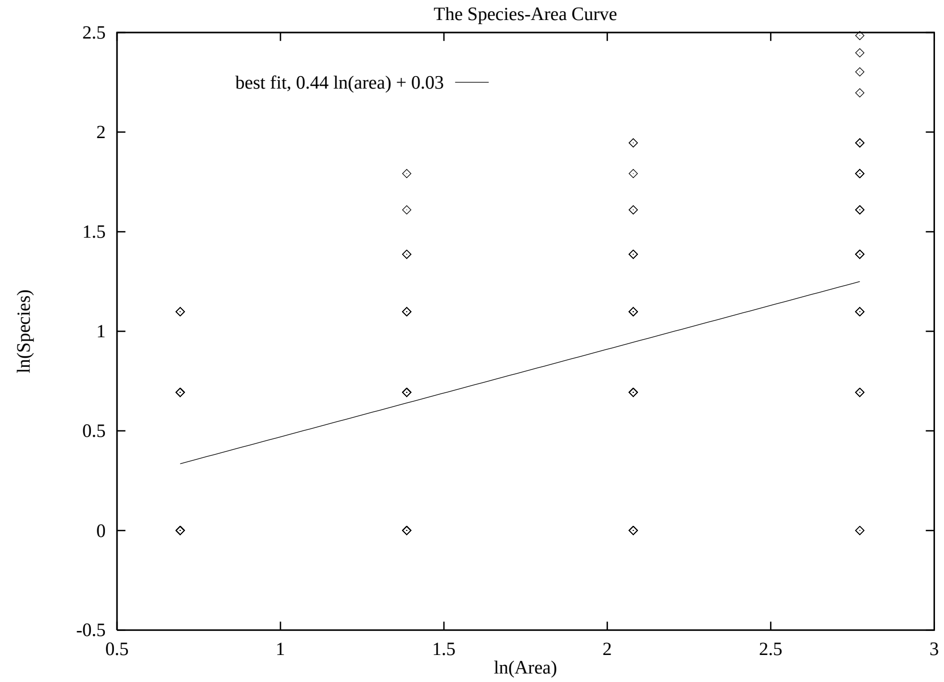


Figure 4-5: The species-area data. The number of patches (area) and the number of species have been scaled by their natural log. Most of the 50 data points for each area value overlap. The line of best fit is $0.44ln(\text{area})+0.03$. The coefficient, $z = 0.44$ is statistically significant ($p < 0.001$), although the intercept (0.03) is not.

4.2.2 Diversity Dependence

These results suggest that there may be some diversity dependence in the processes of diversification. To test for a dynamic equivalent to the diversity equilibrium in island biogeography, we should examine the relationship between biodiversity and speciation events as well as extinction events. The model is able to record the diversity levels as well as the number of new originations and extinctions for every time step. Figure 4-6 shows the observed speciation and extinction rates in the model under a homogeneous environment with 0.9 migratory barriers[6]. The picture from Figure 4-4 has changed, but the presence of a diversity equilibrium has not. Even a slowly increasing speciation rate will, at some point, balance a more rapidly increasing extinction rate to give the observed diversity equilibrium in the model. Figure 4-6 shows the two rates crossing with a standing diversity of somewhere between 5 and 6 species. This is not too far from the mean of 5.04 species observed in the

[6]Since the runs with 0.9 barriers had a higher species diversity than the plain vanilla runs, the accompanying higher variance in species diversity helps us to discern any relationships between diversity and speciation and extinction rates.

model for those runs.

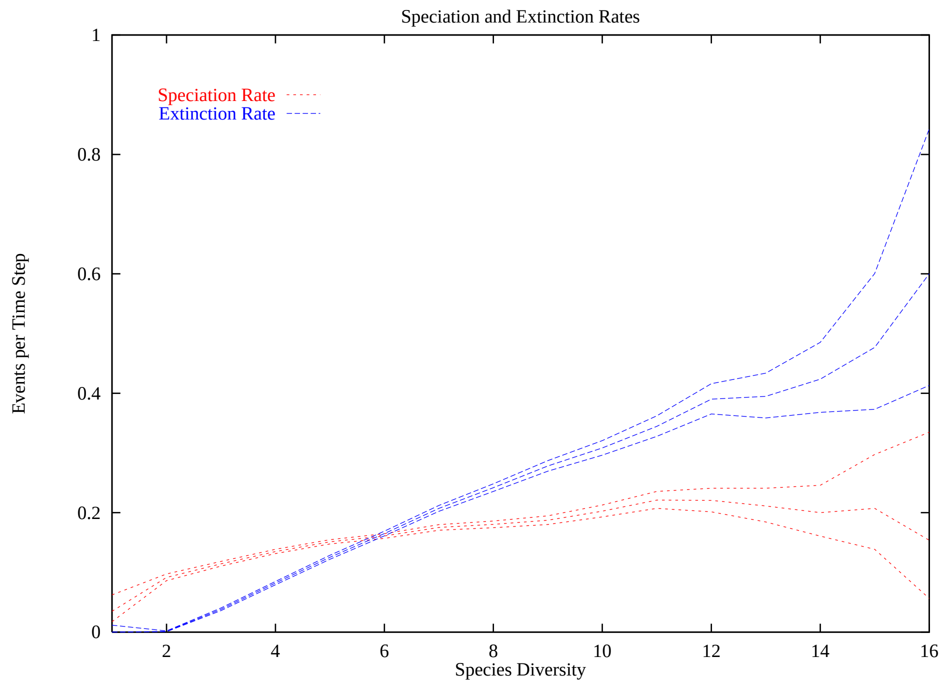


Figure 4-6: Speciation and extinction rates in the model as functions of diversity. This data comes from 50 runs the model with 0.9 migratory barriers in a homogeneous environment. The means for both rates are surrounded by curves representing the 95% confidence intervals for the means. In contrast to the theory of island biogeography, origination rates, i.e., speciation rates, do not decline over the observed range of species diversity. In fact, speciation rates seem to gently increase. However, as long as the speciation rate and extinction rate cross, there will be a diversity equilibrium. At high diversities, species are dying out faster than they are being created. But at low diversities, the opposite is true. The confidence intervals expand at high diversities because the model rarely remained in a high diversity state and so relatively fewer data points could be collected to ascertain the rates.

The phenomena of both speciation and extinction rates rising with low to moderate diversity holds across different parameter settings in the model. Figure 4-7 shows the same data collected for 0.95 migratory barriers and heterogeneous habitats. In this case, the organisms were preferentially mating with other organisms that had similar predation resistance chromosomes. As we shall see in Chapter 7, these factors combine to stimulate intense diversification. Note that even under intense diversification, the extinction rate still eventually out paces the speciation rate as diversity rises.

Valentine (1985) has challenged the conventional wisdom that competition strongly regulates diversity. He proposed a diversity equilibrium model where extinction rates per species is independent of diversity, and only speciation rates per species declines with higher diver-

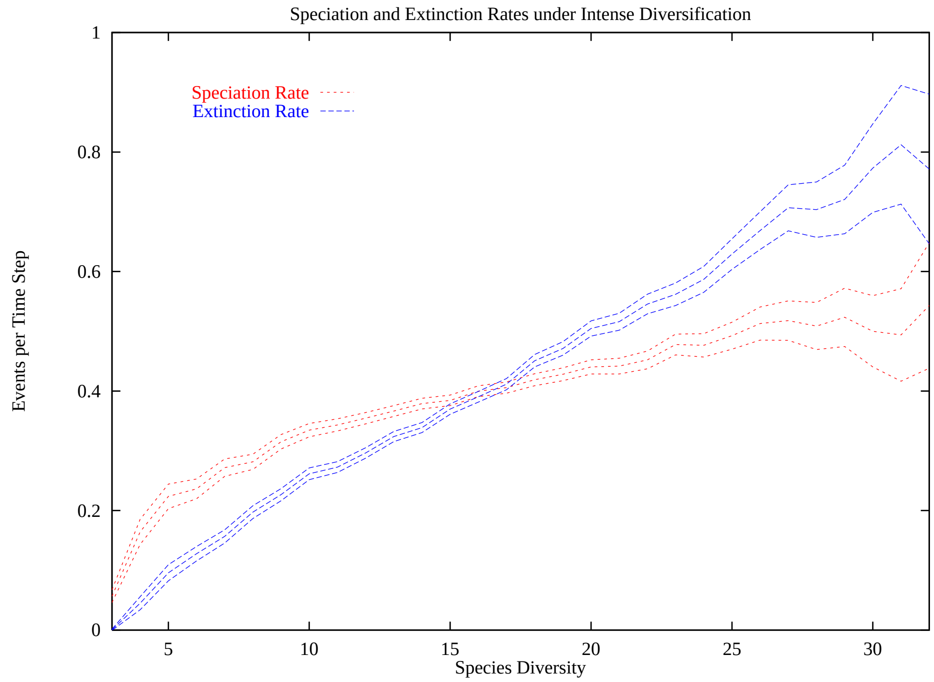


Figure 4-7: Speciation and extinction rates in the model with parameters set so as to maximize animal species diversity. The means for both rates are surrounded by curves representing the 95% confidence intervals for the means. The mean of 16.76 species observed over 50 runs, at the end of 5000 time steps, falls nicely into the equilibrium range indicated by the intersection of the speciation and extinction rates. Again, the confidence intervals widen at high diversities due to the sparse data at those diversities.

sity:

$$\frac{dN}{dt} = \left[a\left(1 - \frac{N}{N_{\max}}\right) - b\right] N \tag{4.1}$$

where N is the number of species, $N_{\max}$ is the maximum number of species sustainable in the environment, a is the intrinsic speciation rate of the group in question, and b is the extinction rate which is determined in part by circumstances independent of diversity. This essentially matches the curves in Figures 4-6 and 4-7. Valentine's data comes from paleontological studies, and so would miss all of the speciations and extinctions of species that never have time to evolve distinctive morphological characters (Larson 1989). Orr (1995) has found that reproductive incompatibilities should accumulate between diverging groups at a rate that is quadratic in the mutation rate. That is, roughly the square of the mutation rate. This implies that reproductive isolation may well precede morphological change in geographically isolated populations, if morphological change is a linear or slower function of mutation rates.

The model suggests that both total extinction rates and speciation rates (per time step, not per species) initially rise with diversity. The data argue for speciation rates more as a gently rising line than the exponentially decaying function of diversity from Island Biogeography. However, the dynamics of speciation and extinction cannot both be linear. At 0 diversity, both speciation rates and extinction rates must logically be 0. If both dynamics were linear, then the two lines would only cross at 0 (assuming the rates are not identical), and so there would be no equilibrium value. However, both Figures 4-6 and 4-7 clearly show a non-zero intersection of speciation and extinction dynamics. On closer inspection, the speciation rates appear to have some curve in the extremes of low and high diversity. If we model speciation rates as a parabola, a quadratic relationship between speciation rates and diversity, the system might be described by Equation 4.2.

$$\frac{dN}{dt} = \left(aN - bN^2\right) - (cN) \tag{4.2}$$

Here cN represents a simple linear relationship between extinction rates and diversity with a 0 extinction rate at 0 diversity. In contrast, $aN - bN^2$ represents a quadratic relationship between speciation rates and diversity. Again, speciation rates must be 0 when there are no species living. As written in Equation 4.2 all the coefficients are positive $(a, b, c > 0)$ and at low diversity speciation rates are greater than extinction rates $(a > c)$.

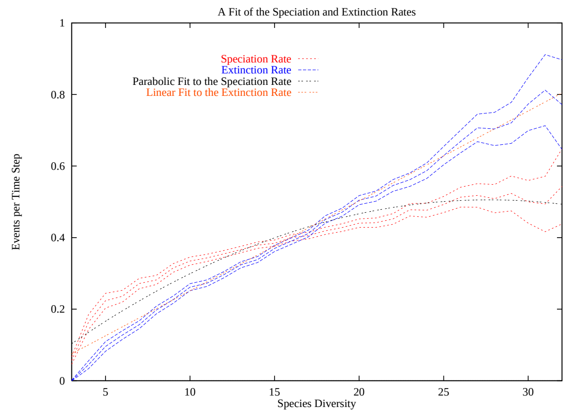


Figure 4-8: A parabola has been fit to the speciation rate curve and a line has been fit to the extinction rate curve from Figure 4-7. They intersect at an equilibrium of 17.27 species, not far from the observed average of 16.76 species observed at the end of 5000 time steps.

A fit of this model to the data from Figure 4-7 is shown in Figure 4-8. This has a stable equilibrium at $\hat{N} = (a - c)/b$ (and a trivial unstable equilibrium at $\hat{N} = 0$). $\hat{N}$ is an equilibrium because when we plug $\hat{N} = (a-c)/b$ into Equation 4.2, we get $dN/dt = 0$. That is, when there are $\hat{N}$ species, the number of species no longer changes. To see that it is stable, consider a small perturbation $\epsilon \ll a-c$ to the diversity such that $N = (a-c+\epsilon)/b$. Plugging this into Equation 4.2 we can see:

$$\frac{dN}{dt} = (a-c-bN)\,N \tag{4.3}$$

$$\frac{dN}{dt} = \left(a-c-b\frac{a-c+\epsilon}{b}\right)\frac{a-c+\epsilon}{b} \tag{4.4}$$

$$= (a-c-(a-c+\epsilon))\,\frac{a-c+\epsilon}{b} \tag{4.5}$$

$$= -\epsilon\left(\frac{a-c+\epsilon}{b}\right) \tag{4.6}$$

Since $\frac{a-c+\epsilon}{b} > 0$ the change in diversity is the opposite of perturbation ϵ. If the perturbation is positive ($\epsilon > 0$), then $\frac{dN}{dt}$ will be negative. That is, the diversity will decrease. So, if we increase N by ϵ, N will then decrease again until it returns to $\hat{N}$. Conversely, if we make ϵ negative, decreasing N, then $\frac{dN}{dt}$ becomes positive and N will rise back up to $\hat{N}$. Equation 4.2 is similar to Valentine's formulation in Equation 4.1, except that Equation 4.2 does not include an asymptote for the speciation rate. Instead, it predicts that speciation rates would eventually decline at high diversity levels.

Fitting a parabola to the speciation rate data from Figure 4-7 results in a significant non-linear component with $b = -0.00066$ and $a = 0.03652$ ($p < 0.001$ for both coefficients with a residual standard error of 0.029). A linear fit of the extinction rate data gives $c = 0.02513$. This predicts $\hat{N} = 17.27$ which is close to the average of 16.76 species observed at the end of 5000 time steps in 50 runs. A logarithmic curve might make more biological sense. The issue hangs on how speciation rates respond to high diversities. If speciation rates decrease at high diversities, then a parabola is probably a better match to the data. However, it is not clear why they should go back down. Certainly extinction rates should increase, but with more species around, it seems reasonable to expect more speciation events as well. If speciation rates continue to increase, then a logarithmic curve might make a better description of the relationship between diversity and speciation rates. The distinction is not crucial for the questions presently under consideration. Since the state of the model rarely ventures into extremely high diversity levels, there is not enough data to argue conclusively for one function over the other. We have examined a parabola simply because it makes a convenient test for non-linearity. If b is significantly different from 0, as it is in this case, then the system is non-linear.

4.3 Percolation and Speciation on a Hypercube

How are we to understand the dynamics of the species in the model? Do extinction rates rise because at high diversities the space of all possible reproductive genotypes gets so crowded with living genotypes that species are pushed together and meld into one? Percolation theory will help to lay down some initial constraints on the dynamics of reproductive genotype space.

There are two important benefits of choosing the reproductive species concept for a definition of a species. First, it can be unambiguously[7] applied to a population of organisms, given enough information about those organisms. Second, its elegance opens the door to a formal analysis of speciation and macroevolution (Gavrilets 1997; Gavrilets & Gravner 1997).

To understand the process of speciation in the model, we need to understand the dynamics of reproductive isolation. The processes of speciation are confined to the genome space of the reproductive chromosomes. Since these chromosomes are composed of 64 bits, this space is a 64-dimensional binary hypercube. The consideration of random graphs in this space stands as a sort of "null" model of evolution. This may help to sharpen our intuition about the dynamics of this space.

The mathematics of random graphs constrained to a spatial geometry is called percolation theory (Grimmett 1989). The central question of percolation theory is, given a probability p that there exists a node at any given coordinate in the discrete space, what is the probability that there is a path between two arbitrary nodes. A path exists if there is a series of adjacent nodes in the space connecting the two end points of the path. In our case, the graphs are formed on a 64-dimensional binary hypercube, so every node has between 0 and 64 adjacent neighbors. In relation to the model, a node represents a reproductive genotype from the current population of the model. The central result of percolation theory is

[7] It should be understood that nothing in biology is truly free of ambiguity. The chance that I might exchange genes with the *E. coli* in my gut is probably non-zero. Does that mean we are part of the same species? Even with complete information about a population of organisms, one might have to choose an arbitrary cutoff in the probability of exchanging genetic material in order to identify reproductive isolation.

that the probability of the existence of a path between an arbitrary pair of nodes undergoes a "phase transition" as p rises. In other words, there is come critical value of p, called p_c, such that, if $p < p_c$ then two arbitrary nodes are not connected (with high probability). When $p < p_c$ the graph is said to be in the "subcritical" state. However, if $p > p_c$ then two arbitrary nodes will be connected (with high probability), and the graph is said to be in the "supercritical" state. For an n-dimensional binary hypercube,

$$p_c = \frac{1}{n-1} \tag{4.7}$$

This is easiest to see by considering the probability of finding a path across the hypercube (Gavrilets & Gravner 1997). Starting at some origin node, we take a step to any neighboring node. For any step along this path, there are n potential neighbors, one of which we just left. Since each potential neighbor has probability p of existing, the expectation is that there should be $p(n-1)$ existing neighbors that we might move to in the next step of the path. So if $p(n-1) > 1$ we should expect that we will never be flummoxed by a complete dead end. In other words, the component in which you travel fills the entire space. By contrast, if $p(n-1) \leq 1$ then, at some point, you are bound to run out of new neighbors to visit and you will find yourself teetering on the edge of the relatively small component that contains your origin node.

In the model, this critical value corresponds to $p_c = 1/63$. In other words, it would take approximately $1/63 \times 2^{64} \approx 2^{58}$ reproductive genotypes randomly scattered about the hypercube before you would expect them to form one connected component. That is, one species. Since the carrying capacity of individuals in the model is 2^{15}, we are nowhere near the critical value p_c. It would take an unreasonably large amount of randomly generated genome diversity before we should expect the constraints of the hypercube to force them into one species.

Other results derived by Gavrilets and Gravner (1997) are also informative as a sort of null model. Most connected components will have $O(n)$ nodes. In the runs of the model, with $n = 64$, 0.99 migratory barriers, and a homogeneous environment, there are an average of 43.79 (standard error = 2.02) nodes per species after 5000 time steps. The 50 data points are shown in Figure 4-9. However, since other chromosome lengths (n's) were not tested, the relationship of connected component sizes to n is unclear.

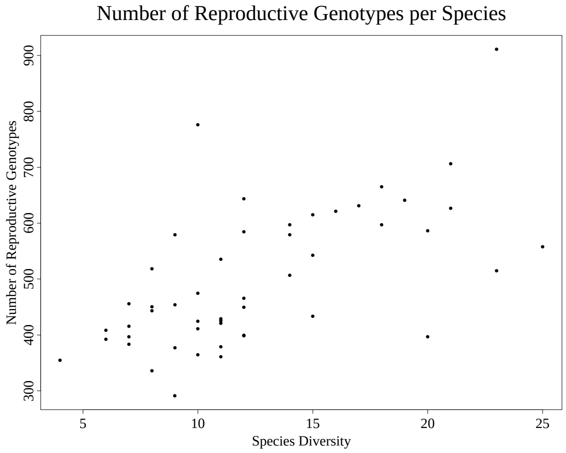


Figure 4-9: The number of reproductive genotypes per species under a homogeneous environment with 0.99 migratory barriers for 50 runs of the model. This parameter combination was chosen because it produced a wide range of diversity values.

These connected components are expected to be far apart from each other. A typical point on the binary hypercube will be $O(n)$ bits different from the nearest node of the closest component. However, the connected components in the model evolve from each other by splitting events on the hypercube. So, unless evolution tends to strongly disperse them, they are probably not randomly scattered about the hypercube as this result assumes. Finally, Gavrilets and Gravner found that typical nodes in the largest component of a subcritical graph tend to be connected by only a single path. The components tend to look like trees, lacking cycles.

The preceding discussion assumes that reproductive genotypes are random, and so the existence of a reproductive genotype is independent of the other genotypes that existed before it. This assumption is violated by the phenomenon of heredity. In the model, new reproductive genotypes arise from extant neighbors on the hypercube, and species arise from the splitting of ancestral species. How do the ecological and diversity dynamics feed back into the process of speciation in the model?

4.4 The Tempo and Mode of Speciation

The issue of the relationship between speciation rates and the current species diversity can be cast into two alternative metaphors for speciation: the eddying stream and the gelatinous blob. One might postulate that at high diversity levels, niche space (not reproductive genotype space) becomes crowded, thus restricting opportunities for new species to exploit empty niches. If we visualize species as streams cascading through time and frequently throwing off little eddies of new incipient species, then we ought to see a divergence between speciation rates and diversity levels. The "niche crowding" hypothesis does not say that speciation rates should go down, only that new species should have a hard time finding a niche in which to thrive. Thus, while speciation rates might remain constant, extinction rates, particularly the extinction rates of new species, ought to increase with diversity levels. An alternative view is that expansion into new niche space occurs prior to the splitting event of speciation. The image might be of a gelatinous mass of individuals, representing a species, oozing and expanding into some neighboring niche space and only then dividing into two separate species. That is, the expansion into niche space is hypothesized to stimulate a speciation event.

4.4.1 Turnover

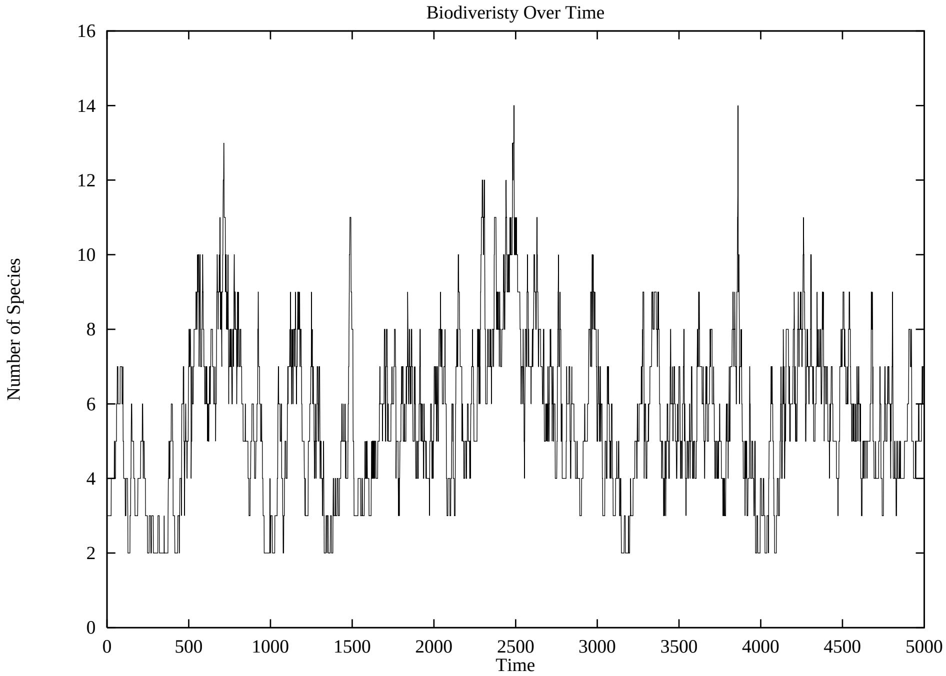


Figure 4-10: The number of species over time from a single run. This is one of the 50 curves that was averaged into Figure 4-2, with 0.9 barriers and environmental homogeneity. Notice that any horizontal lines are very short, indicating that the model did not sustain a particular diversity level for very long.

The evidence from Section 4.2.2 supports the eddying stream view of speciation because speciation rates did not decline with diversity levels[8]. In fact, they rose. We might further contrast the gelatinous blob and the eddying stream metaphors by looking for short-lived species. The eddying stream view sees small populations of individuals being constantly thrown off from the coursing of a successful, populous species. Almost all of these eddies dissipate. Only the rare few find a crevice in the niche landscape through which to squeeze and form the source of a new stream. Yet, if species behave as gelatinous blobs, they would not tend to speciate unless some of the organisms had already colonized a viable niche. New species, under that view, would tend to survive for longer. Figure 4-10 shows a record of the diversity over time. The fact that the curve hardly ever flattens out indicates a constant churning in the biodiversity. It is impossible to have such churning, in a system with only a handful of species, without the frequent occurrence of short lived species.

One way to check for transient species is simply to look at a phylogeny. Does the tree have lots of short branches ending in extinction? An example of a plant phylogeny with the "Christmas tree" shape characteristic of an eddying stream dynamic is shown in Figure 4-11. In this phylogeny, species live for a median of 18 time steps before they either go extinct or split in a speciation event. With no migratory barriers and a homogeneous environment there was an average of 213 speciation events over 5000 time steps but a final biodiversity of only 2.5 species. So only approximately 1% of all new species was still living at the end of the runs. Presumably, this percentage would decrease the longer the model was run and the diversity equilibrium was maintained.

The phylogeny shown in Figure 4-12 was generated under parameter settings designed to maximize diversification. In this case, the median species life span was only 10 time steps before a splitting or extinction event (and a maximum of 491 time steps). In both phylogenies, frequent speciation and rapid extinction is the rule.

Why should we consider a splitting event to end a species' life span? The answer depends fundamentally on one's definition of species.

Consider the small phylogeny in Figure 4-13. A paleobiologist would usually identify one branch of a speciation event as identical to the ancestor because the fossils would probably look the same. Paleobiologists typically only work with fossil data and so must

[8] Figures 4-6 and fig:model-biogeography-maxdiv disagree as to whether or not speciation rates decline at high diversities. The data is sparse enough at high diversity levels that even in Figure 4-6 we cannot be sure that speciation rates decline.

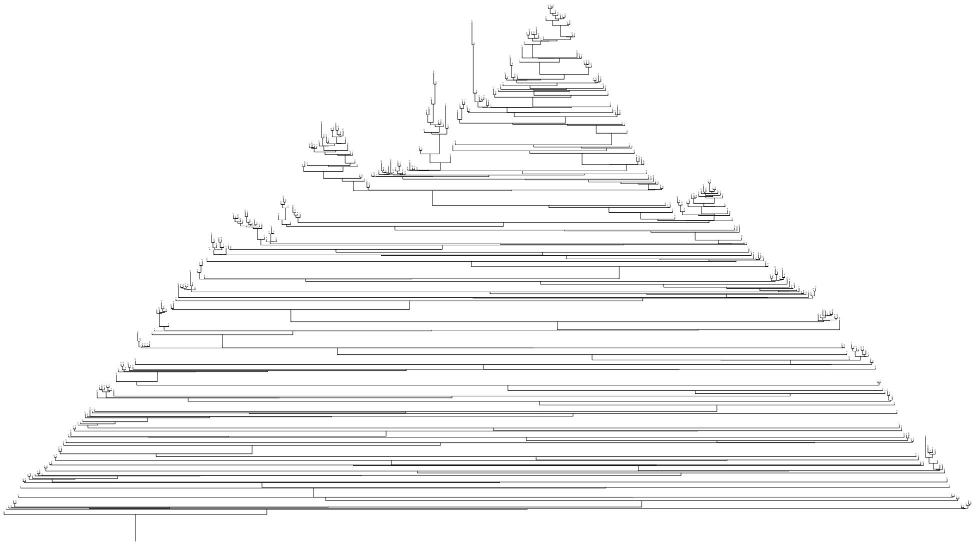


Figure 4-11: A plant phylogeny from a run with 0.9 barriers to migration and a homogeneous environment. Time flows up through the figure. Starting with one plant species at the root, the phylogeny shows many splitting events (the horizontal scale has no meaning), but only rarely do both new species survive for long after speciation. The Christmas tree shape is characteristic of the rapid rise and fall of many species spun off by one or more successful "trunk" species in a phenogram representation of a phylogeny.

use the shapes or morphology of fossils to identify species[9] (Mayr 1942). With knowledge of reproductive isolation, the labeling of only one branch after the speciation event as "ancestral" becomes arbitrary. The two separated populations are equally related to the ancestral species. This is one point where the Systematists depart from the Paleobiologists. When a Systematist reconstructs a phylogeny, the tree is based on the degree of relatedness between the species in question. One can even identify ancestral characters that were preserved by one descendant species but not the other. Still, it is clear that neither of the

[9]The reality is actually worse. In many cases the gross morphology that is easily observable in a fossil is not adequate to reliably distinguish species, or even genera. Paleobiologists generally use families as the unit of measure in the fossil record. The grouping of species into families in Taxonomy introduces a further set of biases to our endeavors to distinguish diversity patterns over time (Robeck et al. 1998).

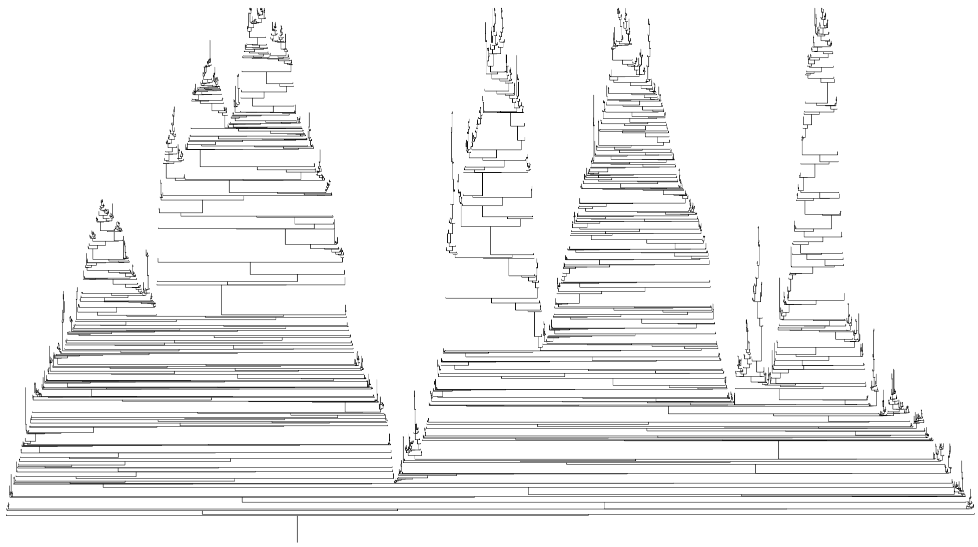


Figure 4-12: A plant phylogeny with maximally diversifying settings for the model. This run had 0.99 barriers to migration, a heterogeneous environment, and positive assortative mating (the tendency to mate with similar organisms) based on predation resistance characters. The Christmas tree shape seen before has diversified into a forest of Christmas trees. Each trunk species can be seen to shed many short-lived species. At the very top of the tree, there are 36 extant species, though they are difficult to resolve by eye.

species after a speciation event is the ancestor of the other. Both descended from a common ancestor. And so the ancestral species, as a coherent entity, disappears with the speciation event (Hennig 1966).

4.4.2 Periodicity

The spikes in the diversity levels shown in Figure 4-10 suggest a possible cyclical pattern to the diversity dynamics that was obscured by the averaging of diversities across runs in Figure 4-2. If the apparent periodicity in the model is more than just a trick of the eye, then there must be something more to the diversity dynamics than the balancing of speciation and extinction rates. A single parameter (one dimensional) analytical model of diversity

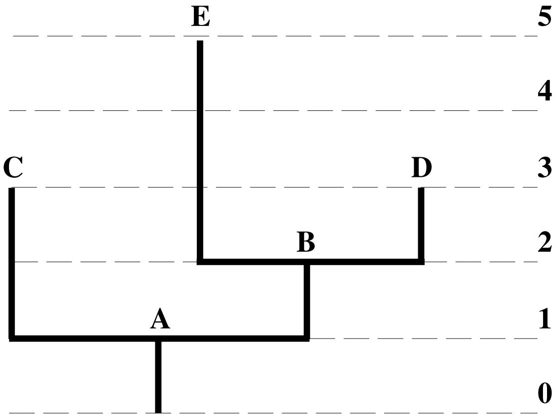


Figure 4-13: The problem of species durations. Is species E 5 time steps old or 3? Did species C and D survive for the same amount of time or did species C live for twice as long as D? To trace a species' origin all the way back to the root of a tree implies that as time progresses the mean species life span must also increase. If we define species by reproductive isolation, and not morphological characters, it is arbitrary and false to say that an ancestral species continues in only one branch after a speciation event. But, by associating the origin of both species with a splitting event implies that the ancestral species has gone extinct when it split.

dynamics, such as Equations 4.1 and 4.2, cannot produce cyclical behavior. A cycle has the property that a single value on one axis (e.g., diversity) can lead to two different behaviors, increasing diversity or decreasing diversity, depending on where the state of the system is on the cycle.

The power spectrum was estimated for each of 50 runs of the model with 0.9 barriers and a homogeneous habitat. This set of runs includes the one shown in Figure 4-10. The power spectrum for that run as well as the average across the power spectra for all 50 runs is shown in Figure 4-14. Each individual run shows strong cyclical behavior with periods on the order of hundreds and even thousands of generations[10]. However, the model does not seem to exhibit a characteristic frequency of cycles. When the power spectra for the 50 runs under the same conditions are averaged together, the spikes smooth out into a curve that increases with the length of the cycles. The presence of these cycles in the model imply that there is at least one other factor interacting with species diversity levels to produce the diversity dynamics. One potential important factor might be the genotypic diversity underlying the species diversity. For any given species diversity, a high genotypic diversity might push the system toward higher species diversity. In contrast, low genotypic diversity

[10]A generation is 3 time steps.

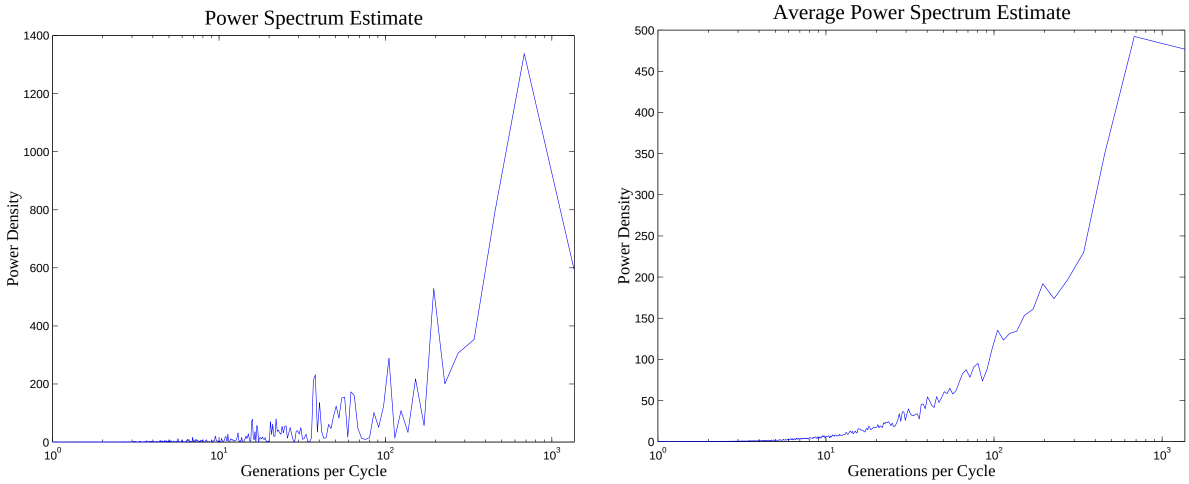


Figure 4-14: An individual (left graph) and the average (right graph) power spectra for diversity over time. The horizontal axis has been scaled logarithmically. The data comes from 50 runs of the model with 0.9 barriers and a homogeneous habitat. The left graph represents the power spectra for the run illustrated in Figure 4-10. The diversity levels over 5000 times steps of a run were analyzed with a power spectrum density estimate using Welch's averaged periodogram method and a window of 4096 time steps. Each individual run exhibits low frequency cycles, as seen in the spikes in the left graph. However, there is no particular frequency characteristic of the model, and so, when averaged across runs, the spikes smooth out. The small spikes in the averaged graph probably come from the inherent limitations of detecting low frequency periodicity. In any given run, we can only sample a low frequency oscillation a few times.

might tend to cause a drop in species diversity. The cause of these cycles remains for the moment an open question.

4.5 Reproductive Barriers

The results that have been presented so far depend on a particular implementation of reproductive barriers. The fact that mating is restricted to one bit differences in the reproductive chromosome constrains the structure and dynamics of the connected components of the reproductive graphs. In most real species, an organism can successfully mate with most other members of its species. The reproductive graph of the organisms is a highly connected graph. We can try to force this into the current representation of reproductive barriers by associating a bit flip in the reproductive chromosome with dramatic genetic changes in real organisms. In this way, most organisms of a real species would share the same "reproductive genotype" under the model representation. However, if the bits of the reproductive genotype represent large scale genetic characters, the mutation rate of 0.008 bits flipped in the reproductive chromosome per new organism is very high. The setting of the mutation rate was a compromise between the speed of evolution in the model (Section 3.4 showed that mutation rates higher than this significantly increase the rate of adaptation) and biological realism.

As an alternative to trying to shoehorn reality into this particular implementation of reproductive barriers, we might consider relaxing the restrictions on mating. Do we still get speciation in the model if organisms can mate with others that have 2 bits different in their reproductive chromosomes? Using a Hamming distance 2 restriction on mating increases the interconnections of the connected components in the reproductive graphs. At the very least we should expect speciation to be less likely when the connected components are more densely interconnected.

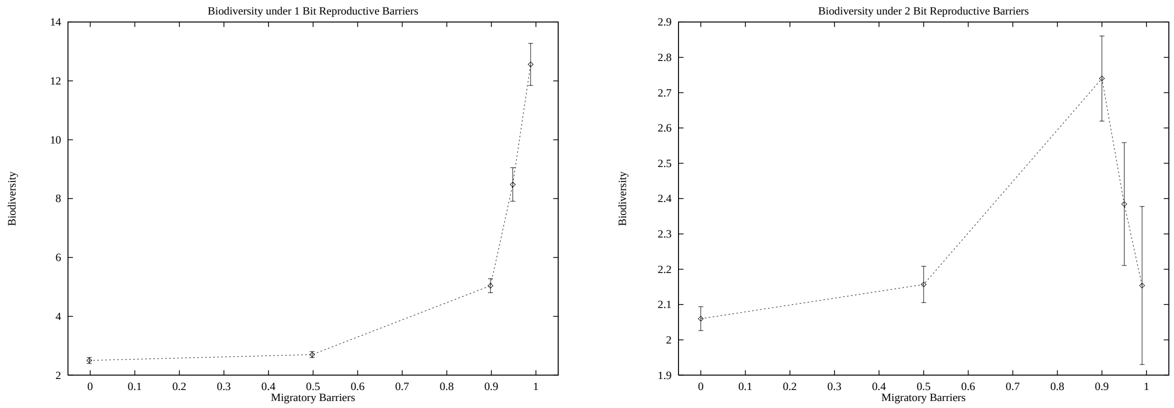
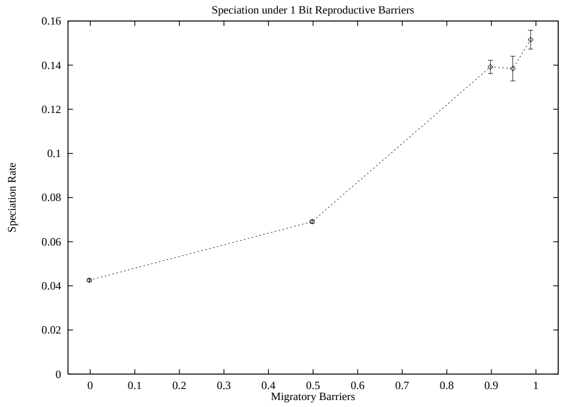


Figure 4-15: Diversity under 2 bit reproductive barriers. The left graph shows the effect of migratory barriers on diversity (with standard error bars) when organisms with at most 1 bit different in their reproductive chromosomes are allowed to mate. The right graph shows the same effect when mating are allowed between organisms with at most 2 bits different. The model was run 50 times under each parameter setting.

Figures 4-15 and 4-16 show that relaxing the restrictions on mating to allow matings between organisms that differ in as many as 2 loci of their reproductive chromosomes still results in speciation events in the model. The data comes from 50 runs of the model under each parameter setting. Diversity levels were recorded after 5000 times steps. There were no differences between the habitats in the patches, so differential selection between patches only arose out of the different genotypes in the populations of the patches.

The gross qualitative behavior of the 2 bit reproductive barriers is the same as the behavior with the more restrictive reproductive barriers. The diversity and speciation rates



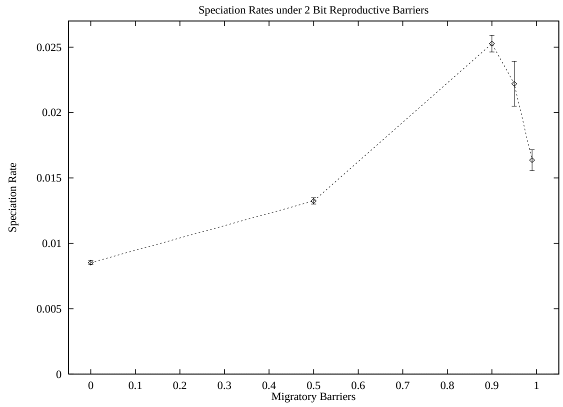


Figure 4-16: Speciation under 2 bit reproductive barriers. The left graph shows the effect of migratory barriers on speciation rates (with standard error bars) when organisms with at most 1 bit different in their reproductive chromosomes are allowed to mate. The right graph shows the same effect when mating are allowed between organisms with at most 2 bits different. The model was run 50 times under each parameter setting.

under 0.9 migratory barriers is significantly greater than the diversity and speciation rates under no migratory barriers ($p < 0.001$). The major difference appears when migratory barriers are high (0.95 and 0.99). At high migratory barriers, the diversity and speciation rates decline for the 2 bit reproductive barriers. The inbreeding, and resulting genetic homogeneity, within the small populations of the patches seems to interact with the density of connections within the gene pools under the less restrictive reproductive barriers to reduce speciation rates and diversity. The exact nature of this interaction remains unclear. The other important difference between the two different mating restrictions is that speciation rates and diversity are lower than the more restrictive case by a factor of about 6. Note that I have not changed the number of loci in the reproductive chromosome or the mutation rate. A fairer comparison might be achieved by doubling the number of loci and doubling the mutation rate when the reproductive barriers are expanded to 2 bit differences. These sorts of explorations must be set aside for future work.

4.6 Conclusions

The attempt to model diversification processes has forced us to adopt a concise understanding of a species. The consequences of restricting the definition of a species to a reproductively isolated gene pool have begun to play themselves out. In doing so, we have shown a dim light into the shadows beyond the lamp post. That light is colored by the particular view of species that we have adopted. So far, three forms have begun to emerge from the darkness.

First, species, as isolated gene pools, are probably being born, and are quickly going extinct, all around us. Though the precise rate of this dynamic depends on the number of genes that influence mating incompatibilities, the mutation rate in those genes, and the degree of mating compatibility between members of a species. Under this view, the true tree of life may be more bushy than we had ever imagined. Such high speciation rates have gone unrecognized until now because these transient species leave behind little morphological evidence of their 15 minutes of fame. Even if a field biologist were lucky enough to sample members of such rare species, she would probably not be able to recognize the fact of its reproductive isolation from its sibling species. Experimental evidence for such transient species will come when we understand the reproductive biology of a species so well that we can efficiently determine which pairings in a well sampled population would produce viable offspring. Given such information, a graph of potential mating relationships could be constructed and the connected components identified. It seems likely that molecular compatibility characters have the best chance of providing such data.

Second, the species-area curve holds under evolutionary time scales where originations are provided by speciation events, not colonization by immigrants from a mainland source community. The organization of individuals into species corrected the unrealistically high exponential rates of genotypic diversity increase found by Jones et al. (1997). I will posit the following prediction. Any reasonable artificial ecosystem that uses components of a reproductive graph to identify species will exhibit realistic exponents in its species-area curve.

Finally, diversity does seem to be regulated. In other words, the model favors the logistic growth description of diversification over the exponential growth description (Walker 1985). Speciation rates seem to rise with species diversity when we adopt the reproductively isolated gene pool definition of species. Yet, even so, an equilibrium is maintained due to the greater increase in extinction rates with diversity. Furthermore, the model quickly achieves an equilibrium level of diversity. If the same is true of the real world, we must ask what factors influence that equilibrium level, and is it possible that the ceiling on diversity might have been raised over time?

Chapter 5

Geography, Habitat, and Productivity

The phenomena that affect the diversification of life can be generally classified into physical, or non-biological, and biological factors. The physical phenomena thought to be important in diversification include changes in geography, climate, and available nutrients. Important biological phenomena include evolutionary innovations that open up the possibility of exploiting new resources, specialization on a small sets of resources or environmental conditions, as well as behavioral interactions within species such as mate choice and sexual selection. This chapter will investigate the impact of manipulations to non-biological parameters on diversification in the model. Chapters 6 and 7 will examine the relative importance of biological phenomena in the diversification of life.

In all of the experiments, there is some baseline condition against which the effects of a manipulation are measured. The default values of the parameters were given in Section 3.6. The most common baseline condition, which I call the "vanilla" condition, was one in which there were no barriers to migration and the environment was homogeneous. That is, all patches had the same habitat, or climate, bit pattern (all 0's). The model is initialized with a plant, a herbivore, and a carnivore species. However, the dynamics of the model are generally dominated by the plant organisms. Why is that?

5.1 The Importance of Being Prey

When two species are bound together in an ecological interaction, they act as important determinants of the selective pressures on one another. In other words, they coevolve. When these interactions are antagonistic, as is the case of a predator-prey interaction, coevolutionary arms races ensue. An adaptation in the prey, making it more difficult to capture and consume by the predator, puts strong pressure on the predator species to evolve an effective response to the prey's defense. Conversely, adaptations in the predator that increase its ability to capture prey, put strong pressure on the prey to evolve further defensive mechanisms to avoid capture. And so it goes, each species upping the ante for the other.

Is there ever a winner? In the model, yes. The stability of the food web was shown under ecological conditions, with no mutation and so no evolution. However, in a single patch simulation, once mutation is turned on, the plants soon evolve predation resistance chromosomes that make them immune to predation and the herbivores starve to death. The predators only have a chance when a multi-patch simulation provides a large enough population base for the predators that beneficial mutations to overcome prey defenses become reasonably likely. This is an evolutionary analog to the ecological "rescue effect" where a local extinction can be reversed by re-invasion of the species from a neighboring patch. Due to the *energy-conversion* parameter, prey populations outnumber predator populations by at least 3 to 1. Prey thus have an evolutionary advantage in their likelihood of discovering useful genetic innovations. This contradicts conventional wisdom in ecology which predicts that individual differences, and their evolutionary effects, tend to stabilize predator-prey dynamics (Crawley 1992). In the real world, the hypothesized advantage of a prey species would be complicated by population sizes and generation times. Small animals that reproduce quickly (like insects) would also gain the evolutionary advantage in numbers that plants enjoy in the model. Both short generation times and high mobility have been implicated in diversification (Marzluff & Dial 1991, for example).

5.2 Barriers to Migration and Habitat Heterogeneity

In the literature that emphasizes physical factors in the diversification of life, most authors have focused on the impact of topography and habitat heterogeneity (Cracraft 1985; Vermeij

1987; Valentine & Moores 1972; Benton 1990; Valentine 1980). Topography, or geography, is primarily important because it can spatially isolate subpopulations. The populations might be divided by a mountain range, a valley, a dessert, a river, an ocean or any other feature that makes migration difficult. All of these features have been abstracted into "migratory barriers" for the model. Once populations are separated, they may genetically diverge either due to random genetic drift or differences in the selective pressures of their separate environments. This brings up the issue of habitat heterogeneity. It seems particularly plausible that subpopulations will diverge if they experience different selective pressures in the form of different habitats. Again, this divergence should be particularly swift if the subpopulations are not exchanging genes through migrant organisms.

The implementation of migratory barriers was described in Section 2.2.3. In brief, a barrier value represents the probability that a new organism fails to emigrate from its current location to a neighboring patch. A heterogeneous environment or habitat was modeled by randomly flipping eight bits in each patch, beginning with all 0's. This results in an expected 5.8 bits set to 1 in a patch's habitat pattern and an expected difference of 10.3 bits between any two patches' climates, or habitats (see Appendix B.1 for a derivation of this difference). In all the following experiments a "heterogeneous environment" refers to this condition of flipping 8 bits in each patch's climate bit patterns.

Figure 5-1 shows the effects of migratory barriers on diversity levels for both homogeneous and heterogeneous habitats. While there is a clear increase in diversity with increasing migratory barriers, the introduction of habitat heterogeneity has surprisingly little impact. An analysis of variance across barrier values of $0, 0.5, 0.9, 0.95$,and 0.99, and habitat variation of 0 or 8 bits flipped in each patch, shows that only the migratory barriers act as a significant determinant of biodiversity ($p < 0.001$). Neither habitat variation, nor the interaction of habitat with migratory barriers are statistically significant factors at the $p = 0.05$ level. T-tests show that habitat variation is only a significant factor in diversity levels when the migratory barriers are 0.9.

Speciation rates seem to be more sensitive to the effects of migratory barriers and habitat heterogeneity, as shown in Figure 5-2. For example, speciation rates at 0.5 barriers are significantly higher than speciation rates with no barriers ($p < 0.001$). In an analysis of variance, both migratory barriers ($p < 0.001$) and habitat heterogeneity ($p < 0.05$) were statistically significant, although their interaction was not ($p = 0.058$). However,

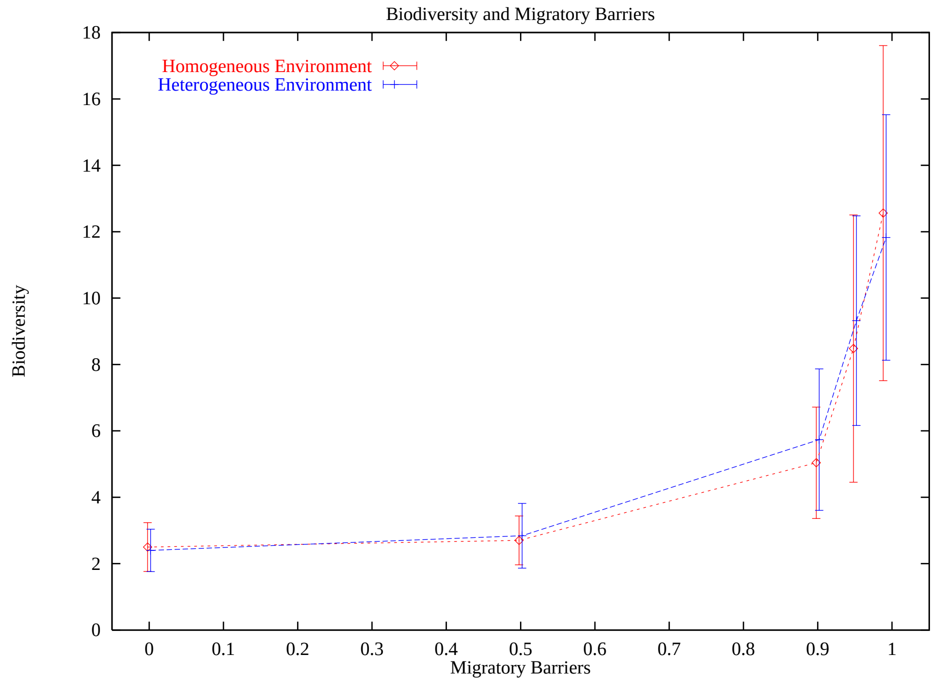


Figure 5-1: The effect of migratory barriers and habitat heterogeneity on biodiversity. The means of 50 runs in each condition are shown along with standard error bars. Migratory barriers have a strong impact on diversity only when those barriers are high. Habitat heterogeneity appears to slightly increase diversity but the effect is not statistically significant.

the apparent crossing of the homogeneous habitat curve and the heterogeneous habitat curve between barriers of 0.95 and 0.99 is significant. At a migratory barrier of 0.95, the heterogeneous habitat incites significantly more speciation events than the homogeneous habitat ($p < 0.001$). Yet, when the migratory barriers are raised to 0.99, the homogeneous habitat leads to a higher speciation rate ($p < 0.001$).

It may appear surprising, particularly to population geneticists, that patches need only be isolated by 0.9 barriers in order to see an effect of habitat heterogeneity. One common result of population genetics is that a very little migration[1] between populations is sufficient to prevent genetic divergence. This is generally measured by the fixation index F which can be interpreted as the chance that two alleles, randomly chosen from a population, are "identical by descent." That is, the two alleles are identical copies of a single allele from an organism that was an ancestor to both of the organisms that were randomly sampled. With 0.9 migratory barriers in the model, we can expect to exchange approximately 5 migrants with a neighboring patch every generation. This means a migration rate of only about

[1] $O(1)$ migrants per generation.

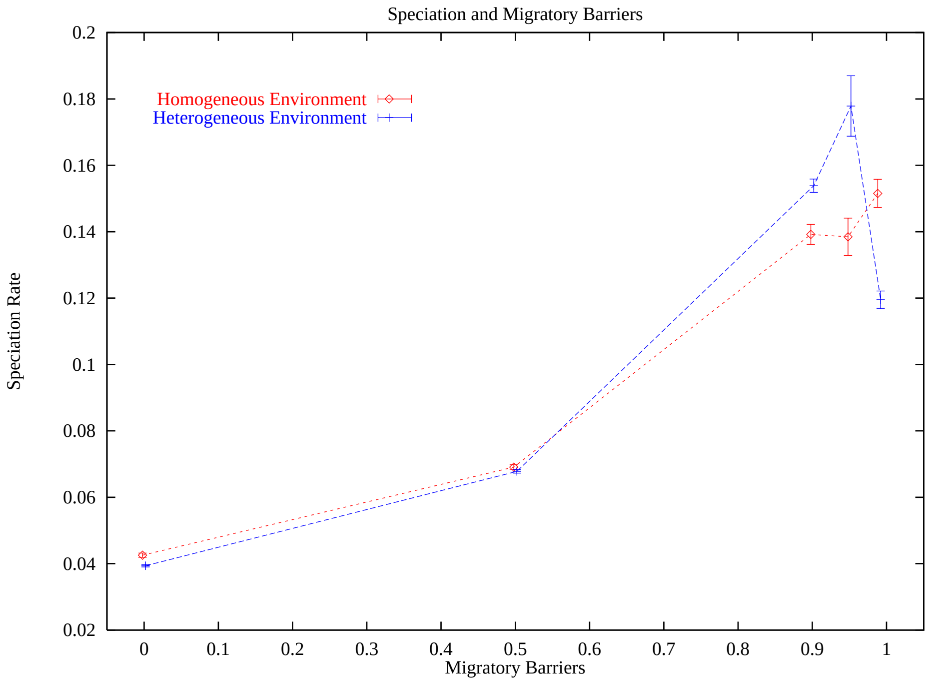


Figure 5-2: The effect of migratory barriers and habitat heterogeneity on speciation rates. Speciation rates also increase with migratory barriers, and even show a response to low (0.5) barriers. Here, habitat heterogeneity has a statistically significant impact on speciation rates.

0.005 between two neighboring patches and a fixation index of $F \approx 0.09$. See Section B.2.3 for a rough derivation of these values. A fixation index of 0.09 is considered to indicate a moderate level of divergence between populations (Hartl & Clark 1997, p.119) and so is consistent with the observed effects.

The combination of migratory barriers and habitat heterogeneity should have an important interaction. Heterogeneity in the environment means that a species that specializes on one habitat may be a superior competitor in that patch but an inferior competitor in another patch that includes the same community of species. The migratory barriers allow the populations of the superior competitors in a patch to expand without losing too many new members to emigration. So the combination of high migratory barriers and habitat heterogeneity allows a variety of species to specialize on a particular patch and coexist comfortably with other species specializing on different patches. It is interesting that this does not emerge as a dominant effect in the model. Even without habitat heterogeneity, the barriers allow reproductively isolated populations to coexist even though they have essentially the same prey search pattern and so inhabit the same niche. A locally dominant population is buffered from competition with rare migrants because the migrants have difficulty finding mates (Bernstein et al. 1985).

5.3 Flux

Some researchers have argued that it is not the mere presence of topographical or climatic heterogeneity that drives diversification. Rather, it is the *changes* in the topography and climate that have stimulated speciation events (Cracraft 1985; Vermeij 1987). These fluctuations might be random or systematic, and where it is appropriate, I have simulated both types of dynamics.

5.3.1 Change in Migratory Barriers

In the previous experiments, the migratory barriers were static. They stayed at the same value throughout a run. The populations of organisms lived within the subdivided environment and, depending on the intensity or height of the barriers, migrants would occasionally slip out of one subpopulation and into a new one. This is a view of organisms acting on top of the environment. What if the environment acts upon the organisms? Some changes in the physical environment may actively divide a population. A change in course of a stream, a forest fire, a falling tree, or other environmental phenomenon might cut off a group of organisms from the rest of a population. Alternatively, geographical changes might also unite previously isolated populations.

Increasing Barriers

Benton (1990) identifies "increased provinciality" as a potential explanation for increasing diversity over time. Given the results of the previous section, this seems reasonable. Certainly when the model was run under higher migratory barriers the diversity increased. Two questions might be posed. First, does an increase in barriers over time lead to an increase in diversities over time? Second, would increasing barriers result in higher diversities than constant barriers?

To address these questions, I let the migratory barriers increase linearly from an initial value of 0 to a final value of 0.99. Figure 5-3 shows the change in diversity levels over time as the barriers increase. Fifty runs of the model resulted in a mean biodiversity of 8.7 and a mean speciation rate of 0.0853 per time step. These are both significantly higher than the biodiversity (2.7) and speciation rate (0.0691) when the migratory barriers are held constant at 0.5. The question arises, is the significant factor here the change in the migratory barriers, splitting apart populations, or do the high biodiversity and speciation rates derive from the final stages of the runs when the migratory barriers were high? Figure 5-3 essentially mirrors the curve of Figure 5-1 and shows that the high diversity levels is only a product of the last part of the run. It is difficult to say whether this is due to the cleaving of populations by the rising barriers or to migrants leaping those barriers and establishing new colonies in neighboring patches. That is perhaps an arbitrary distinction. However, organismal migration generally occurs on a much faster time scale and frequency than geographical changes to migratory barriers. So it is reasonable to say that the presence of the barriers, not their arrival, is the most important factor in diversification.

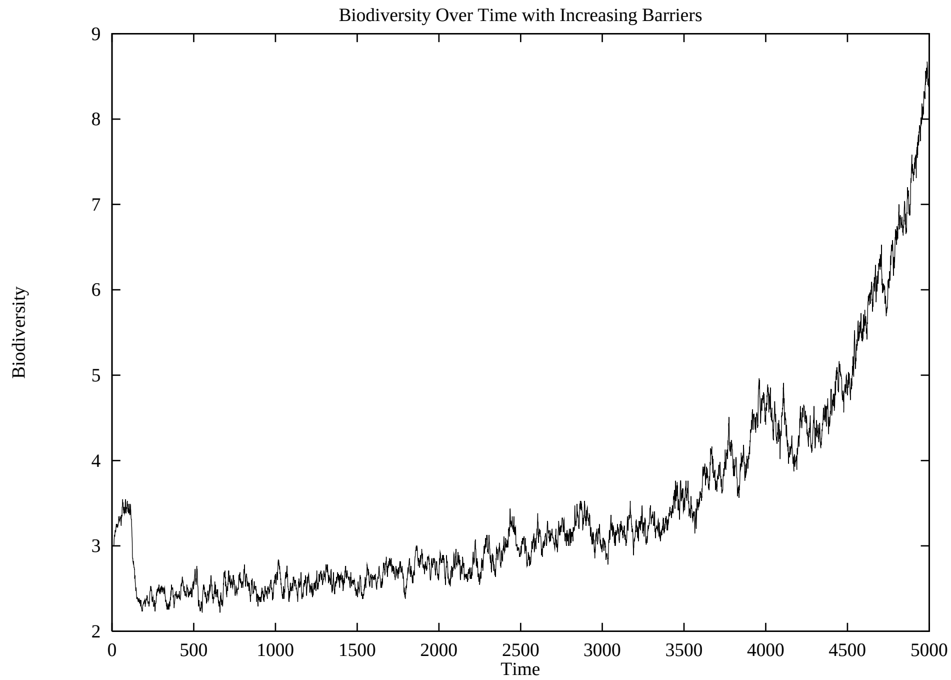


Figure 5-3: The effects of increasing barriers over time on species diversity in a heterogeneous environment.

Figure 5-3 actually shows data averaged over 55 runs with a heterogeneous habitat. This resulted in 8.48 species which was not significantly different from the homogeneous habitat condition with 8.7 species. Speciation rates in the homogeneous (0.0853) and heterogeneous (0.0841) environments were also not significantly different.

Random Fluctuations

Cracraft (1985, p.799) suggests that the important factor in diversification is change in the environment. Unlike the previous experiment, the directionality of that change is unimportant to Cracraft. The environment need not become more fragmented over time, it need only change over time. Perhaps the frequent isolation and then reuniting of populations has a sort of "churning" effect that stirs up diversification. Topographical flux was implemented by allowing the migratory barrier in each patch to individually, randomly change its value by 0.05, either increasing or decreasing by that amount, every time step. In other words, the barrier values of the patches took random walks with 0.05-sized steps. The barriers were not allowed to go below 0 or above 1.

The random walks of the barriers in the patches should average out to a barrier value of 0.5. Thus the constant barrier of 0.5 is the appropriate baseline against which to compare the effect of changing barriers. The introduction of randomly fluctuating migratory barriers generates significantly greater biodiversity (mean = 3.06, $p < 0.05$) than constant barriers of 0.5 which resulted in an average of only 2.7 species after 5000 time steps. Similarly, randomly fluctuating barriers stimulated a significantly higher speciation rate of 0.0782 as compared to a speciation rate of 0.0691 for 0.5 barriers ($p < 0.001$).

Oscillating Barriers

The previous experiment assumed that the migratory barriers in the patches changed independently. This need not be the case. If organisms of a species were constrained to cool climates then long term changes in temperature might alternately confine them to high altitudes when the climate warmed, and release them into the valleys when the climate cooled. The barriers in a region might change in concert. To simulate this, I set the migratory barriers to follow a sine wave with a period of 500 time steps, a minimum of 0 and a maximum of 1, as shown in Figure 5-4.

The resulting biodiversity of 2.74 species was not significantly different from the diversity under 0.5 barriers. However, the speciation rate was (mean = 0.0865, $p < 0.001$). The lack of a significant effect in diversity levels may be due to the fact that all the migratory barriers of the patches were lower than 0.5 for the last 250 time steps of a run. This contrasts with the randomly fluctuating barriers in which at least a few of the 16 patches probably have

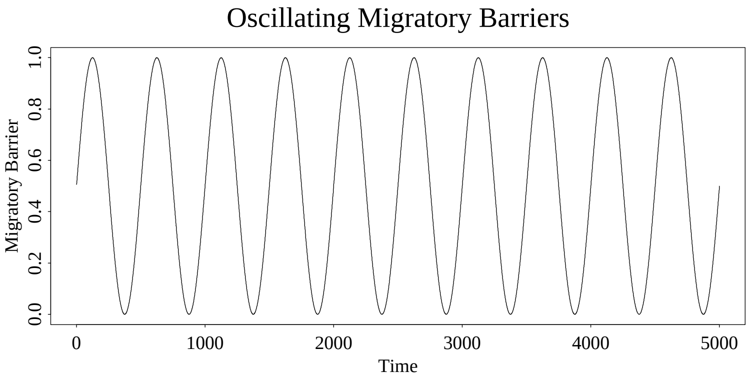


Figure 5-4: The sine wave function for oscillating barriers.

high migratory barriers at any given time.

5.3.2 Habitat Heterogeneity in Time

Cracraft's (1985) emphasis on change was not only directed towards topographical change and migratory barriers. He assumed that most species would be well adapted to their environments and so it would require environmental change, including changes in habitat, to disrupt ecological communities. This disruption might well drive some species extinct but it also might provide new opportunities for other species to diversify.

Random Habitat Fluctuations

I simulated change in habitats by changing the climate bit patterns in the patches over time. All the patches began with the same habitat bit pattern. However, within a patch, and for every time step, there was a $32/5000 = 0.0064$ probability that one bit in that patch's climate pattern would flip. This leads to an expected 32 bit flips in a patch over the entire run of the model. Each patch changed independently of the others. The expected number of bits that differ between any two patches, as a function of time, is plotted in Figure 5-5. This averages out to a 12 bit divergence between patches.

To see if habitat change, and not just habitat heterogeneity, stimulates diversification, we can compare the randomly fluctuation habitat to the "heterogeneous environment." Recall that in the heterogeneous environment the habitat bit patterns in the patches were static though different from each other by an average of 10.3 bits. So we should expect a slightly higher average heterogeneity from the randomly fluctuating habitats. The ran-

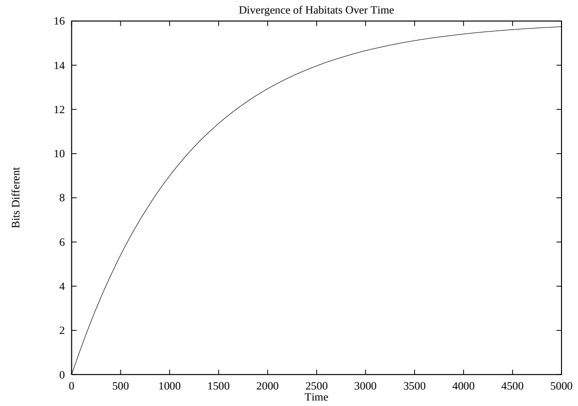


Figure 5-5: The expected number of bits that differ between patches as a function of time, when habitats are fluctuating randomly. This has an asymptote at 16 and a mean value of 12 for the first 5000 time steps.

domly fluctuating habitats result in an average of 2.57 species which is not significantly different from the 2.4 species produced by the static heterogeneous environment. The speciation rates in the two conditions (0.03978 and 0.03939 respectively) are also statistically indistinguishable.

Seasonal Fluctuations

Changes in the climate can play out in both changes to an organism's habitat as well as changes in the productivity of the autotrophs or plants. Productivity in an ecosystem is the rate of energy flow into the autotrophs (Rosensweig 1995). In the model productivity directly corresponds to the *energy-input* parameter. That is, the number of energy units a plant may absorb in a time step if it is perfectly adapted to its habitat. I investigated the effects of fluctuating productivity levels by altering the *energy-input* parameter over time. Specifically, the *energy-input* followed a sine wave with a minimum of 0.2 (a minimum of 0 would have been cataclysmic) and a maximum of 1.8, so as to maintain an average equal to the *energy-input* of the baseline runs. I examined sine waves with a period of 4 time steps, to simulate a sort of seasonal scale of change in productivity. I also examined sine waves with a period of 500 time steps to simulate global warming and cooling dynamics. I will refer to this longer time scale fluctuation as the “ice age” condition. There were no migratory barriers in these runs and the habitats were kept homogeneous. The “vanilla” parameter settings act as the comparison or baseline condition.

Under the seasonal fluctuations in productivity, only 2.3 species survived. This is not significantly different from the biodiversity in the vanilla runs of the model. The diversity dropped further to 2.23 under the ice age conditions. This is significantly lower than the vanilla biodiversity ($p < 0.05$). The speciation rates for both the seasonal fluctuations (0.0381) and the ice age fluctuations (0.0281) are significantly lower than the speciation rates in the vanilla condition ($p < 0.001$). When productivity is low, survival becomes more difficult for the organisms. This probably drives some species extinct, while high productivity periods are not able to compensate with raised raised diversity levels.

Both the seasonal and the ice age fluctuations in productivity were also run with 0.95 migratory barriers and environmental heterogeneity. Perhaps the diversity levels are just too low in the vanilla conditions to see any effects of oscillating productivity. Under these higher diversification settings of the parameters, seasonal fluctuations to productivity have no significant effects on either diversity or speciation rates. The ice age condition also has no effect on diversity, but it has a significantly lower speciation rate (0.0865) than the comparison condition (0.1779) with 0.95 barriers, habitat heterogeneity, and a constant *energy-input* of 1.

5.4 Productivity

The effects of productivity on diversity is a hot topic in Ecology today (Rosensweig 1995; Tilman 1987; Goldberg & Miller 1990). In real ecosystems productivity is often manipulated by altering nutrient levels, rainfall, or sunlight. The basic question has been the relationship between productivity and diversity. However, experiments to investigate that relationship have been hobbled by practical constraints that dictate the manipulation of only small spatial scales over short periods of time. Rosenzweig (1995, p.347) writes:

> [T]he experimental results now available cast little or no light on patterns at larger scales of time and space. None of these experimental results take evolution into account.

The previous investigations into the effects of fluctuating productivity were a little hasty. The more fundamental question is, how are diversity and speciation affected by productivity? This can be tested by running the model under constant *energy-input* levels and measuring the results.

The model was run with *energy-input* values of 1, 3, and 9[2]. A plant may reproduce once for every 3 energy units it accumulates. This means that a plant could reproduce more than once in a time step when the *energy-input* is 9. All three of the *energy-input* levels were run under both homogeneous environments with no migratory barriers as well as heterogeneous environments with 0.95 barriers. The 0.95 barrier value was chosen so as to maximize animal species diversity. While total diversity might have been maximized under 0.99 barriers, such high barriers cause the animal populations to crash. Since productivity in the plants may well have an effect on diversity through the herbivores, 0.95 migratory barriers seemed a more appropriate choice.

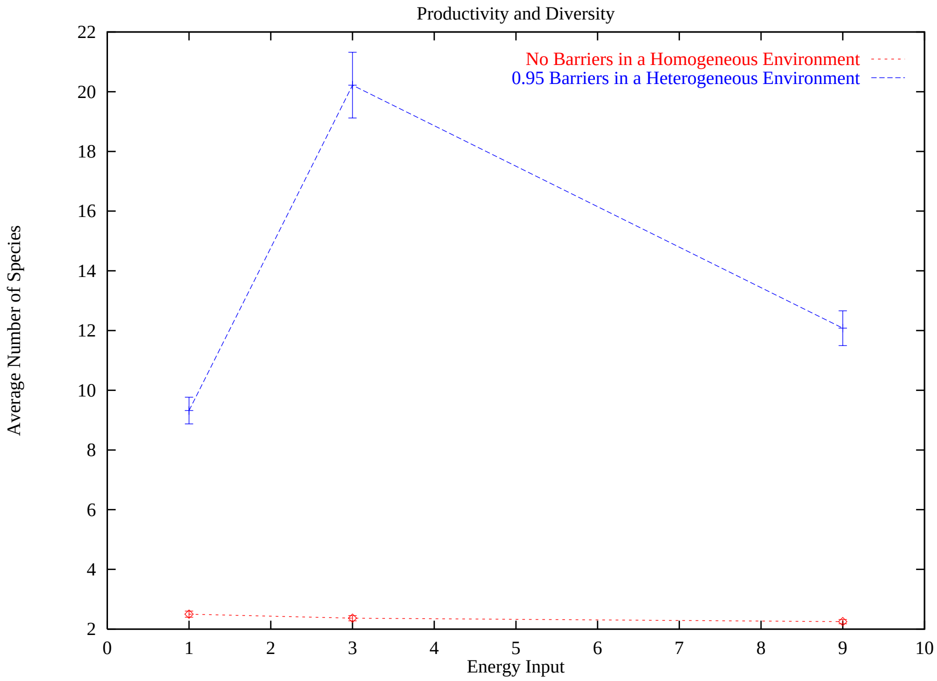


Figure 5-6: The effects of productivity on diversity. The lower curve shows the effects, or the lack of effects, of productivity in a homogeneous environment with no migratory barriers. The upper curve shows that diversity has a non-linear relationship to productivity in a heterogeneous environment with 0.95 migratory barriers. The means are surrounded by standard error bars derived from 50 runs of the model for each parameter setting.

The results for species diversity levels and speciation rates are shown in Figures 5-6 and 5-7, respectively. In a homogeneous environment with no migratory barriers there is no significant difference in diversity levels between the higher productivity conditions and

[2]I chose an exponentially increasing series of parameter settings to maximize the likelihood of finding an effect. Before the experiment, we don't know if the important dynamics only show up at a particular scale in the parameter. The factor of 3 derives from the *energy-conversion* parameter which sets the generation time to 3 time steps.

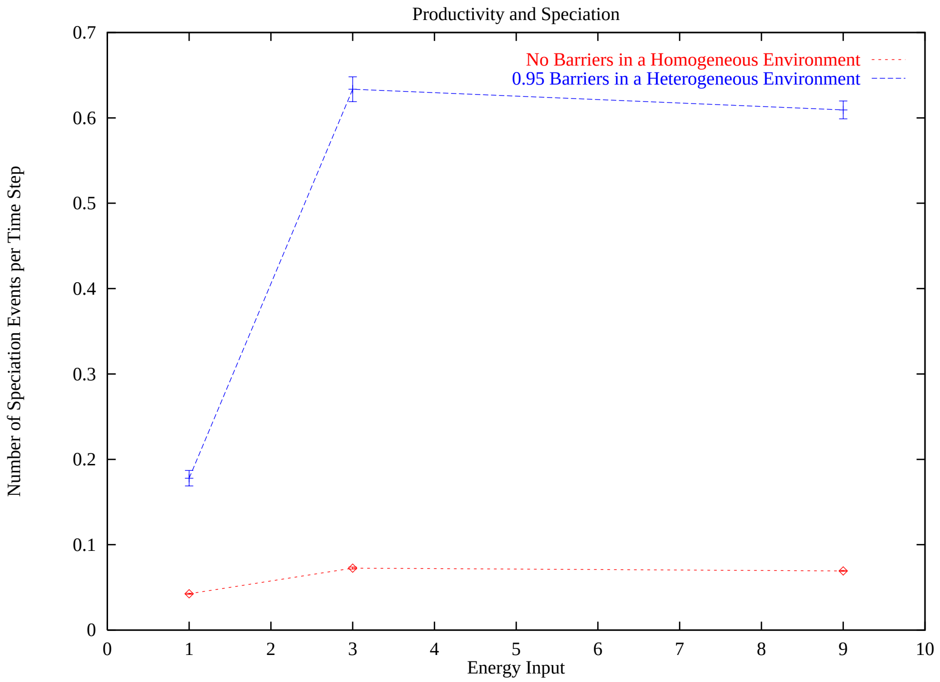


Figure 5-7: The effects of productivity on speciation rates. As before, the upper curve represents the dynamics of a heterogeneous environment with 0.95 migratory barriers. The lower curve has no migratory barriers and a homogeneous environment. The means are surrounded by standard error bars derived from 50 runs of the model for each parameter setting. In all cases, the elevated productivity levels lead to speciation rates significantly greater than the baseline ($energy\text{-}input = 1$). However, for both curves, there is no significant difference between speciation rates with energy inputs of 3 and 9.

the "vanilla" condition ($energy\text{-}input = 1$). However, the lower curve in Figure 5-7 shows that raising productivity did significantly increase speciation rates ($p < 0.001$), though only slightly. When the context is changed to include 0.95 barriers and habitat heterogeneity, the increased speciation rates bear fruit in significantly higher biodiversity levels for the raised productivity conditions ($p < 0.001$).

Productivity seems to have an effect on speciation, but not on diversity levels unless the environment includes heterogeneous habitats and migratory barriers. It is particularly interesting to note that the diversity curve in this case supports Rosenzweig's (1995, pp.345–372) hypothesis that there is a unimodal or "humped" relationship between productivity and diversity at a regional spatial scale.

Rosenzweig (1995) reviews 9 hypotheses for the unimodal relationship between productivity and diversity. Most of these can be rejected as explanations for the results because they depend on factors that do not vary across the experimental conditions in the model.

For example, the amount of area in high and low productivity experiments was held constant, so one cannot argue that low diversity at high productivity is an artefact of the rarity of highly productive areas. Nor can one explain the decrease in diversity due to productivity homogenizing the environment (Tilman 1987). There is no such relationship between productivity and habitat heterogeneity in the model. The model is run over evolutionary time scales, so we can dismiss arguments that are based upon short term ecological disruption. A hypothesized relationship between productivity and disturbance is irrelevant here because disturbance is not modeled. Finally, competition between taxa is also irrelevant because each trophic level begins diversifying from a single species and so is filled by a single taxon. The only remaining explanations are based on arguments that competition must be more intense at intermediate levels of productivity than at either extreme. Yet, the *energy-input* settings of 3 and 9 result in almost identical total population numbers of $20,735$ and $20,725$, respectively. This suggests that the effect is not simply due to competition for space.

The explanation for the unimodal relationship between productivity and diversity rests on the phenomenon of migration in the model. There are two countervailing forces in the model. First off, an increase in productivity, which increases the number of new organisms generated per time step, raises the amount of genetic upheaval. Mutations are tabulated at the point of reproduction[3]. Also, crossover between the parental chromosomes introduces new combinations of genes to the population. Because of the finite carrying capacity of the patches, these new organisms are replacing their parent's generation at a faster rate under high productivity levels. So an increase in productivity should stimulate genetic change and thus diversity. However, an increase in productivity also leads to more migration between patches. As we know from the experiments in Sections 5.2 and 5.3.1, the prevalence of migration between patches has a dramatically negative affect on diversity. To see that productivity affects the amount of migration between patches, first consider the vanilla condition of the model. One third of the organisms, on average, will reproduce in a given time step, because they can collect a maximum of 1 energy unit per time step, and the *energy-conversion* parameter specifies that it requires 3 energy units to reproduce. In the worst case, from the perspective of the migrants, the population is at carrying capacity and only half of the new organisms will survive, replacing other organisms in the patch. So, at

[3]We may assume that mutations actually accumulate in the germ line of an individual throughout its life span. However, these need only be tabulated when one of the mutant germ cells forms a new organism.

the least, 1/6 of the population in the next time step will be composed of new organisms[4]. If we raise the productivity by increasing the *energy-input* to 3, then roughly all of the plants will reproduce in a time step and half of the next generation will be composed of new organisms. With an *energy-input* of 9, the parental population will be swamped with approximately three times their number in new organisms. At some point, the boost to diversity given by the genetic turnover in the populations will be offset by the homogenizing effect of migration between patches. Extreme productivity does not homogenize the habitat as proposed by Tilman (1987). Rather, the subpopulations across the environment become homogenized through migration. This is why we only see the unimodal relationship between productivity and diversity under high migratory barriers.

5.5 Conclusions

Despite claims of the importance of a dynamic environment, barriers to migration overshadows all other factors. The relative effects of all the manipulations can be seen in Figure 5-8. Each of the effects has been normalized against its baseline. The red bars indicate manipulations that result in significantly different diversities or speciation rates relative to the baseline.

The central message in the results is that barriers to migration have dramatic effects on both speciation rates and diversity levels. The results provide little support for the importance of habitat heterogeneity. Speciation rates are much more sensitive to manipulations than diversity levels. In many cases an increase in speciation rates did not lead to a significant change in diversity. This emphasizes again the common conflation of speciation with the success of the new species. Species diversity levels must be a function of both speciation rates and factors that help to maintain those new species.

The comparison of the bar charts in Figure 5-8 is particularly illuminating. The productivity experiments with *energy-input* of 3 and 9 in a homogeneous environment with no migratory barriers results in high speciation rates but has no effect on diversity levels. The increase in speciation rates is probably due in part to larger population sizes under the

[4] In the 0.95 migratory barrier condition that was explored here, most of the new organisms in the next generation are not migrants, due to the high probability that the barriers will prevent them from leaving. Yet, the number of migrants will always be a fixed proportion of the new organisms (in this case approximately 4.9% as derived in Appendix B.2.3). As the proportion of new organisms in a patch increases, so does the proportion of migrants.

high productivity condition. When *energy-input* is 3 and 9, the average total population at the end of 5000 time steps was $20,735$ and $20,725$ respectively. Both of these are significantly higher than the average of $17,507$ organisms produced by the vanilla conditions of the model ($p < 0.01$). The same effect of population size on speciation rates can be seen at the opposite end of the spectrum. When the *energy-input* oscillates over a long period in the "ice age" condition, the average final population size is $13,372$ which is significantly below the vanilla condition ($p < 0.01$), and the speciation rate is also significantly lower. However, this explanation for differences in speciation rates only works for the experiments that manipulated the *energy-input* parameter. The oscillating barrier condition maintained significantly fewer organisms than the constant 0.5 barrier condition ($p < 0.05$), yet it supported a significantly higher speciation rate.

The benefits of isolating subpopulations have long been known in the field of evolutionary algorithms (Tanese 1989; Belding 1995; East & Rowe 1996; Cohoon et al. 1987). When these algorithms are used to try to find optimal solutions to various test problems, a "distributed" algorithm, with individuals segregated into subpopulations along with occasional migration between populations, they tend to out perform algorithms with fully mixed populations. The distributed algorithms typically attain higher maximal fitness values (Tanese 1989; Belding 1995; East & Rowe 1996), as well as maintain greater genotypic diversity (East & Rowe 1996).

The importance of geographical isolation of populations reinforces Orr's (1995) result that populations will quickly diverge and form new species if they are separated. It also supports the dominant allopatric theories of speciation, but without any reference to differences in the selective pressures on the different populations. The central role of migratory barriers in speciation and diversity is not a new discovery. It is thought to be the main explanation for the estimated 800–900 different species of fruit fly amongst the Hawaiian islands (Andersson 1994, p. 209). However, the relative importance of migratory barriers, to the exclusion of all other factors tested thus far, is a startling result of the model. Still, we have not yet begun to dig into the heart of the matter. It is oft said that the most important aspect of an organism's environment is other organisms. It is now time to turn to the relative importance of biotic factors that may influence diversification.

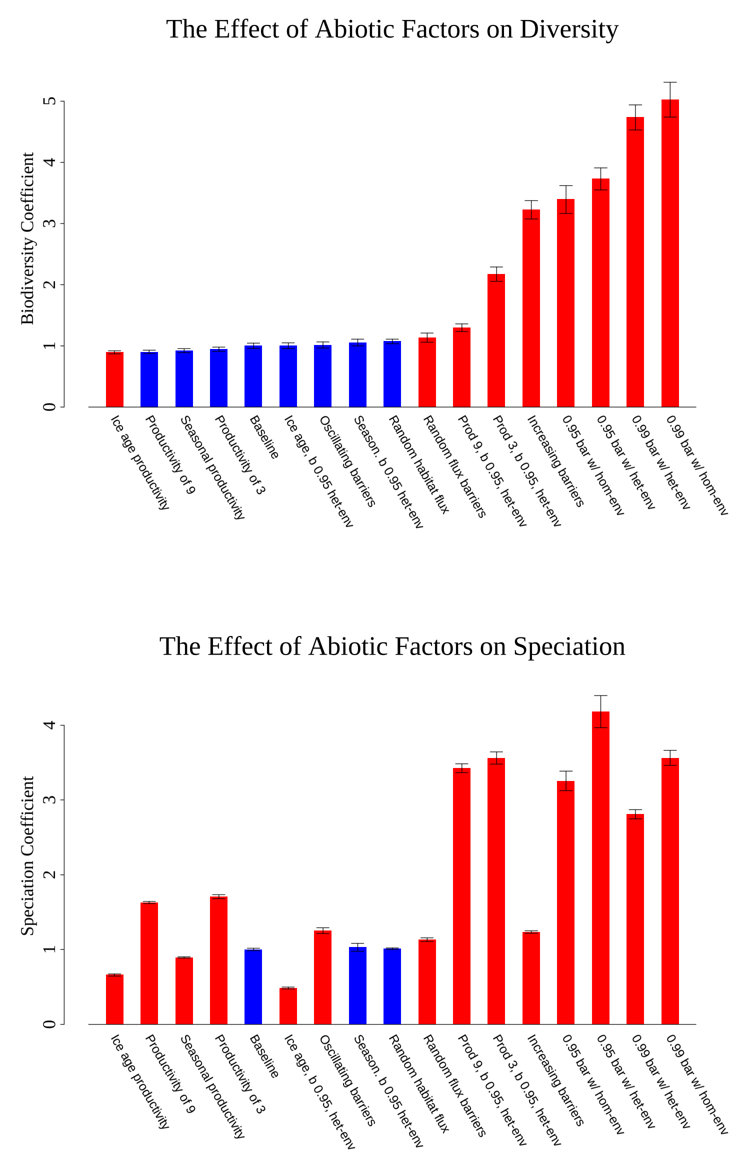


Figure 5-8: A summary of the non-biological effects on species diversity (top) and speciation rates (bottom). Each effect has been normalized by its relevant baseline. The bars indicate the magnitude of an effect relative to the baseline. The blue bars indicate that the manipulation had no statistically significant effect. Red bars indicate a significant effect, with $p < 0.05$. Abbreviations include "b" and "bar" for migratory barrier, "het-env" for heterogeneous environment, "hom-env" for homogeneous environment, "season" for seasonal, and "prod" for productivity.

Chapter 6

What's in a Niche: Specialization and Innovations

As scientists, and perhaps more generally as humans, we tend to look under the lamp post. We focus on what is easily discernible, and often fail to notice that important questions have been left behind in the dark. Even within our pool of light, we act like crows. We are particularly attracted to the bright shiny objects. In the study of evolution this has meant focusing on the evolutionary innovations in taxa as causes of diversification (Slowinski & Guyer 1993). We are drawn to the drama of our creation from the relatively homogenous primordial soup, as well as the flowering of diversity all around us. What could be more compelling in that story of emergence than the innovations that have repeatedly unlocked the gates to a new Eden.

6.1 Niche Space

In the previous chapters, I have often referred to adaptive space. Adaptive space is imagined to be a many dimensional space which maps an organism's phenotype (one dimension for each relevant phenotypic character) onto a fitness value. The fitness value is an additional dimension and is conceptualized to be the "altitude" of a point in space. Thus the language of fitness "peaks" and "valleys."

Ecologists and evolutionary biologists often work with a related conceptual tool, niche space. This is the space of all possible niches. But first, what do we mean by a niche? Colwell (1992) has reviewed the often subtle controversy between two basic interpretations of the term. Under one interpretation, a niche is an aspect of the environment, an "opportunity for survival and reproduction." Colwell calls this the *environmental niche* concept. In contrast, the *population niche* concept identifies a niche as an aspect of a population in relation to its environment. It is, at heart, just "an ecological description of the phenotype of some particular population or species." It is impossible for a *population niche* to be empty. Under the *population niche* concept species do not adapt to a niche, rather they merely adjust their niche. The distinction hinges in part on whether you want to emphasize the action of the environment on the evolution of species (the *environmental niche* concept) or the action of the species to define and shape its own environment (the *population niche* concept). The work of Valentine and Walker, which models the dynamics of taxa filling and relinquishing discrete locations in niche space, implicitly subscribes to the *environmental niche* concept (Valentine & Walker 1986; Walker 1985).

The metaphors of niche space and adaptive space imply the use of the *environmental niche* concept. One might try to force a *population niche* interpretation upon niche space by characterizing how populations flow across that space. But if there is no representation of differences in the viability or fitness in that space, we are simply talking about character space, with special reference to ecological characters.

For any given ecological model, it may be difficult to describe what the viable locations are in niche space, however, the formalizations of the model will bound that space. For example, in our model, we know that the genomes[1] include 160 bits and so there are 2^{160} potentially different organisms. Due to the generalism chromosome, many of these organisms will be ecologically equivalent, though not evolutionarily equivalent. The same mutation in two ecologically equivalent organisms will not produce the same behavior in the mutant organisms. Only a few of the niches in the model can be determined from a description of the initial state of the model. Given the bit patterns of the habitats in the patches, it is easy to describe viable genotypes for plant organisms. However, since organisms eat each other, the success of most genotypes will depend on the current population of both prey and their predators. Furthermore, the composition of these populations evolves over time and so viable niches will shift over time. A niche is thus a characteristic of the environment only if we avoid a naive interpretation of "environment" which fails to consider

[1] Recall that there is no distinction between genotype and phenotype in the model. The bits of the "genome" alternatively stand for the inheritable characters of an organism's phenotype.

the ecological context of an organism.

If we are careful to avoid the conflation of speciation and the survival of the new isolated gene pools (see Chapters 4 and 5), the concept of niche space is a useful tool for thinking about the factors that determine how many species can coexist in the environment. Two major dynamics stand out from this perspective. If you want to feed more guests, you can either cut the pie into smaller pieces, or you can bake more pies. That is, species may specialize on a small fraction of the niche space. This fine grained division of niche space allows the packing of more species into a fixed amount of the space. More intriguing, however, is the possibility that a species may evolve some innovation that opens up vast new tracts of niche space that were previously uninhabited. The descendants of that species are then free to diversify into all those unfilled niches. Think of plants colonizing land, or the evolution of the hard egg shells that allowed reptiles to sever the bonds to water in their life cycles.

Benton (1990) noted a correlation between diversification and an expansion in the habitats and diets of tetrapods (four legged vertebrates). In addition, Benton cites the increased diversity in comparable groups of tetrapods over the same surface area as evidence for increased specialization.

Bambach (1985) has compiled a vast description of the changes in diversity of the classes of marine life.

> [I]ncrease in diversity is achieved only by organisms that develop features that permit them to utilize more ecospace (the multidimensional hypervolume of ecologic resources such as food, mode of life, behavior, physiologic tolerance of ambient conditions, etc.). (Bambach 1985, p.240)

Among the more striking examples of innovations that have allowed the utilization of more ecospace are the evolution of multicellularity, predation, and burrowing behavior to colonize the sediments[2] just below the sea floor.

Ausich and Bottjer (1985) also note a correlation between diversity and "tiering" in their suspension feeding marine communities. Tiering is the vertical range between the deepest burrowers to the highest altitude organisms living in and about the sea floor. Over time this range has extended in both directions. Ausich and Bottjer argue that this is indicative of an

[2]Marine biologists refer to this as an "infaunal" life style.

expansion into new niches as well as more finely subdividing previous niches. However, they explicitly avoid the claim that there is a causal connection between tiering and diversification (Ausich & Bottjer 1985).

These theories depend on the connection between the evolution of ecological traits and the formation of species. Under the reproductive species concept, this implies a connection—perhaps a completely indirect connection—between the genes that code for ecological characters and the genes that code for reproductive characters. If such a connection exists, it is not well understood (Schluter 1996). The assumption is that once the niches have sculpted the ecological characters of population, reproductive isolation from other populations will follow. One might even postulate selection for mutations that prevent an organism from mating with dissimilar organisms. Organisms that try to mix their genes with organisms that lacked their specific ecological adaptations may well produce poorly adapted hybrid offspring. The accumulation of mutations that prevent these hybrid matings is called speciation by reinforcement (Liou & Price 1993). In the following experiments, we will be concerned with the occupation of niches described by the characters that effect resource utilization and predator-prey interactions. The results do not bear on a wider conception of "population" niches that include such characters as reproductive and dispersal behavior or other aspects of an organism's life history.

6.2 Specialization

How has the capacity to specialize, to narrow one's occupation of niche space, affected the evolution of diversity? Specialism may interact with diversification through natural selection against hybrids. If two subpopulations specialize on different resources or habitats, then hybrids from matings between the two subpopulations may be ill fit to survive on either resource (Schluter 1996). This selection against hybrids might isolate the subpopulations enough that genetic drift in the reproductive characters would establish pre-mating barriers and thus cause a speciation event.

In the model, we have a direct handle on specialization through the generalism chromosome of the organisms. By setting portions of this chromosome to 1's, producing wild-card positions in the prey search pattern, and preventing those loci from mutating, we can enforce various degrees of generalism on the organisms. In other words, we can "clamp" a portion of the generalism gene such that it always stays "on." This may be compared to the "vanilla" condition of the model. The change from the restricted condition to the vanilla condition shows how the increased ability to subdivide niche space influences diversification.

The model was run with the upper 10 bits of all the organism's generalism chromosomes fixed to 1's. This prevented specialization in that range of the genome. The results listed in Table 6.2 includes data for runs with no migratory barriers in a homogeneous environment as well as runs with 0.95 migratory barriers and a heterogeneous environment. These results, like all of the results exhibited in this chapter, are the averages of at least 50 runs of the model under each set of parameter values.

Condition	**Species Diversity**	**Speciation Rate**
Restrictions on specialization	2.62	0.04349
Allow specialization	2.50	*0.04254
Restrictions on specialization with 0.95 barriers and hetero-environment	6.70	0.20154
Allow specialization with 0.95 barriers and hetero-environment	***9.32	*0.17786

Table 6.1: The results for allowing specialism to evolve. The * indicates statistical significance at the $p < 0.05$ level. The *** indicates significance at the $p < 0.001$ level. There are significant effects in the heterogeneous environment condition with 0.95 migratory barriers for both an increase in biodiversity ($p < 0.001$) and a decrease in speciation rates ($p < 0.05$). In the vanilla condition, the only significant effect of allowing specialization is a slight reduction in speciation rates ($p < 0.05$).

Specialism seems to have a particularly strong effect on diversity levels when the population is subdivided by barriers to migration and the patches have different habitats. Diversity is boosted despite the slight drop in speciation rates. This means that extinction rates dropped even further than speciation rates. A decline in extinction rates is consistent with the common ecological dynamic of species avoiding competition by specializing on relatively unexploited resources. Though it should also be noted that specialization can also make a species more vulnerable to extinction due to changes in the available resources.

6.3 Evolutionary Innovations in Niche Occupancy

A similar experiment may be performed to examine the effect of allowing the organisms to explore previously restricted areas of adaptive space. By preventing mutations in the middle 10 bits of the prey template and generalism chromosomes, we prevent organisms from being able to exploit any resources in that area of adaptive space. Similarly, by preventing mutations in the predation resistance chromosome, we prevent organisms from escaping predation in that area of adaptive space. In combination, these restrictions prevent the exploitation of novel resources as well as the escape from predation pressures. Note that in a heterogeneous habitat, some of the bits of the habitat bit patterns will be set to 1 in this restricted area. This means that while the clamps are on, plants will not be able to exploit the full range of resources in their environment. The innovations that effect resource utilization and predation interactions are only a subset of the evolutionary innovations that may affect diversification. The results from Chapter 5 imply that innovations that affect dispersal should have important consequences on the diversification of a species. However, here I am seeking to test the hypothesis that expansions in niche space occupancy, the exploitation of previously unutilized resources and the avoidance of predation, increases diversification.

Once again, I compared the results of evolution under these restrictions against the baseline conditions with no restrictions on adaptive space. Table 6.3 shows the data for runs under a homogeneous environment with no migratory barriers as well as runs under a heterogeneous environment with 0.95 migratory barriers.

Condition	**Species Diversity**	**Speciation Rate**
Restrictions on innovation	2.42	0.04348
Allow innovation	2.50	0.04254
Restrictions on innovation with 0.95 barriers and hetero-environment	8.67	0.21515
Allow innovation with 0.95 barriers and hetero-environment	9.32	***0.17786

Table 6.2: The results for allowing evolutionary innovations in predation interactions and resource utilization. *** indicates a significant difference from the baseline with $p < 0.001$. The only effect occurs under a heterogeneous environment with 0.95 migratory barriers. In this case, evolution in the expanded adaptive space results in a suppression of speciation.

The only observed effect appears when migratory barriers are high and there is habitat heterogeneity. In this case, releasing restrictions on innovations significantly reduces speciation rates. The change in species diversity levels is not significant. It seems that allowing the exploration of untrammeled niche space does not stimulate diversification, but may even suppresses it.

6.4 Extreme Restrictions on Niche Space

The preceding results are based on restricting 10 out of 32 bits in each of the ecological chromosomes. Perhaps the loss of 10 bits just is not particularly restrictive to the evolution of the ecosystems. We are dealing with only an average of 9.32 species in the most diverse case. Perhaps the set of 2^{22} possible genotypes for each ecological chromosome is large enough that the restrictions on innovations in predation interactions never really impact the evolution of the species. Would we see an effect if niche space were to be drastically reduced? A pair of experiments were run under a heterogeneous environment with 0.95 migratory barriers. However, this time, 28 of the 32 bits in the ecological chromosomes were fixed. In the case of the specialization experiment, 28 bits of the generalism chromosome were fixed as 1's. In the case of the predation innovation experiment, 28 bits of the predation resistance chromosome, the prey template chromosome, and the generalism chromosome were fixed as 0's. Thus, each restricted chromosome could only take on a possible 16 different states.

Condition	**Species Diversity**	**Speciation Rate**
Extreme restrictions on specialization with 0.95 barriers and hetero-environment	6.92	0.20096
Allow specialization with 0.95 barriers and hetero-environment	***9.32	*0.17786
Condition	**Species Diversity**	**Speciation Rate**
Extreme restrictions on innovation with 0.95 barriers and hetero-environment	8.08	0.212644
Allow innovation with 0.95 barriers and hetero-environment	9.32	***0.17786

Table 6.3: The results for extreme restrictions on niche space. The data come from running the model with 0.95 barriers to migration and a heterogeneous environment. *** indicates a significant difference from the baseline with $p < 0.001$ and * indicates $p < 0.05$.

Table 6.4 shows the results for the extreme restrictions on niche space. The extreme restrictions have the same effects as the mild restrictions. Restrictions on specialism reduces diversity but increases speciation rates. While, restrictions on innovations in the predation interactions has no significant effect on diversity levels yet significantly boosts speciation rates.

6.5 The Cost of Generalism

It is time to revisit an early assumption. Perhaps the reason restrictions on innovations in niche space have so little, or even a negative, impact on diversification is that there is no advantage to specialization in the model. This seems counterintuitive, and indeed, for most of the history of ecology we have assumed that generalists must be worse at exploiting resources as compared to a specialist. Yet, this assumption is controversial.

> Although intuitively appealing, this concept of a reduction in efficiency with niche width has proved notoriously difficult to confirm empirically. Indeed, where a general trade-off between diet breadth and efficiency has been investigated, experimental confirmation has been relatively unusual. (Dall & Cuthill 1997)

The debate has shifted to searching for trade-offs between specialism and generalism that have gone unnoticed in these efficiency studies. Dall and Cuthill (1997) argue that generalists should not be able to track all of the relevant ecological variables of their different habitats as well as a specialist that remains within its specialty. In addition to the efficiency of resource utilization, other important ecological pressures include the avoidance of predators, the location and courtship of mates, and direct competition with other organisms in the environment.

Before we add an artificial trade-off to the model, we should be clear about what it would tell us. Up until now the results argue that with no trade-off between generalism and other beneficial traits, the capacity to expand into unexplored niche space does not stimulate diversification. If the addition of a trade-off changes this, such that innovations do have a significant impact on diversity, we can only argue for their importance predicated on an assumption that has resisted empirical confirmation. While innovation of ecological characters may seem like an intriguing, and even compelling factor in the evolution of diversity, it rests on a rather shaky limb that the empiricists have been busily sawing.

Selection for specialization of a preference for a resource does not require increased efficiency in one habitat at the expense of another. It only requires that some traits vary in their effects across habitats (Fry 1996). For example, the proboscis of a mosquito may be a good instrument for extracting blood from many kinds of animals. However, if it is particularly effective for extracting blood from a human, then that species of mosquito will be selected for specialization on human hosts. Dall and Cuthill's (1997) discussion of the costs of generalism caution us that we need not be simplistic in the design of a trade-off between generalism and specialism. The costs of generalism may play out in increased susceptibility to predation, the consumption of suboptimal prey, and reproduction, in addition to the more traditional view of a loss of efficiency in resource utilization.

The choice of implementation for a trade-off between generalism and some other benefit is, as usual, somewhat arbitrary. Furthermore, the simplicity of the model imposes its own set of constraints. Four possible elaborations of the model suggest themselves. First, we might make it more difficult for generalists to find a mate. Second, we might make it easier for predators to locate generalists. Third, we might raise the threshold of "meals" required for a generalist to reproduce. This would effectively lower the efficiency with which a generalist transforms prey matter into generalist matter. And finally, we might make it more difficult for generalists to gather resources.

The first suggestion, that generalists have a harder time finding a mate, lacks a biological motivation and seems exceptionally arbitrary. The second suggestion, that generalists are exposed to predation is at least biologically motivated. A generalist would have to defend against predators in all of the habitats in which it thrives. It is probably difficult to evolve cryptic coloration or chemical defenses that work well across a heterogeneous environment. However, since most of the diversification in the model happens at the autotroph level, and herbivores do not generally thrive, it is unclear that predation represents a strong enough selective pressure to have much of an effect. The third suggestion is to raise the *energy-conversion* parameter for generalists from 3 to some higher number. While this need not be an integer, by allowing reproduction with probability proportional to any fractional part, this seems like an awfully crude hammer with which to hit the generalists. In comparison, the fourth suggestion seems better in that it also changes resource utilization efficiency, but in a more flexible and subtle manner.

Consider penalizing a generalist by reducing the chance of consuming a prey organism in proportion to the number of generalism bits that were used to match the prey's predation resistance chromosome. Under this suggestion, a generalist would suffer no penalty if its prey template chromosome exactly matched the prey's predation resistance chromosome. This meets Fry's (1996) criterium that a trait should vary in its effect across habitats or resources. More interestingly, it spans all of the three genetic effects on generalism that Thompson (1994) describes. Recall that a gene that allows the utilization of a new resource may have one of three effects. It can act independently of genes for utilizing other resources. In the model, the addition of a wild-card bit to an organism's generalism chromosome allows the utilization of new resources without affecting its utilization of other resources. A gene may also positively contribute to the effects of other genes for resource use. This is a generalization of the previous effect. In the model, a single bit mutation may allow the utilization of one or many new resources depending on the nature of the resources. For example, if two species of prey both tend to have 1's in a particular locus of their predation resistance chromosomes, then a mutation to a 1 in a predator's prey template gene would facilitate the exploitation of both of those species. Finally, a gene can interfere with the effects of other genes for resource utilization. Flipping a bit in the prey template chromosome, depending on the resources available in the environment, may reduce the predator's ability to utilize some resources at the exchange of facilitating the use of others.

A trade-off was introduced between generalism and the efficiency of capturing and utilizing food sources. Recall, from Equation 2.2, that the probability that a predator consumes a prey organism is a function of the match between the predator's search pattern and the prey's predation resistance characters. A trade-off was implemented by penalizing the match of a generalist predator to its prey by half a bit for every wild-card position in the predator's search pattern that was used. A wild-card position was considered useful if that loci would not have matched the prey in the absence of the wild-card bit. For example, consider a single locus in the ecological chromosomes. If a predator had a 1 in that position of its generalism chromosome (a wild-card) and a 0 in that position of its prey template chromosome, then if it tried to capture a prey organism with a 1 in that position of its predator resistance chromosome, the predator would only score 0.5 for the match at that position. However if the predator had a 1 in both its generalism and its prey template chromosome, then it would have scored a full 1 bit match at that position. So a generalist only suffers a disadvantage when its generalism is relevant to its survival. This is a fairly weak form of a trade off. A generalist can be just as good as a specialist at utilizing a resource if they have identical prey template chromosomes. However, now a generalist cannot be as effective as two different specialists. The generalist will suffer a penalty when trying to exploit any resource that does not exactly match its prey template.

Condition	Species Diversity	Speciation Rate
Restrictions on specialization with 0.95 barriers and hetero-environment	5.58	0.08158
Allow specialization with 0.95 barriers and hetero-environment	***9.32	***0.17786
Condition	**Species Diversity**	**Speciation Rate**
Restrictions on innovation with 0.95 barriers and hetero-environment	10.48	0.25312
Allow innovation with 0.95 barriers and hetero-environment	9.32	***0.17786

Table 6.4: The results for extreme restrictions on adaptive space with a trade-off between generalism and prey capture. The data come from running the model with 0.95 barriers to migration and a heterogeneous environment. *** indicates a significant difference from the baseline with $p < 0.001$.

Once again, the same comparison was made to the results of the model with no restrictions on specialism or innovation. However, this time, in an attempt to up the ante so as to find *any* effect, 28 loci of the 32 bit chromosomes were restricted. The organisms could only freely evolve in 4 of the 32 bits in their ecological chromosomes. However, there is a difference this time when specialism is restricted. Previously, when 10 bits of the generalism chromosome were clamped at 1, mutations in the predation resistance chromosomes at those 10 loci had no effect. Now, given the trade-off between generalism and specialism, the states of those bits in both the predation resistance and the prey template chromosomes matter. To maximize the chance of finding an effect, the model was run with 0.95 migratory barriers as well as a heterogeneous habitat. The results are given in Table 6.5. The combination of trade-offs and these extreme restrictions on specialization dramatically reduces diversification in the model. Or, vice versa, allowing specialization increased both biodiversity and speciation rates. Yet again, the opening of adaptive space for evolutionary innovations in resource utilization and predation interactions only reduces speciation rates and has no effect on diversity levels. In summary, the addition of a trade-off between generalism and prey capture, along with extreme restrictions on adaptive space do not produce any qualitatively different results from the mild restrictions on adaptive space examined in Sections 6.2 and 6.3.

6.6 Heterotroph Diversities

Restrictions to adaptive space can have an interesting effect on the coevolution of predators and prey. When there are no restrictions on adaptive space, very high barriers (e.g., 0.99) lead to the extinction of the animals. It seems that the small population of animals in a patch is not adequate to the task of tracking the evolution of plants as they change their predation resistance chromosomes. When the barriers are lower, the animals have a combined population size large enough that mutations will arise to counter the plant's adaptations in the evolutionary arms races.

The average biodiversity with no restrictions on adaptive space and 0.99 barriers is 11.83. All of those species are plants. The average biodiversity with the same barriers but mild restrictions on adaptive space (preventing mutations in 10 loci of the predation resistance, prey template, and generalism genes) leads to the evolution of a whopping 37.51 extant species. Needless to say, these two conditions are significantly different ($p < 0.001$). Yet if we look at this condition more closely we see that only 10.63 out of 37.51 species are plants[3]. There is no significant difference between the number of plants in the two different conditions. It is the remaining average of 26.88 animals that makes all the difference. The animals thrive under mild restrictions in adaptive space. This effect makes it appear that an expansion of adaptive space dramatically lowers species diversity. In contrast, the animals do not die out when migratory barriers are set to 0.95. And so the comparison between the conditions with and without restrictions on adaptive space, under 0.95 migratory barriers, is a better indicator of the general dynamics of the ecosystem.

6.7 Conclusions

Figure 6-1 summarizes the results from all the experiments on specialization and evolutionary innovations in resource utilization and predation interactions. The first thing to notice is that there were no parameter settings under which the availability of additional niche space significantly increased diversification. In fact, allowing the populations to ex-

[3]On a few occasions the model recorded more extinctions than speciations amongst the animals. However, the total number of speciations minus the extinctions matched the final biodiversity. Clearly something went wrong in the labeling of the species. This is a bug, also known as a "feature" in computer science, in the algorithm for tracking the trophic level of the species. It is a bug in the algorithms that record the data, not in the algorithms that produce the data. I have excluded these cases from my analyses of animal diversities. Yet, the results that distinguish plants from animals should be treated with caution.

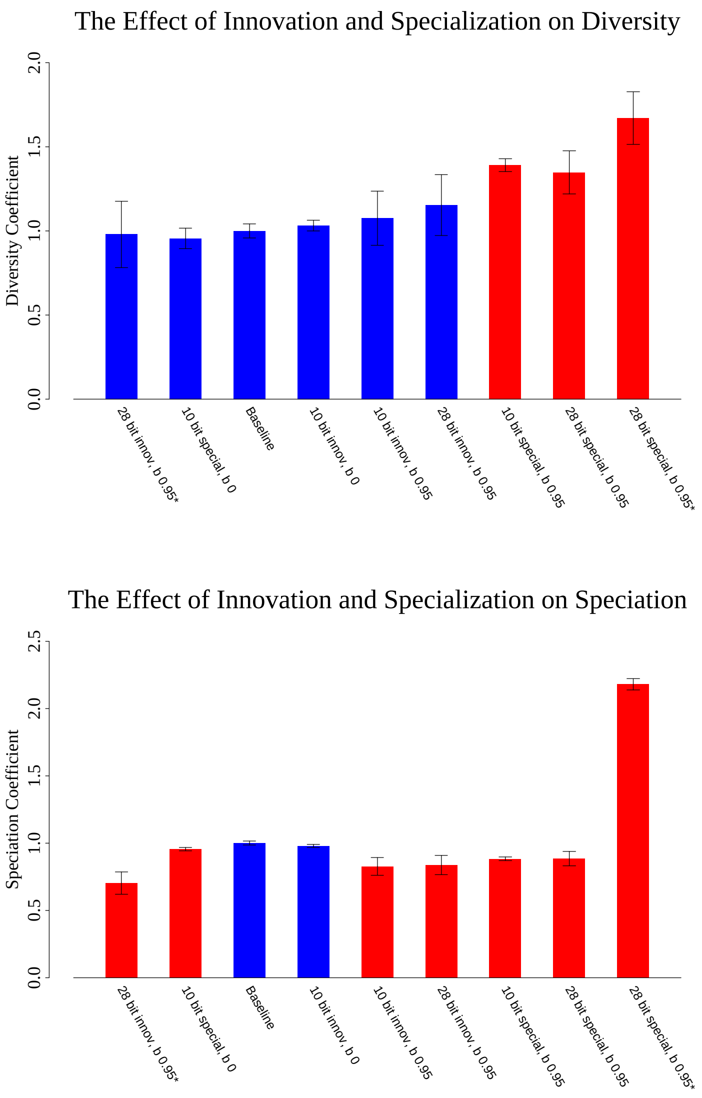


Figure 6-1: A summary of the effects of innovation and specialization on species diversity (top) and speciation rates (bottom). Each effect has been normalized by its relevant baseline. The bars indicate the magnitude of an effect relative to the baseline. The blue bars indicate that the manipulation had no statistically significant effect. Red bars indicate a significant effect, with $p < 0.05$. The conditions have been labeled by the number of bits that were fixed or clamped, out of a total of 32. Abbreviations include "b" for migratory barrier, "innov" for innovation and "special" for specialism. The * marks the runs in which there was a trade-off between generalism and prey capture.

plore new niche space often significantly reduced speciation rates. In contrast, even mild restrictions on specialism had a significant effect on diversification. Specialism seems to facilitate the maintenance of diversity. McPeek (1996) argues that a trade-off between generalism and specialism is necessary to force species to segregate in niche space and thus diversify. The dynamics of the model make this position untenable. Increased capacity for specialism generally lead to an increase in biodiversity. However, the results suggest a restriction of the assertion to speciation rates, not diversity. The increased capacity for specialization generally lowered speciation rates while maintaining greater diversity. The only observed example of an increase in speciation rates was in the runs with a trade-off between generalism and specialism.

The fact that innovations in resource utilization and predation interactions do not seem to increase diversification may be due in part to the simplicity of the model. Heard and Hauser (1995) hypothesize three basic mechanisms by which an innovation might increase speciation. (1) It may allow a lineage to escape competition. (2) It may increase individual fitness. This in turn should increase the population sizes in species with that innovation and so reduce the probability of extinction. One might also add that increased population sizes may provide more opportunities for speciation through mutation and emigration. (3) An innovation may allow more specialization. While innovations in the model can easily boost fitness or increase opportunities for specialization, escaping competition is more difficult. Though it is technically true that a mutation in a prey template chromosome or a generalism chromosome could open up resources that were previously unused, the reality of the matter is that such a mutation would probably just give a lineage an advantage over its competitors. In almost all cases, such a mutation would merely help the lineage exploit a resource that was already under moderate use. Even a slight match of a plant to its habitat gives it a chance of survival. This means that there is probably never a patch completely devoid of plants. So the only reasonable case of an unexploited resource is a plant species that has no predators. In this case, that plant species would likely drive its competitors, who suffer predation, to extinction, and thereby drive the predators extinct as well. An innovation that truly provides new resources in the model is probably rare.

The space limitations in a patch, along with the *energy-conversion* parameter restrict the number of viable trophic levels in the model. This may prevent the kinds of shifts in ecosystem organization that Bambach (1985) identifies as important correlates of increasing diversity in marine life. The model may not allow for that kind of complexity.

Both specialism and innovation have been investigated in the context of predation and resource consumption. While this is fairly reasonable for specialism, evolutionary innovation certainly extends far beyond such a restrictive domain. If it is true that innovations in predator-prey relationships have little impact on diversification, innovations may have dramatic impacts on diversification by way of other aspects of an organism's life history. In fact, the model clearly predicts that an "innovation" that reduces migration between populations of a species should significantly contribute to the future diversification of that species. Cracraft (1990) argues against the idea that an ecological innovation that opens up new adaptive space or facilitates specialism has been a causal factor in the diversification of life. He suggests that for an innovation to stimulate speciation, it must directly impact the processes of speciation through an increase in population isolation, mutation rates, or the fixation of mutations in populations. Most proposed "key innovations" fail to satisfy these conditions (Cracraft 1990). Hunter (1998) adds that there is little evidence for selection directly on characters responsible for speciation. The results of the model seem to support Cracraft's position on innovations, but contradict his dismissal of specialism as an important factor in diversification.

In order to reject Cracraft's position, future work might focus on models in which an evolutionary innovation might allow a lineage to escape competition. Alternatively, the evolution of characters that affect speciation and extinction rates might be examined to test their relative importance against the more typical key innovations of ecological characters.

Another important avenue of research would be to study the evolution of characters that allow reproductive specialization (Hunter 1998). Reproductive specialization might be implemented through characters for mate compatibility or mate preference. The next chapter will examine the effects of enforcing various mate preferences on the populations of organisms.

Chapter 7

Assortative Mating

7.1 What has Sex got to do with it?

Quite a lot, actually. Evolution is based on the differential survival and reproduction of organisms. Yet, the focus on natural selection and the drama of life and death has tended to eclipse the fact that reproduction is an equally crucial, and potentially shaping, dynamic in evolution (Cronin 1991). After all, survival with no reproduction has the same effect on the genetic composition of the next generation as dying before having the opportunity to reproduce.

Perhaps, given that sexual reproduction defines the boundaries of gene pools, it is not surprising that the dynamics of sex can have an important impact on diversity. Bernstein et al. (1985) point to an important implication of sexual reproduction in addition to the sharing of genes in a gene pool. In order to reproduce sexually, one must find a mate. This may be no problem if a species is quite common and so potential mates abound. However, if the species is rare in an area, it may be quite difficult to locate a potential mate. In other words, sexual reproduction includes a cost for low population density. A superior competitor that arrives in a region either through mutation or migration may still die out if only because the probability of finding a mate is low (Bernstein et al. 1985). All other things being equal, the penalty for low density may tend to kill off the outliers in character or geographical space and so cluster the remaining organisms into discrete bunches. This alone may explain the perception of species as coherent units. The supposition that outliers will be tend to die out rests on the assumption that these outliers tend to be reproductively differentiated from the more common organisms and that this differentiation makes it harder for them to successfully mate.

Andersson (1994, pp. 205–226) reviews the importance of sexual selection in speciation. He finds evidence that "mechanisms of mate recognition, which often seem to be based on sexually selected traits, may be crucial in speciation" (Andersson 1994, p. 223). Specifically, evidence for the importance of sexual selection in diversification comes from the mating behaviors of passerine birds, anurans (frogs), Hawaiian fruit flies and crickets, African cichlid fish, insect pollinated angiosperms (flowering plants), and others. Many of these cases may simply be cases where the organisms are using perceptual cues to selectively mate with similar organisms. That is, cues that facilitate positive assortative mating.

The very existence of sexual reproduction, which was invented about 1 billion[1] years ago (Cowen 1990, p.67), introduces a new dimension to the process of evolution. No longer is it adequate just to survive and accumulate resources. With sexual reproduction, an organism must find and be accepted by a mate. These pressures are every bit as important in the evolution of a lineage as the pressures to avoid predation and find nourishment. Recent models have shown that the evolution of mate preferences can stimulate rapid speciation without any geographic barriers to divide the population (Todd & Miller 1997; 1991). This chapter will examine the extent to which the pressures of mate selection lead to the division of gene pools and the formation of new species.

7.2 Sexual Selection

There are many forms of sexual selection. In general it encompasses all factors that influence the quantity and quality of mating opportunities of organisms. One archetypal arena in which these factors play out is in the choices, or selection, of mates by organisms. However, it is only considered sexual "selection" if the choice of the mates comprises a selective force on the organisms. This means that mate choices lead to differential reproduction due to the expression of genetic differences in the organisms.

Sexual selection has been implicated in the evolution of some of the most glorious exemplars of nature, including most of the spectacular plumage on male birds, most famously the peacock. The power of sexual selection lies in a positive feedback loop. Let us consider two genes. One that encodes a female preference for bright plumage in their mates. The

[1] Give or take half a billion.

other encodes the development of that plumage in the males. Then a male with bright plumage will enjoy an average number of mating opportunities with the females who do not care about plumage, but an additional bonus number of mating opportunities with the choosy females. This explains how the genes for colorful males spread, but not how the females genes for that preference spread. That this was a serious hole in Darwin's theory of sexual selection went largely unnoticed for half a century (Cronin 1991). And so the significance of R. A. Fisher's solution in 1915 was not fully appreciated until quite recently. Cronin quotes a characteristically modest reflection by, perhaps the most brilliant living theoretical biologist, John Maynard Smith:

> "In the extensive publications marking the centenary of the *Origin of Species*, the only explicit treatment of sexual selection was Maynard Smith (1958a); although I did describe a possible mechanism of female choice in *Drosophila subobscura*, it is clear that I had not read or understood Fisher." (Cronin 1991, p.244)

Fisher's solution was to show how selection could indirectly impact preference genes. Given some proportion of choosy females in the population, we know that colorful males have an advantage over their drab brethren. This means that any female that mates with a colorful male will tend to have colorful sons. These sons will pass on to their many offspring the choosy mother's genes as well as their father's colorful genes. Thus a genetic correlation arises, which biologists call linkage disequilibrium, between genes for colorful males and genes for choosy females. The spread of the female preference genes creates an even stronger selective effect benefiting the colorful males and the choosy females. So the spread of both coloration and female preference for coloration leads to a positive feedback loop or "runaway process." The mathematical description of this theory was worked out by Kirkpatrick (Collins & Jefferson 1992, summarize and extend this work). Computational models have been able to elaborate the mathematical version by showing that the predictions are robust to relaxing the simplifications that made the mathematics tractable (Collins & Jefferson 1992). Specifically, Collins and Jefferson showed that the addition of local mate choice, mutations, finite populations, geographic isolation of subpopulations, and diploidy do not qualitatively change the results. Unfortunately, the literature has been more supportive of explanations for the evolution of female mating preferences based on the direct impact of the preference genes upon the female's fitness, and not her offspring's fitness (Kirkpatrick & Ryan 1991).

If this canonical form of sexual selection is the ornate arch of a Gothic cathedral, then simple assortative mating is an arch from Stonehenge. It is a general, and often non-directional, form of selection. Assortative mating arises from organisms pairing up in non-random ways. These non-random pairings may be the result of some structuring of the population (e.g., sedentary organisms will tend to mate with their neighbors) or from some form of mate choice. The famous Hardy-Weinberg equilibrium[2] for the proportion of homozygous and heterozygous individuals in a population is based on the assumption of random mating. This assumption is patently false in our species for many of the characters that humans most readily perceive (Hartl & Clark 1997). That is, mating is only random with respect to certain characteristics or loci, not others. In fact, biologists routinely check for non-random mating with respect to a locus by looking for deviations from the Hardy-Weinberg equilibrium.

Non-random mating with respect to a locus is also called assortative mating. It comes in two flavors, positive and negative[3]. Positive assortative mating is the tendency to mate with individuals with similar characteristics or alleles. This tends to lead to an excess of homozygotes in comparison to Hardy-Weinberg equilibrium. Humans exhibit positive assortative mating with respect to many characters, including height and hair color (Hartl & Clark 1997). In contrast, negative assortative mating is the tendency to mate with dissimilar individuals and so it tends to lead to an excess of heterozygotes. An extreme form of this is found in many plants which prevent fertilization by pollen that carry the same alleles as the plant at "self-incompatibility" loci. They select for pollen from plants that are different from them at these loci.

Todd and Miller (1997) introduced mating preference genes to an evolutionary model and

[2]In an almost apologetic letter to *Science* in 1908, the mathematician Godfrey Hardy pointed out the following phenomenon (developed by Weinberg independently in the same year). If the proportion of one allele (A) in a population is p, and the proportion of the other allele at that locus (a) is $q = 1 - p$, then you can expect to get p^2 of the AA homozygotes in the next generation, along with $2pq$ of the Aa heterozygotes, and q^2 of the aa homozygotes. This is easiest to see by considering the gametes, the sperm and eggs, that will combine to form the next generation. We know that on average p of the sperm will carry the A allele while p of the eggs will also carry the A allele. So the chance of an A-sperm coming together with an A-egg is just $p \times p$. The same reasoning can deduce the proportions of Aa heterozygotes and aa homozygotes, with a note that some of the pq heterozygotes are formed by the union of an A-sperm and an a-egg, while other pq heterozygotes are formed by the union of an a-sperm and an A-egg. This results in $2pq$ heterozygotes.

[3]Positive and negative assortative mating are also sometimes called assortative and disassortative mating.

found that populations tended to speciate sympatrically. More specifically, a subpopulation would converge on a particular phenotype as well as a preference for that phenotype in a mate. Different subpopulations converged on different preferences. So the subpopulations naturally formed reproductively isolated gene pools. However, because phenotypes were only specified by two loci, the preferred phenotypes in the subpopulations often drifted towards each other and the gene pools of those subpopulations merged. Would the same dynamic of spontaneous sympatric speciation play out in an ecological model with many loci?

Assortative mating was implemented in the model by choosing a random location in the array of organisms and inspecting the first eight potential mates after that point. A potential mate is a live organism in the same patch whose reproductive chromosome is at most one bit different from the organism initiating the mating. The best of the eight potential mates was then selected for mating, where "best" was interpreted differently for negative and positive assortative mating. For positive assortative mating, "best" meant the closest match of the chromosomes of the two mates. For negative assortative mating, "best" meant the greatest mismatch. In the case of a tie, the first of the top ranked mates was chosen. The particular chromosomes that were used to determine similarity differed across the experiments. This is a simplified version of Todd and Miller's (1997) sexual selection in that the phenotype and the preference are encoded in the same genes.

Does this form of assortative mating constitute a kind of sexual selection? An organism will differentially gain mating opportunities based on its genotype (which is identical to its phenotype). Given a population's genotypes and the form of assortative mating, we can predict which genotypes will leave more copies of themselves in the next generation. So it is selection in that sense. However, over long time scales, there are no particular characters that are specifically being selected. We would not expect to see a peacock's tail evolve under assortative mating.

The only form of positive feedback possible under assortative mating, is the convergence of subpopulations into homogeneous mating pools with positive assortative mating. If we consider genotype space for a moment, any clusters of similar organisms will tend to all mate with each other, while isolated outliers will be forced to mate with relatively distinct organisms. So the clusters of similarity will grow, fed by the assimilation of the outliers. This is not the kind of runaway process we expect from positive feedback. There is a clear limit to the process once the populations become completely homogeneous. Whether or not this limit is reached will depend on factors like the mutation rate and whether or not the homogeneous subpopulations can grow large enough that the individuals can easily find each other when it comes time to mate.

7.3 Assortative Mating and Speciation

How might assortative mating affect speciation? It has long been recognized that positive assortative mating facilitates a sort of disruptive selection (Andersson 1994). Under positive assortative mating, a population will soon sort itself into a set of relatively homogeneous subpopulations in which individuals prefer to mate with other individuals of their subpopulation to the exclusion of mating with outsiders. Note that positive assortative mating in the model, and often in nature, does not create isolated gene pools in and of itself. If an organism in the model cannot find a similar mate, it will mate with a dissimilar organism. Given such, slightly porous, behavioral reproductive isolation we might expect the reproductive chromosomes of the subpopulations to drift apart. Analogous to geographical isolation, positive assortative mating should separate the subpopulations, allowing them to diverge[4] and eventually become species in their own right. So we should expect positive assortative mating to boost speciation rates.

Conversely, negative assortative mating should tend to bind populations together, and prevent them from forming isolated subpopulations. Any new mutant will instantly become "popular" and the mutation will tend to be mixed with the other genes, rather than forming a distinct subpopulation. This should suppress speciation.

In all the following cases the baseline for comparison is the original random mating data presented in Chapter 5 with the matching level of migratory barriers and habitat heterogeneity. With no migratory barriers and a homogeneous environment there was a baseline average of 2.5 species living at the end of 5000 time steps and a speciation rate of 0.0425 speciation events per time step. However, random mating in a heterogeneous environment with 0.99 barriers resulted in an average of 11.83 species and 0.1195 speciation events per time step. As before, a heterogeneous environment means an average of 10.3 bits

[4]Divergence will occur mainly in the characters that are not being selected through assortative mating, given appropriate conditions for divergence. These conditions include either divergence selection or small enough populations that genetic drift can play an important role in the population's evolution.

different between the habitat bit patterns of any two patches.

7.3.1 Selection on Predation Resistance Characters

Two forms of assortative mating were examined: selection on a mate's predation resistance chromosome and selection on a mate's reproductive chromosome. The degree of similarity between mates was based on a count of the number of bits that differed between the relevant chromosomes of the two mates.

Negative Assortative Mating

Under negative assortative mating based on predation resistance characters, an organism would choose, from up to 8 potential mates, the organism that had the least similar predation resistance chromosome. Offspring were then generated in the normal way, their chromosomes being determined by two point crossover between the parental chromosomes, along with possible point mutations. Under a homogeneous environment with no migratory barriers, negative assortative mating based on predation resistance characters resulted in only 1.5 species and 0.0160 speciation events per time step. Both of these values are significantly below the equivalent measurements of the random mating condition ($p < 0.001$ in a two-sided T-test). Similarly, under 0.99 migratory barriers and a heterogeneous environment, this form of negative assortment resulted in 5.18 species and 0.0615 speciation events per time step. Again, these are significantly below the results for random mating ($p < 0.001$). A heterogeneous environment with 0.95 migratory barriers produced an average of 3.98 species and a speciation rate of 0.04988. Both of these are below the values observed under random mating in a heterogeneous environment with 0.95 barriers ($p < 0.001$).

Positive Assortative Mating

Positive assortative mating on predation resistance chromosomes is analogous to negative assortment, except that mates are chosen out of the 8 random suitors so as to maximize the number of bits that match between the organism's predation resistance chromosome and that of its suitor. Under a homogeneous environment and no migratory barriers, this led to the survival of 3.09 species and a speciation rate of 0.0608. With a heterogeneous environment and 0.99 barriers, these statistics rose to a whopping 44.27 species and 0.7127 speciation events per time step. All of these are significantly higher than the corresponding values under random mating ($p < 0.01$ in the first case, and $p < 0.001$ in the other three). A heterogeneous environment with 0.95 migratory barriers similarly led to an average of 16.76 species and a 0.3874 speciation rate ($p < 0.001$).

7.3.2 Selection on Reproductive Characters

Sexual selection on the predation resistance chromosome is tantamount to choosing a mate based on its predation characteristics. It is also reasonable to consider the selection of a mate based on its reproductive characters like a mating dance or secondary sexual displays. Because reproductive isolation is based on these characteristics, assortative mating in reference to them should have an impact on speciation dynamics. An initial hypothesis might suggest that, like assortative mating on predation resistance, negative assortative mating should tend to glue species together. In contrast, we might expect positive assortative mating to fragment populations and thus boost speciation rates.

Negative Assortative Mating

Negative assortative mating on reproductive characters was implemented in a similar fashion to assortative mating on predation resistance characters. The only difference is that a potential mate can have at most a one bit difference in its reproductive chromosome compared to the chromosome of the initiating organism. So negative assortative mating means looking at eight potential mates and taking the first one that has a different reproductive chromosome. If none of the eight had a different reproductive gene, the first potential mate was chosen. The reproductive chromosome of the offspring was identical to the reproductive chromosome of one of the parents, assuming no mutation. The offspring had a 50/50 chance of matching a particular parent's reproductive chromosome.

In a homogeneous environment with no migratory barriers, negative assortment on reproductive characters led to an average of 2.24 species. This is significantly below the 2.5 species generated by the random mating condition ($p < 0.05$). Similarly, negative assortment led to a speciation rate of 0.0148, significantly below 0.0425 for random mating ($p < 0.001$). However, the situation reversed when the model was run with 0.99 barriers to migration and environmental heterogeneity. Under these conditions, negative assortative mating on reproductive genotypes leads to an astonishing average of 29.42 species and a speciation rate of 0.2732. These are both significantly *greater* ($p < 0.001$) than the corresponding values, 11.83 and 0.1195, for random mating. Similarly, in a heterogeneous environment with 0.95 migratory barriers, negative assortment on reproductive characters produces a significant ($p < 0.001$) increase in the number of species (17.22) and the speciation rate (0.2260).

Positive Assortative Mating

As before, positive assortative mating was the opposite of negative assortative mating on reproductive characters. The initiating organism looked for an exact match with its reproductive chromosome amongst the eight potential mates. If none was to be found, it accepted the first potential mate with a one bit difference.

In a homogeneous environment with no reproductive barriers, positive assortative mating based on reproductive characters resulted in an average of 2.2 species and a speciation rate of 0.0148 species per time step. These are both significantly lower than the random mating condition ($p < 0.05$ and $p < 0.001$ respectively). Yet, when the organisms are evolving in a heterogeneous environment with 0.99 migratory barriers between patches, the biodiversity rises to 30.00 species and the speciation rate rockets to 0.2765 species per time step. A heterogeneous environment with 0.95 migratory barriers results in a biodiversity of 15.26 species and a speciation rate of 0.2338. All of these are statistically significant ($p < 0.001$).

7.3.3 Summary of the Results

Figures 7-1 and 7-2 summarize the results. The influence of the different kinds of assortative mating on species diversity (Figure 7-1) and speciation rates (Figure 7-2) have been normalized by the results for random mating. In all cases, the influence of assortative mating is statistically significant at the $p < 0.05$ level, as indicated by the red bars in the figures.

As one might guess from Figure 7-1, biodiversities under positive and negative assortative mating on reproductive characters are statistically indistinguishable. The same is true for speciation rates in Figure 7-2. When there are no migratory barriers and a homogeneous environment, both conditions lead to significantly less speciation than the random mating case and lower biodiversity. However, with 0.95 barriers to migration and habitat heterogeneity across patches, both conditions lead to higher speciation rates and biodiversity relative to random mating. Yet, both conditions remained statistically indistinguishable.

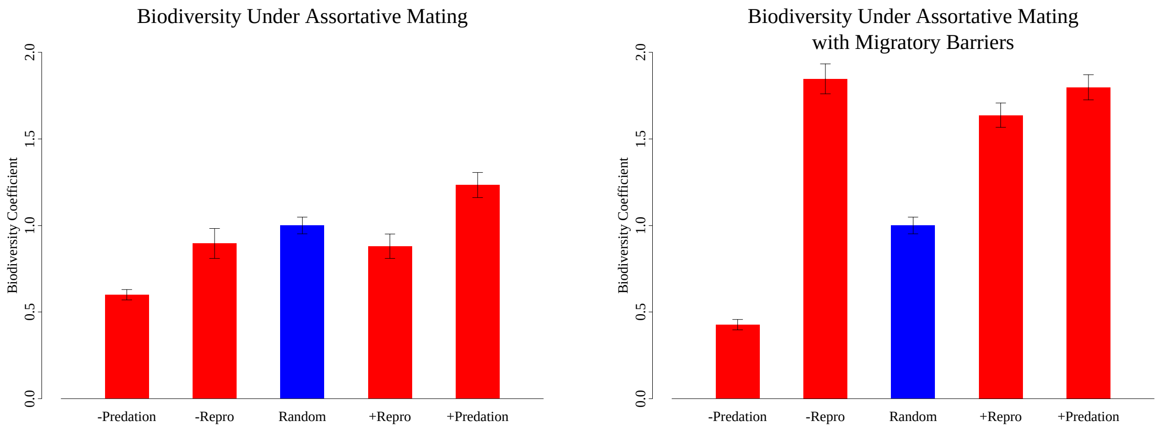


Figure 7-1: The effects of assortative mating on species diversity in a homogeneous environment with no migratory barriers (left) and in a heterogeneous environment with 0.95 migratory barriers (right). "Repro" stands for either positive (+Repro) or negative (−Repro) assortative mating on reproductive characters. "Predation" indicates either positive (+Predation) or negative (−Predation) assortative mating on predation resistance characters. Red bars indicate statistically significant effects that distinguish assortative mating from the baseline of random mating. The effects of each condition have been normalized against the baseline value. For example, under a heterogeneous environment with 0.95 migratory barriers (right hand bar chart) positive assortment on predation resistance characters increases biodiversity by a factor of 1.8.

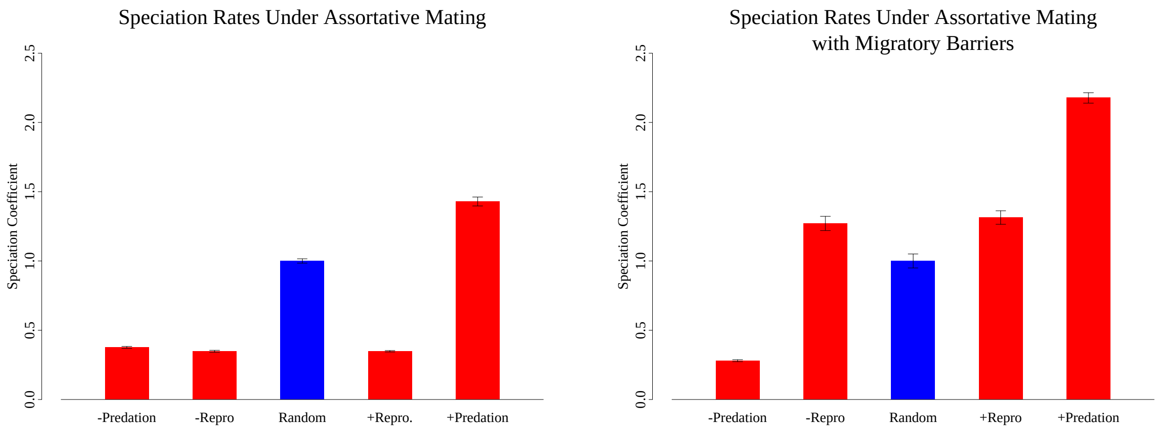


Figure 7-2: The effects of assortative mating on speciation rates in a homogeneous environment with no migratory barriers (left) and in a heterogeneous environment with 0.95 migratory barriers (right). Again, red bars indicate statistically significant effects that distinguish assortative mating from the baseline of random mating. The effects of each condition have been normalized against the baseline value.

The same is true for 0.99 migratory barriers in a heterogeneous environment.

Assortative mating based on predation resistance characters has the predicted effects. Negative assortative mating depresses speciation rates and the resulting species diversity. Positive assortative mating stimulates speciation and the maintenance of diversity. The presence of more species in the positive assortative mating condition does not necessarily indicate the colonization of more niches. It may result from the same number of "successful" species surrounded by a plethora of young species that are quickly dying out. If this were the case, a histogram of the population sizes of the species would show many small, a few large, and a very few medium sized species populations. However the histograms in Figure 7-3 show that the high diversity of the positive assortative mating spans all possible population sizes.

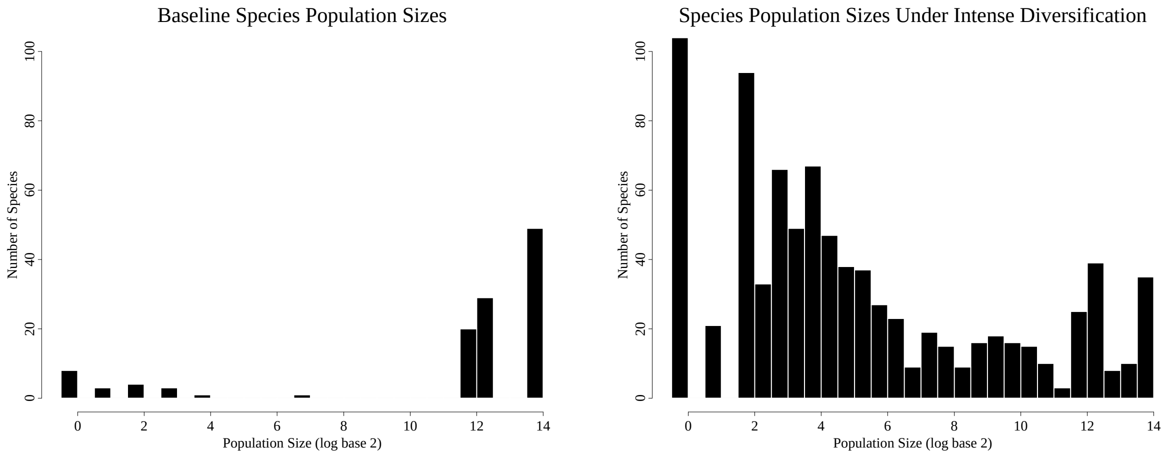


Figure 7-3: The distribution of species population sizes for both the baseline "vanilla" conditions (left) and strongly diversifying conditions (right) summed over 49 trials. Population sizes have been scaled logarithmically. The right histogram is based on running the model with positive assortative mating based on predation resistance characters in a heterogeneous habitat with 0.95 migratory barriers. The baseline condition, with random mating, no migratory barriers and a homogeneous environment produces a bimodal distribution. Typically an extremely large species is accompanied by a few very rare species. In contrast, the intense diversification condition leads to a distribution that spans all the possible population sizes.

7.4 Discussion

Why should positive and negative assortative mating based on reproductive characters behave similarly? And why should they interact so dramatically with habitat heterogeneity and migratory barriers?

To understand the dynamics of speciation we must examine the graph components that represent species. The growth and fragmentation dynamics of these graphs are complicated due to the underlying population interactions that generate the graphs (see Appendix A for a discussion of some related open problems). However, a few things can be deduced about the effects of assortative mating when selection is based on the reproductive genotypes.

7.4.1 Negative Assortative Mating on the Hypercube

First, consider the simple case of a "populous" node, where many organisms share the same reproductive genotype. If a mutation in the offspring of one of these organisms creates a new node, what are the expected dynamics of the population on this two node graph? It is perhaps easiest to see the effects of negative assortative mating. If all the organisms are trying to mate with different reproductive genotypes, then many of the organisms in the populous node will mate with the mutant with the new genotype. Half of these matings will produce organisms with the rare genotype, and half with the populous genotype. Similarly, the new genotype organism will mate with an organism that has the common genotype. The end result is a net flow from the populous to the rare genotype. It is clear that this flow will continue until the two genotypes reach an equal popularity. This is a stable equilibrium because any perturbation will stimulate a new flow from the more populous genotype to the less populous genotype. However, the story does not end there.

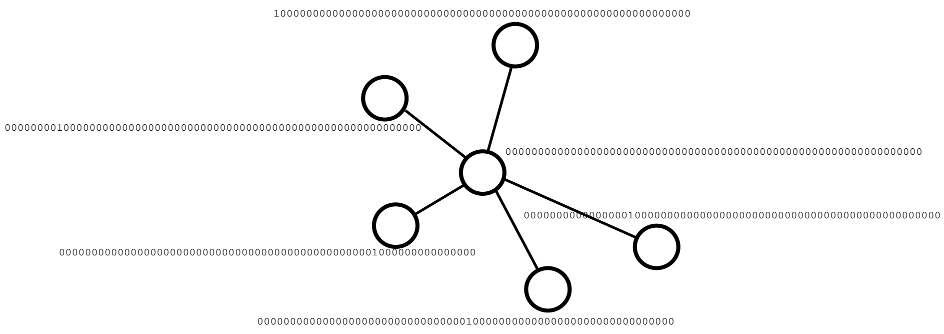


Figure 7-4: A star graph of reproductive genotypes. Each node has been labeled with a representative 64-bit genotype. The presence of a node indicates that an organism exists with that genotype. So the appearance and disappearance of nodes is determined by the underlying population dynamics of the organisms that have those reproductive chromosomes.

Consider the dynamics of a star graph, a graph with one central node and many neighbors, but with no connections between the neighbors, as shown in Figure 7-4. Recall from Section 4.3 that in a random graph in a subcritical state, the largest component will tend not to have cycles (Gavrilets & Gravner 1997). A star graph is thus a reasonable archetype to analyze[5]. Each of the satellite, or leaf, nodes has only one neighbor, and so all of the organisms in the leaf nodes will choose to mate with the organisms of the central node. In contrast, organisms in the central node will tend to mate with organisms from all the neighboring nodes. If there are a total of m organisms in the leaf nodes, then approximately $m/2$ organisms will flow into the central node every generation. In contrast, if there are n organisms in the central node, it will tend to distribute $n/2$ organisms evenly across its neighboring nodes. It is clear that this balances out when $n = m$, or when the number of organisms in the central node equals the sum of all the organisms in its neighboring nodes. So with negative assortative mating, the equilibrium distributes the population across the nodes of a component in proportion to the number of neighbors a node has. The result is that the leaf nodes have the fewest organisms and so are the most susceptible to dying out. However, a species only splits if a non-leaf node dies. Furthermore, the flow of organisms into a node when its population is pushed below equilibrium, acts as a buffer against the disconnection of a component. If a non-leaf node starts to become rare, all of its neighbors will begin contributing individuals to that node, boosting it back up again. In particular, the neighbors with genes that are ecologically successful are the most likely to contribute to the endangered node. So that node will be infused with relatively successful ecological chromosomes. In sum, due to the low populations in the leaves and the buffering effects in the non-leaves, we should expect speciation events to be suppressed by negative assortative mating. We will leave the discussion of the interaction between migratory barriers, environmental heterogeneity, and negative assortative mating until later.

7.4.2 Positive Assortative Mating on the Hypercube

Now consider the case of positive assortative mating. As before we will start with a single populous node and a new mutant neighbor node with only one organism. The organisms in the populous node will only mate with themselves, while the new mutant will not be able to find a similar mate and so will also mate with an organism in the populous node. The result

[5] Although a star graph is probably not a realistic representation of the mating relationships in a real species. See Section 4.5.

of this mating is that there is only a 50% chance that a new organism will enter the mutant node. Even if the mutant node's population grows a little, since the organisms of this node will be relatively rare, there is still a good chance that they will not find each other when they scan over eight potential mates in their patch. So positive assortative mating biases against the growth of new mutant nodes, and many will tend to get reabsorbed by the populous nodes. Liou and Price (1993) found the same dynamic in their model of positive assortative mating. The rarer population tended to go extinct.

This dynamic remains the same in a star graph. If a mutant node ever does grow to a point where the constituent organisms can find each other as mates, the dynamics of the populous node and the new node will be relatively independent. Each will rise and fall on the merits of the ecological chromosomes in their constituent organisms as well as the chance events of death through competition for space. Thus positive assortative mating is also biased against the survival of leaf nodes, and so depresses speciation, but for completely different reasons as compared to the bias of negative assortative mating.

7.4.3 Random Mating on the Hypercube

Finally, consider the dynamics of random mating on a two node graph. If there are n organisms in the first node and m in the second, we can derive the population sizes of the next generation as:

$$\begin{aligned} n' &= n\frac{n}{n+m} + \frac{n}{2}\left(\frac{m}{n+m}\right) + \frac{m}{2}\left(\frac{n}{n+m}\right) && (7.1) \\ &= n && (7.2) \\ m' &= m\frac{m}{n+m} + \frac{m}{2}\left(\frac{n}{n+m}\right) + \frac{n}{2}\left(\frac{m}{n+m}\right) && (7.3) \\ &= m && (7.4) \end{aligned}$$

The first term in Equation 7.1 corresponds to the number of matings between organisms in the populous node. The second term, $\frac{n}{2}\left(\frac{m}{n+m}\right)$ is the number of organisms contributed to the populous node from matings between nodes, initiated by organisms with the common genotype. Finally, the last term is the number of organisms contributed by matings between nodes that were initiated by organisms with the mutant genotype. The result $n' = n$ means that, despite the interbreeding, the population sizes of the two nodes should remain the same. This does not hold for the star graph case. If we call the population size of the central node n and the populations of the leaf nodes m_i, then

$$n' = n\frac{n}{n + \sum_i m_i} + \frac{n}{2}\left(\frac{\sum_i m_i}{n + \sum_i m_i}\right) + \sum_i \frac{m_i}{2}\left(\frac{n}{n + m_i}\right) \tag{7.5}$$

and since

$$\sum_i \frac{m_i}{2}\left(\frac{n}{n + m_i}\right) > \frac{n}{2}\left(\frac{\sum_i m_i}{n + \sum_i m_i}\right) \tag{7.6}$$

we get

$$n' > n \tag{7.7}$$

So there is a tendency, though a weak one, for the less populous nodes to get absorbed by their neighbors. This tendency is not nearly as strong as it is in positive assortative mating. Nor does random mating produce the kind of buffering effects against node death that negative assortative mating exhibits. Thus, random mating, stuck half way between either extreme, ends up exhibiting higher speciation than either positive or negative assortative mating.

7.4.4 Migratory Barriers and Habitat Heterogeneity

Why, then, does the relationship between random and assortative mating reverse under high migratory barriers and habitat heterogeneity? The complexities of the underlying dynamics mitigate against a full analysis. However, a few hypotheses may be laid on the table. For a start, negative assortative mating is generally thought to boost variation within a population (Cronin 1991, p.237). Since variation is the meat upon which selection feeds, this ought to interact well with disruptive selection pressures, speeding a subpopulation along to a new adaptive "peak." In contrast, positive assortative mating ought to increase the variation between populations. This may interact well with selection pressures that differ between populations.

The effect of high migratory barriers is the divide the total population that we have been describing by the reproductive graphs, into 16 subpopulations which are nearly independent. A speciation event can be caused by any one of these subpopulations evolving to a state in which its reproductive graph is disconnected from the reproductive graph represented by the agglomeration of the other 15 subpopulations. Furthermore, since each subpopulation experiences a significantly different habitat as compared to the others, selection becomes a disruptive force. This is clearly illustrated in the random mating case when comparing the speciation rate under a homogeneous environment with no barriers (0.0425) to a heterogeneous environment with 0.99 barriers (0.1195). Is there any reason to suspect that both positive and negative assortative mating should interact with this sort of disruptive selection?

In both positive and negative assortative mating, the interaction with habitat heterogeneity and migratory barriers probably depends on the fate of "rare" mutant nodes in the total population's reproductive graph. A node that represents few organisms in the total population, may have a local majority within its patch. Purely by the variance of the sampling that divides the total population into 16 subpopulations, it may occasionally be the case that within a patch a "mutant" node's population will out number its neighbor node's population. This is made all the more likely because of the different pressures in the patches. Ecological chromosomes that help their organisms in one patch will often hinder their organisms in another patch. So a "mutant" node with a small population that has relatively good ecological chromosomes for its patch may well outgrow the population in a neighboring node that has more constituent individuals in the total population. Once a "rare" node outnumbers its neighbors, the dynamics of both negative or positive assortative mating will work in its favor. It will tend to spawn more satellite, leaf nodes and so become the center of its own star graph. Its satellite nodes will tend to contribute more organisms to it than it gives up to them. In short, it becomes the center of an isolated reproductive graph in its patch. Thus, the patch's reproductive graph dynamics are effectively separated from the total population and given the pressures of selection, it is likely to drift apart from the reproductive graphs in the other patches.

7.5 Conclusions and Implications

There are three important conclusions from these experiments. First, assortative mating has dramatic effects on speciation rates and diversity levels. Second, the nature of those effects depends on what characters organisms are using to select their mates. If the selected characters do not interact with reproductive barriers, except through the assortative mating, then positive assortment dramatically boosts speciation while negative assortment suppresses it. The importance of positive assortative mating for diversification in the model matches evidence that mate choice based on ecological characters has promoted the rapid diversification of three-spined stickleback fish (Schluter 1996). If the selected characters influence reproductive barriers through channels independent of mating preferences, then we are faced with the third and by far the strangest conclusion. The effects of assortative mating on reproductive characters changes depending on the relative isolation and degree of disruptive selection on subpopulations. If the organisms are evolving under homogeneous conditions as one large population, then assortment by reproductive characters depresses speciation regardless of whether the assortment is positive or negative. However, if subpopulations are isolated and experience different habitats, then both forms of assortment dramatically boost speciation relative to random mating. Notice that these conclusions make no reference to the cost of wasting a mating attempt on an incompatible mate. There is no obvious force of natural selection in the traditional sense of the struggle for survival. It is rather a dynamic of the sexual selection in the model and its interaction with the genes that control characters for reproductive isolation.

The results carry implications for systematists and paleobiologists looking for explanations for the explosive radiation of a group. When searching for a "key innovation" that allowed the group to diversify, one should look for characters that allowed or affected mate choice, as well as characters that stimulated geographical isolation of subpopulations. Angiosperm (flowering plant) evolution may provide a particularly interesting test case. The evolution of pollination by animals (usually insects) acts as a form of positive assortative mating. Those plants that share similar attractive features for a pollinator will tend to pollinate each other. There is some evidence that animal pollination indeed boosted diversification in angiosperms (Eriksson & Bremer 1992). Whether or not the effects on diversity of animal pollination is analogous to positive assortative mating is debatable.

There is some reason to think that the degree of positive assortative mating should reduce or at least stabilize over time, due to coevolution with parasites. The more homogenous the hosts, the easier it is for parasites to evolve strategies for exploiting the host population. A mutation in a parasite that gives it an advantage in its utilization of its current host will tend to help it and its offspring exploit the other hosts in a homogeneous population. In fact, one of the theories for the origin and maintenance of sexual reproduction is based on the idea that sexual reproduction helps to maintain genetic variation in a population (Hamilton et al. 1990). These same pressures should tend to reduce the prevalence of extreme forms of positive assortative mating. Particularly positive assortative mating based on characters that relate to an immune system.

The vulnerability of homogenous populations to parasitism is part of the reasoning that Brown (1997) uses to suggest that there is selection for maximizing heterozygosity in offspring. Brown also cites developmental stability, and attractiveness (based on an organism's symmetry) as further benefits of heterozygosity. He predicts that various forms of negative assortative mating should be prevalent across all sexual species (Brown 1997). If it is true that negative assortative mating is widespread, the model would suggest that in most cases this has depressed speciation. The only exception being in cases where a species is fragmented into relatively isolated populations experiencing a variety of selective pressures. In this case, if assortment is based on characters that are independently involved in reproductive isolation, selection for heterozygosity may actually stimulate speciation.

The results also have implications for computer science, in the design of evolutionary algorithms. The field of evolutionary algorithms is characterized by the attempt to evolve solutions to a given problem through successive applications of mutation, crossover (sexual reproduction), and selection of the best individuals in the current population of solutions (Goldberg 1989; Mitchell 1996; Banzhaf et al. 1998). One common problem in our attempts to evolve solutions to complex problems, is premature convergence. Often, the entire population of solutions becomes homogenized, with all the members representing small variations on one solution that is only locally, and not globally, optimal. When a population loses variation, selection is useless. Without variation it is impossible to say one solution is better than another. There is no way to shuffle genes between individuals through sexual reproduction to produce new variations. Ideally, we would like the population to maintain genetic diversity such that a broader area of the solution space may be searched and thereby raise the probability of discovering the global optimum. A few studies have shown that the performance of a genetic algorithm improves with the prohibition of mating between similar (Eshelman & Schaffer 1991) or related (Craighurst & Martin 1995) solutions. This fits well with the intuition that negative assortative mating boosts genetic variation within a population.

The opposite effect, positive assortative mating, has also been explored and shown to be effective (Deb & Goldberg 1989; Booker 1985). If similar solutions preferentially mate, then the population should fracture into subpopulations (Todd & Miller 1991). In a genetic algorithm, this can be exploited to maintain subpopulations on multiple fitness peaks. A simple implementation of this positive assortative mating would be to choose a few potential mates for a solution and then select the mate with the fewest bits different from that solution (i.e., the shortest Hamming distance). Deb and Goldberg (1989) used both negative assortative and positive assortative mating to improve the performance of a GA on a fitness landscape with many peaks. Negative assortative mating was a side effect of their "sharing" scheme which reduces a solutions fitness if there are similar solutions in the population. Since mating is a probabilistic function of the fitness values, this encourages solutions to mate with dissimilar solutions. They also introduced explicit positive assortative mating to fracture the population by selecting similar solutions for mating from the population. They refer to these subpopulations as "species."

7.5.1 Towards an Experimental Test of the Predictions

These interactions between the characters for mate selection, population structure and habitats seem hopelessly complicated. With this in mind, I will conclude with the description of an experimental test of the model's predictions. A potential test of the hypotheses rests on the behavior of *Wolbachia* bacteria (Werren 1997). Members of the *Wolbachia* genus infect the reproductive tissues of arthropods. They have the surprising effect that males infected with a particular strain of *Wolbachia* can only successfully fertilize females that are also infected with that same strain. So, hosts that are infected with different strains of *Wolbachia* will appear to be members of different species until they are treated with antibiotics and demonstrate that they can, in fact, successfully reproduce. It seems that the *Wolbachia* in the male host alters the sperm of its host. These deformed sperm are incapable of fertilizing an uninfected female[6]. However, if an egg contains the same strain of *Wolbachia*, the bacteria seems to "rescue" the deformed sperm, allowing it to combine with the egg and form a viable zygote. Given their effects on reproductive isolation, it is perhaps not surprising that *Wolbachia* has been implicated in the rapid speciation of some taxa (Giordano et al. 1997; Rousset & Solignac 1995; Hurst & Schilthuizen 1998) and are found in two of the most

[6]A zygote forms but typically does not survive past the first few cell divisions. The mechanisms for this failure are still being worked out (Werren 1997).

specios groups, insects and mites (Giordano et al. 1997).

The restriction to mating between organisms infected with the same strain of *Wolbachia* may be sufficiently similar to positive assortative mating to allow a test of the model. In this case, the "character" being selected by the hosts is the type of sperm or eggs in their potential mates. This suggests that populations infected with a variety of strains of *Wolbachia* may be evolving under conditions similar to positive assortative mating on reproductive characters. The model predicts that subpopulations should evolve reproductive isolation relative to each other if they are geographically segregated and experience somewhat different selective pressures. Paradoxically, the model also predicts that if the population is kept in one large homogenous environment, it should resist reproductive segregation relative to a randomly mating, uninfected population. This can be tested with population cages of *Drosophila* fruit flies.

The fruit flies carrying the different strains of Wolbachia should also carry genetic markers for easy identification. Typically these include characters like red or white eye color mutations. The experiment requires at least four conditions. In the first control condition, flies carrying the markers but free of *Wolbachia* infection would be kept in one large, homogeneous cage. This corresponds to the "vanilla" condition in the model. The second condition would separate uninfected flies into a system of cages with some form of low level migration between populations. Each cage would have different food sources and environmental conditions so as to provide a heterogeneous environment with migration barriers. Both these conditions, lacking *Wolbachia* infection, should exhibit random mating. The two experimental conditions would be the same as the controls except that all populations would have a mixture of infections of the different *Wolbachia* strains, each associated with a genetic marker. The populations would be allowed to evolve for a sufficient amount of time such that mutations might introduce other forms of reproductive isolation, besides the *Wolbachia* infections. Some mutagen might be added to speed up this process. At any time we could check for incipient speciation by treating the infected flies with antibiotics and then testing for hybrid viability in crosses between flies with different genetic markers. The prediction is that flies that evolved under the heterogeneous environment along with *Wolbachia* infections should demonstrate more reproductive isolation than flies that had not been infected. These uninfected flies should, in turn, demonstrate more reproductive isolation than the infected flies that were kept in a homogeneous environment.

Note that I have turned the traditional speciation experiment on its head. The typical speciation experiment enforces some degree of post-zygotic barrier, such as killing the hybrid offspring between two populations. It then looks for the evolution of positive assortative mating in the resulting subpopulations (Rice & Hostert 1993). I am suggesting enforcing the pre-zygotic barriers, and then, after removing the pre-zygotic barrier with antibiotics, testing for the evolution of post-zygotic barriers.

A weakness of this design is the ambiguity in the analogy between *Wolbachia* infection and positive assortative mating based on reproductive characters. If it is more analogous to positive assortative mating on non-reproductive characters, then the infected flies raised in the homogeneous environment should demonstrate more reproductive isolation than the uninfected flies. However, they should still demonstrate less isolation than the uninfected flies raised in the homogeneous environment with high migratory barriers. The attraction of using *Wolbachia* is that mate "choice" can be manipulated by the application of antibiotics.

Chapter 8

Conclusions

> Whatever topic you select, your project should have three parts: the hypothesis, the part that goes after the hypothesis, and the conclusion (this should be the same as the hypothesis). (Barry 1998)

The preceding chapters have shown how a configuration model can illuminate dynamics that we have never seen before, but suspect must be out there somewhere in the darkness. The results of such models, and this one is certainly no exception, should be treated as guiding hypotheses for future investigation. To accept the results as the truth about biology would require accepting the abstraction of the model as a representation of the essential features of a biological system. Biological systems are too complex, and these models are also probably too complex to allow us any such confidence. Instead, I have presented evidence for why MoD may be a reasonable representation of ecological and evolutionary dynamics. We may thus, with caution, take the results as support for or against the hypotheses for the diversification of life.

8.1 Primary Results

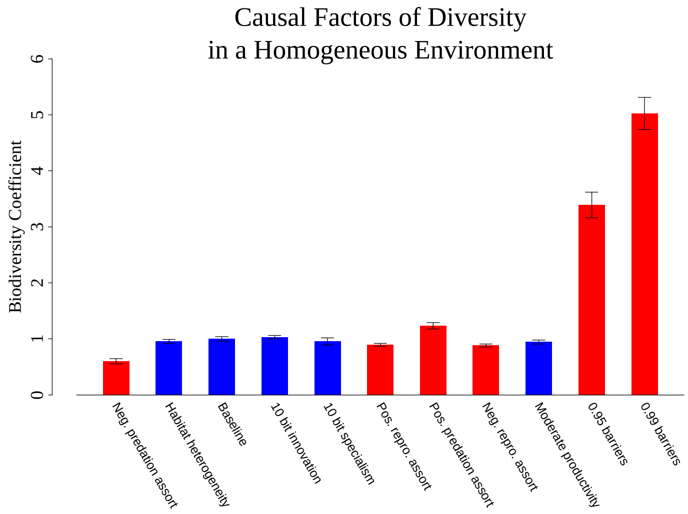
The most important innovation of MoD is its implementation of a species. While neither reproductive barriers (Saruwatari et al. 1994) nor reproductively isolated populations (Todd & Miller 1991) are novel, their combination is. This implementation of the reproductive species concept has allowed us to see that speciation may be far from a rare event. It is possibly occuring all around us every day, given this particular definition and implementation of species. What is rare is the survival and differentiation of those incipient species such that a naturalist might notice them.

There are five general statements that can be made about diversification in MoD. First of all, by far the most important factor that influences diversification is the isolation of populations from one another. In the model, this is represented by geographical isolation. Second, positive assortative mating can dramatically stimulate diversification. Positive assortative mating is similar to geographical isolation in that it also serves to isolate subpopulations. Third, the subdivision of niche space through the evolution of specialization can have significant effects on the processes of diversification. Fourth, evolutionary innovations in resource utilization and predation interactions have no discernible effect except to lower speciation rates. This result poses a challenge to the entire industry of looking for key ecological innovations to explain the diversification of various groups. However, as Cracraft (1990) has pointed out, the connection between ecological pressures and speciation is only indirect and so of questionable importance. Finally, the most direct form of divergent selection in the model is caused by habitat heterogeneity. Yet, this seems to have little effect on diversity. The relative importance of geographic isolation and unimportance of habitat differences directly contradicts evidence from laboratory experiments on flies (Rice & Hostert 1993).

An overview of the importance of the causal factors that have been examined appears in Figures 8-1 and 8-2. Only a representative subset of all the experiments is depicted. The heterogeneous habitat condition with high migratory barriers provides a better gauge of diversity dynamics than the vanilla condition. Diversity levels are too low in the vanilla condition to allow the effect of a manipulation to be detected.

The divergence between diversity and speciation rates can be seen in both the manipulations to productivity and specialization. When productivity is boosted to moderate levels, speciation rates dramatically increase but there is no effect on diversity levels. Extinction rates rise in tandem with speciation rates in this case. In contrast, allowing organisms to specialize results in a decrease in speciation rates but an increase in the standing diversity. The ability to specialize lowers extinction rates.

The counter-intuitive results of Chapter 7 show up most clearly in Figure 8-2. Both positive and negative assortative mating based on reproductive characters dramatically suppress speciation rates (and diversity) in a homogeneous environment (top chart). Yet,



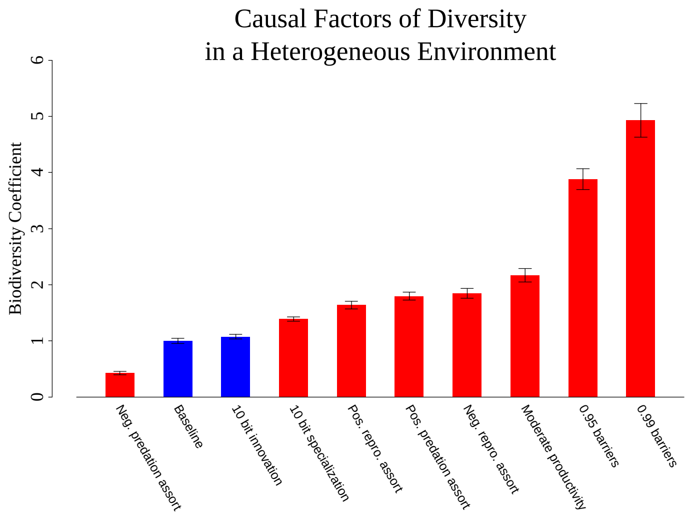


Figure 8-1: A summary of manipulations to species diversity. The top chart shows the manipulations in the context of a homogeneous environment with no migratory barriers. The bottom graph shows the same manipulations in the context of a heterogeneous environment and 0.95 migratory barriers. Each effect has been normalized by its relevant baseline and include standard error bars. In the case of the manipulations to the barriers, the relevant baseline is either the homogenous or heterogeneous habitats with no migratory barriers. The bars indicate the magnitude of an effect relative to the baseline. The blue bars indicate that the manipulation had no statistically significant effect. Red bars indicate a significant effect, with $p < 0.05$.

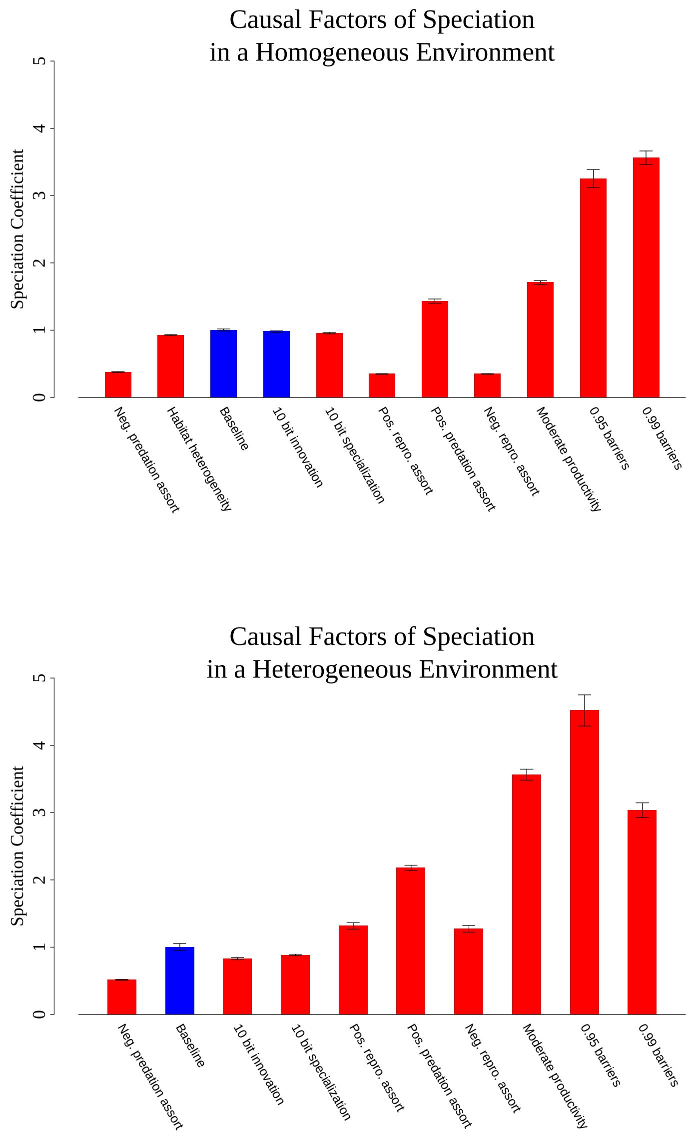


Figure 8-2: A summary of manipulations of speciation rates. The top chart shows the effect in the context of a homogeneous environment with no migratory barriers. The bottom graph shows the same manipulations in the context of a heterogeneous environment and 0.95 migratory barriers. Standard error bars have been added to the means of the effects. Differences between these graphs and those in Figure 8-1 illustrate the fact that diversity is determined by the interaction of both speciation and extinction rates. Each condition has been normalized by its relevant baseline. In the case of the manipulations to the barriers, the relevant baseline is either the homogenous or heterogeneous habitats with no migratory barriers. The blue bars indicate that the manipulation had no statistically significant effect. Red bars indicate a significant effect, with $p < 0.05$.

in a heterogeneous environment with migratory barriers (bottom chart), both forms of assortative mating significantly increase speciation rates (and diversity). This can only be understood through the dynamics of the reproductive graphs. Assortative mating based on the reproductive characters biases the dynamics of the graphs towards the maintenance of large internal nodes. Since only the death of all the organisms of an internal node can cause speciation, these biases tend to reduce speciation. However, in a subdivided population, which nodes are "peripheral" and which are "internal" can sometimes change. This may cause the reproductive graphs in the patches to pull apart from one another. The divergence of the graphs leads to an increase in speciation rates. While this may appear particular to the reproductive graphs used in MoD, the dynamics derive from the way in which relative rarity interacts with assortative mating and so may have broader relevance to other implementations of species and perhaps biology in general.

Fluctuations in an organism's physical environment does not appear to be important as a cause of diversification in and of itself. Amongst the abiotic influences, only geographical isolation showed a major, monotonic effect on diversity. Changes that increase the geographical isolation or fragmentation of populations are expected to stimulate diversification.

The model implies that there is a negative feedback dynamic in diversification that causes diversity levels to be regulated once they reach a sort of diversity carrying capacity. This is not due to a decline in total speciation rates per time step, as diversity rises. On the contrary, speciation rates rise with diversity. However, extinction rates also rise with diversity, and, crucially, they rise faster than the speciation rates. The implication that diversity is similarly regulated in the real world does not answer the question of where we are on that logistic curve. If, as Benton (1995) would have it, global diversity is increasing exponentially, that is still consistent with the early part of a logistic curve. On the other hand, if one hypothesizes that the diversity ceiling will be reached quickly, then the apparent continuous increase in species diversity must be explained by arguments for why the ceiling on diversity might have been raised over time.

It is difficult to raise the ceiling. Let us consider each of the hypotheses for diversification that MoD supports. Arguments for the importance of geographical isolation would have to explain why populations have been increasingly fragmented and isolated over time. Geologically this seems implausible. At the large scale, we know that continents have both fragmented and come back together over time. At a smaller scale, there is no obvious trend. Biologically, one would have to posit less and less dispersal over the history of life. If anything, with the development of multicellular life, the opposite is more likely to be true.

Proponents of sexual selection as a diversifying force would have to argue that the ceiling on diversity has been raised through the ever increasing capacity or tendency of organisms to discriminate between their potential mates. This argument might rest on the proposition that, on average and across species, organisms have become more structurally or behaviorally complex. That is a common assumption. However, it has yet to be proven. Certainly there is a higher variance in biological complexity now than there was 3.5 billion years ago. But to be fair, since we are counting species, and we have chosen a species definition based on sexual reproduction, we should restrict our attention to changes in complexity in sexual organisms. Again, we might argue that the variance in complexity has increased, but it is not at all clear that average complexity has increased. Not to mention the average propensity to be choosy about one's mate.

The argument for specialization as the factor that has elevated the diversity ceiling is very similar to the argument for sexual selection. We would have to argue that the capacity to specialize has increased over time. If the simple ability to specialize was a constant source of diversity, then we could not explain any regulation, not to mention the release of regulation, of species diversity. One must posit that the capacity to specialize at a given time led to a diversification which was eventually muted by a rising extinction rate. Then, further capacity to specialize evolved, allowing the packing of even more species into the niche space, and so diversity increased again until it hit a new, higher ceiling. An argument for the increasing capacity to specialize would probably have to be based on arguments of increasing developmental and behavioral complexity. Again, this is a tenuous position.

Finally, let us turn to the favored hypotheses regarding niche space. Perhaps niche space has been repeatedly enlarged by evolutionary innovations. Can niche space be enlarged? Niche space, at least *environmental niche* space, is the space of potential niches. Proponents of evolutionary innovations in the utilization of resources would not argue that the potential niche space was enlarged, just the amount of it that was occupied. An innovation allows organisms to expand into virgin territory. The ceiling effect must come from an additional assumption. We must assume that there are some form of barriers or topology in niche space that prevent species from expanding freely in that space. In other words, some innovations must be less likely than others. This may be a reasonable assumption given the complexities of biological development[1]. The hypothesis, then, is that organisms were constrained to a valley in niche space and that occasionally, an innovation arose allowing the lineages to break free from the valley into the surrounding lands. However, we must hypothesize that the organisms would soon run up against new obstacles in this territory that prevented any further expansion. Until, of course, the next innovation opened the gates to a successive radiation. The challenge comes in relating the filling of adaptive space with the splitting of gene pools. That connection is not at all clear and has been challenged by the model.

The only reason to argue for changes in the ceiling to diversity is the belief that life has butted up against that ceiling at various times in the past. If we narrow our attention to any one clade, or branch of the tree of life, we will often see it diversify up to some point and then, often, decline. However, we often cannot even tell if a specific clade reached a species diversity equilibrium because the fossil record generally does not allow resolution at the species level. The typical compromise is to count families as proxies for species diversity. This adds strong biases to the perception of the underlying diversity (Maley et al. 1997). Furthermore, the model, and its results, are based on species coevolving in an ecosystem. Most taxa that people study are not large enough to encapsulate more than one trophic level of the ecosystems that shaped those species. In summary, the jury is still out on whether diversity regulation has yet to play an important part in the history of life. Until there is better evidence for the regulation of diversity across trophic levels in evolutionary time, we should shave with Occam's razor, and favor the simpler assumption that the world has yet to reach the diversity ceiling.

In other words, MoD argues for a world view in which speciation is generally more likely than extinction. This diversification is presumed to be based on geographical isolation of populations, assortative mating, and specialization. Of course, there may well be other important factors in diversification that were not tested in MoD.

The results imply that when biologists seek an explanation for an adaptive radiation in some key innovation of that taxon, they should try to weigh potential changes in dispersal as well as secondary sexual characteristics equally with, if not more than, apparent ecological innovations. Furthermore, paleobiologists should give greater weight to reproductive

[1]For example, it is probably much less likely that my future children will develop an extra set of arms than they will develop a superior intelligence compared to their father.

characters in their attempts to subdivide fossils into groups of related individuals. Similarly, biologists of living organisms might look more closely at species that display strong sexual selection. What we think is an individual species may well be a complex of cryptic species.

8.2 Related Results

In the course of the experiments some of the complexity of predator-prey interactions has emerged. An initial, naive implementation of predation failed because rare prey were too easy to find by common predators. Three trophic levels were only stabilized when interference competition between predators was introduced.

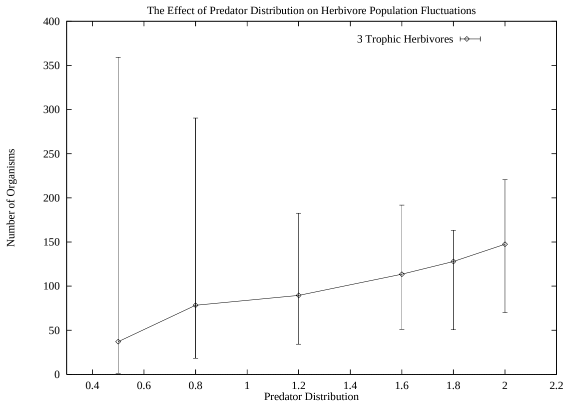


Figure 8-3: The average herbivore population sizes as a function of the *predation-distribution* parameter. The error bars indicate the average maximum and average minimum population sizes for 50 runs. As *predation-distribution* increases, adding more competitive interference between the herbivores, the oscillations in the herbivore population diminish and the average population size rises.

Chapter 3 illuminated some counter-intuitive dynamics in predator-prey interactions. When parameters of MoD were changed so as to make life "harder" for predators, the predator population, averaged over time, increased. This occured when the prey were more protected from predation by raising the *predation-distribution* parameter, lowering the probability of finding prey with the *prey-location* parameter, or restricting the patrolling of the predators with the *search-radius* parameter. All of these effects dampen the oscillations between the predator and prey populations and so stabilize both populations. The net result is an average boost to predator population size, as indicated by Figure 8-3.

Similarly, Chapter 6 illuminated a further complexity in the coevolution of predators and their prey. In general, MoD showed that the prey have an advantage in the evolutionary arms race. With a larger population, the prey have more opportunities for evolving novel forms of resistance to predation. Yet, when we restrict adaptive space, when we constrain the number of possible prey innovations, in some cases this allows the predators to keep up in the race. This implies that evolutionary models of predation are critically affected by the number of dimensions in the adaptive space.

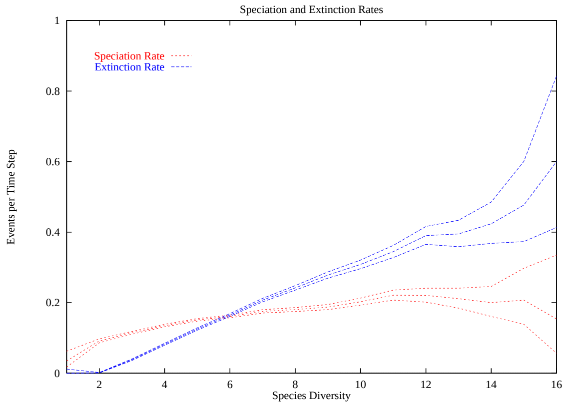


Figure 8-4: The interaction of speciation and extinction rates as functions of diversity. The means for both rates are surrounded by curves representing the 95% confidence intervals. In contrast to the theory of island biogeography, origination rates, i.e., speciation rates, do not decline with rising diversity.

Chapter 4 investigated the species-area curve from island biogeography in an evolutionary context. Rather than introducing new species through immigration from a mainland population, new species were generated by the model through speciation events within the ecosystem. This results in speciation rates that increase, roughly linearly, with diversity as shown in Figure 8-4. The only reason this does not lead to "run-away" diversification is that extinction rates also rise with diversity. Extinction rates rise faster than speciation rates. Where these two lines cross, the ecosystem settles down into a stable equilibrium diversity level. The scaling of diversity with area in MoD showed a far more realistic species-area curve than previous attempts (Jones et al. 1997). The difference probably derives from the fact that Jones et al. substituted genotypic diversity for species diversity. The more realistic species-area curve in MoD inspired the prediction that any ecological model that uses a reproductive graph definition of species will produce realistic species-area curves.

The experiments on the relationship between productivity and diversity provided another example of a typical ecological experiment transplanted into an evolutionary context. Experiments that have studied the effects of productivity on diversity levels last at most a few years (Tilman 1987; Goldberg & Miller 1990). Thus, any rise in diversity levels must come from immigrants entering the area of study. In contrast, MoD looks at the evolution new species in response to changes in productivity levels. The results match the ecological observations of a unimodal relationship between diversity and productivity. However, the explanation for this dynamic in the model is unlike any of the hypotheses reviewed by Rosenzweig (1995). Increases in productivity boost reproduction rates. This produces two countervailing forces for diversity. On the one hand, larger populations and faster turnover accelerates diversification through greater genetic change in the populations. On the other hand, the production of more individuals increases the number of migrants between populations. This combines with greater turnover in the populations to produce higher proportions of migrant genes entering populations. The barriers between populations are thus effectively lowered and the populations are less likely to diverge.

Finally, the results of Chapter 7 show that the effects of assortative mating on diversity are subtle. Speciation may be stimulated or suppressed depending on whether organisms are choosing mates based on characters that are independently involved in reproductive isolation or predation interactions. Furthermore, the effects may change depending on the sizes of the population. When a population is fragmented, the sampling error between the subpopulations means that a phenotype that is rare in the consolidated population may enjoy the benefits (or penalties) of being common in an occasional subpopulation. Assortative mating is a particularly simple form of sexual selection. Biologists only just begun to explore the complexities of sex.

8.3 Caveats

Biology has largely been a science of specifics. In an attempt to explore generalized dynamics, the model was constructed at a high level of abstraction. The reality of the situation is that there are probably many species and ecosystems that differ from the assumptions of MoD in crucial ways. So, we must first predicate the model, and all of its results, on the set of assumptions upon which it was built. Furthermore, MoD represents only one implementation, or instantiation of those assumptions. As a computational model, it is vulnerable to details of its construction that may significantly impact the results, unbeknownst to its creators. Thus, any attempt to replicate the assumptions, structure, or results of the model would be a valuable contribution.

MoD itself suffers from a number of weaknesses in its representation of the requirements for testing theories of diversity. We have been strict in our focus on species, and not morphological diversity. The complexity of the modeled ecosystems has been severely constrained. There are no parasites or decomposers in the model. The set of resources and the complexity of the underlying physical environment is minimal, practically vestigial. There are no nutrient cycling dynamics. Nor are there any representations of body size or mass of the organisms. Besides complexity issues, the spatial and temporal scales of MoD are not what one would hope. The results bear on diversification over thousands of generations within an isolated region containing only tens of thousands of individuals. Species diversity levels in MoD are nowhere near the magnitudes observed on Earth. In a time scale of millions of years, an extremely rare but dramatic innovation might indeed change the course of natural history. Though, probably not in this model, given its rather simplistic representation of ecology.

Finally, it is not clear that the different hypotheses for diversification have really received a fair hearing. How should we compare the degree of assortative mating with the degree of geographical isolation, or the degree of constraints on niche space? I have attempted to compare representative conditions for each manipulation, but computational constraints have precluded against a thorough exploration of parameter space[2]. In particular, divergent selection may not have been very strong in MoD. An organism perfectly adapted to the habitat in one patch and entering another patch in a heterogeneous environment could expect to find an average of 10.3 bits differences in the new habitat. However, that organism was likely to have wild-cards in its generalism loci for some of those bit positions. So we may be underestimating the impact of real habitat heterogeneity. Still, the competitive exclusion results of Section 3.3 did show that a single bit difference in a plant can make all the difference to that species' survival. So habitat heterogeneity was certainly not a trivial factor in the adaptation of species. But adaptation is a different issue from diversification.

[2] I should note that I have reported all of my investigations of parameter space. There is often the worry in such models with huge numbers of parameters, that the model is over-fit to the desired results. Has the experimenter explored a lot of parameter space and chosen one of a very few parameter settings that produce the desired results? In this case, no. Section 3.2 presented a fair amount of parameter space. For each parameter there was a reasonable portion under which the model behaved well. I chose one point in that space and pushed on, without further tuning.

8.4 Future Work

There are three distinct forms of future work that might result from this initial exploration. First off, there are many remaining questions that might be addressed through elaborations of this and other models. Secondly, as was mentioned above, MoD should be validated through alternative implementations. And finally, the results have provided hypotheses to be tested in the lab and field.

8.4.1 Further Explorations

As was mentioned above, it would be nice to explore the dynamics of the model over a wider spatial scale. It would also be interesting to look at population size effects. We know these are important because of the fact that 0.99 migratory barriers lead to the extinction of the animals. Unfortunately, the problem with expanding the populations within a patch is that the time required to compute intra-patch interactions grows quadratically ($O(n^2)$), whereas expanding the number of patches only increases processing time linearly ($O(n)$). Modifying the mating algorithm to only search a fixed number of locations in the organism array of a patch would change intra-patch computation to $O(n)$ processing time, but might also dramatically affect the growth of rare species, perhaps realistically (Bernstein et al. 1985). Still, this could be avoided with a better data structure for organizing the organisms within a patch, such as a hash table based on the reproductive chromosomes. Related to both the population sizes and the number of patches is the desire to run the model for more time steps. However, the model did appear to settle down to a diversity equilibrium fairly rapidly. So running the model for longer would only be justified under elaborations where rare events (like significant innovations) might change the diversity dynamics.

The environment might be made richer by introducing multiple habitat patterns representing different resources that might support autotrophs. This would allow the autotrophs to specialize within a patch. An elaboration of the complexity of the environment might mesh well with alterations that would provide for evolutionary innovations that would truly open up unoccupied niche space. If there were multiple resources in the environment, then a lineage could escape competition by evolving the use of an unexploited resource. Space within a patch might also be subdivided into separate compartments which could serve as differentiated resources themselves.

We might allow evolutionary innovations in the characters that have a direct impact on the creation and maintenance of isolated gene pools. Organisms could include a gene for dispersal tendencies. We then might examine the effects of reduction in dispersal tendencies upon the diversification of that lineage. Similarly, organisms could be given characters for mate preferences, rather than enforcing a particular kind of assortative mating on all organisms. We then might look at how geographical fragmentation interacts with the evolution of mating preferences and how, together, these factors affect diversification.

The coevolutionary dynamics of MoD have not yet been studied. It would be interesting to know if an enforced speciation event in a prey species would tend to result in a later speciation event in its predators. We could ask the opposite question. How does a speciation event in a predator species affect the diversification of its prey?

Further work might be carried out on different model platforms. The reproductive graph implementation of species could be easily ported to ECHO. We might then re-examine the species-area curve in ECHO to see how the curve changes when we abandon genotypic diversity as the measurement of species. The same could be done in most other artificial life models by adding reproductive chromosomes to the organisms.

Finally, a full analysis of the reproductive graphs in MoD proved intractable. Appendix A lays out some of the open problems in the analysis of reproductive graphs. The results of this, and similar, models depends crucially on the implementation of a species. Section 4.5 made a start at examining the effect of modifying the restrictions on mating between organisms. There remains a host of questions to be answered. How does the number of loci in the reproductive chromosomes (or dimensions of the hypercube that constrains the reproductive graphs) interact with diversification dynamics as well as reproductive barriers. Would an expansion to more than just two possible alleles at a locus change the dynamics qualitatively? How do these manipulations interact with the segregation of the total population into semi-isolated subpopulations? A better approach to these problems might be to construct a model whose analysis would be more tractable. This would mean avoiding such complications as examining 8 potential mates in assortative mating. It would probably be better to either implement a model where all potential mates are examined, or perhaps use some probabilistic function with “nice” analytical behavior.

8.4.2 Alternative Implementations

One of the most powerful theoretical validations of the model's results would be the creation of an alternative model that matched the requirements from Section 2.1 and still produced the same qualitative results. If, instead, that alternative model should produce qualitatively different results, we would have a basis upon which to seek the explanation for that difference. Chapter 2 describes the details of my implementation in sufficient detail that this specific model should be replicable. We might then begin altering the two models, the replicate and the alternative model, bringing them closer together until the crucial detail is found that changes the dynamics of the model.

This suggests a further approach to understanding the fundamental issues. Starting from a model, such as this one, we might systematically remove pieces, bit by bit, until the dynamics qualitatively change. Such an investigation by amputation should also help to identify the necessary pieces of a model of diversification.

Finally, there may well be a distribution model that could be constructed to test these same hypotheses. The discovery of such a model would be an important advance because it may provide us with further analytical tools for understanding and generalizing the dynamics.

8.4.3 Experimental Tests

Support or, more importantly, criticism of the model's results might be gained through traditional observational techniques in paleobiology and systematics. Has the evolution of leking behavior, where a female chooses a mate from a crowd (lek) of males on display, been associated with shifts of diversification rates in phylogenetic trees? More indirectly, have the elaboration of secondary sexual characteristics been associated with shifts in diversification?

One of the fundamental predictions of MoD is that proto-species are being formed at a relatively high frequency. Subpopulations are probably being reproductively isolated and then going extinct before they are able to evolve differentiating characters for us to notice. To test this, ideally, we would like to understand the reproductive system of a species so well that we could construct a reproductive graph of the individuals in a population. Perhaps from a DNA sample or similar non-lethal assay of a pair of individuals, we could predict the probability of their successfully mating and producing viable offspring. If we knew this, then we could sample a population, construct a graph of who can mate with whom, and then detect if there are reproductively isolated gene pools within that population. This sounds a bit far fetched now, but with the level of genetic detail we have on a number of model systems, such as yeast (*Saccharomyces cerevisiae*) or perhaps even fruit flies (*Drosophila*), such a level of understanding might not be too far distant in the future. It might be possible today in some plants where we can examine the so called self incompatibility loci. These are the genes that determine which pollen can fertilize which ova.

Finally, more ambitious experiments that closely mimic the manipulations of the model might be carried out in a laboratory setting. Section 7.5.1 suggested an experiment on fruit flies. Yeast is probably a better model organism for most of these experiments because it has a much faster generation time than fruit flies. Although, it lacks the behavior complexity necessary for some sexual selection experiments. It seems strange that most speciation experiments have been carried out on flies (Rice & Hostert 1993), though it may be an indication of the level of our knowledge of mating compatibility in yeasts. They now seem relatively simple. Still, an automated system could be set up to deliver migrant spores between populations and different populations might be exposed to different selection pressures through toxins, temperatures, and nutrient sources. Mutagens might be used to boost the exploration of adaptive space while viruses might stand in for predators, if natural predators are inconvenient. Analogs of most of the experiments from Chapter 5 should be implementable. Productivity is easily manipulated through the input of nutrients to the cultures. An important benefit of working with yeast is that we should be able to ascertain the genetic basis for any reproductive barriers that evolve between populations.

8.5 Valediction

I hope that, by now, the value of computation models in theoretical biology is obvious. We have been able to throw out some hypotheses, such as the importance of flux in the environment, and emphasize others that have suffered from a paucity of attention, such as sexual selection. Computational models give us another looking glass through which to see nature and also reflect upon some of the biases in our other tools. Yet, this new looking glass is certainly no less distorted than the others. It is merely distorted in different ways. The models, and their abstractions, provide a method for searching for the systematic and fundamental in biology. Is there anything there to be discovered? Will biology always remain the study of the particular? That is the question.

Appendix A

Open Problems in Reproductive Graphs

We would like to know the expected number of components (species) in an evolving reproductive graph. These are graphs with nodes representing the different reproductive genotypes in the current population of the model. The graph starts with one node for each initial species. Over time, neighbor nodes appear due to mutations in the reproductive genes of the live organisms. At the same time, nodes disappear due to the death of all the organisms that carry the reproductive genotype represented by that node. These graphs are constrained to a 64-dimensional binary hypercube in the model. However, generalization to n-dimensional quadrary hypercubes may have more relevance to biological species, where each dimension corresponds to the four possible nucleic acids at a locus (G,A,T, and C).

One way to think about this is to consider the space of all possible graphs on the hypercube. If the binary hypercube has n dimensions, then there are 2^n possible nodes and so the power set of these nodes represents all 2^{2^n} possible graphs. However, it is not true that in one time step the model could transition to any other possible graph. The evolution of these graphs are constrained by the fact that only neighbors of current nodes may appear in the next time step. Thus, we might construct a Markov process with 2^{2^n} states along with the probabilities of making transitions between all these states.

The determination of these probabilities is non-trivial. Even if we ignore the bias to the probabilities introduced by predation (e.g., some organisms are more likely to die in a time step than others), it is still true that the probability of the loss of a node is a function of the number of organisms represented by that node. Furthermore, the probability of the appearance of a new node is a function of the number of organisms represented by neighboring nodes. In fact, the story is even more complex. Let us ignore the effects of the limits on searching for a mate, as well as the spatial partitioning of organisms into patches. Given these simplifications, the probability of a new node appearing is actually a function of the number of organisms in the neighboring nodes as well as the number of organisms in the neighbors of those neighbors. To see this, consider a two node graph with one populous node and a new mutant node. Now consider all the potential neighboring nodes to the mutant that are not neighbors of the populous node. The chance of these nodes appearing in the next time step is not simply a function of the number of organisms (1) in the mutant node. If the "populous" node also only has 1 organism, then there will be at most 2 new organisms produced in the next time step. Those new organisms might be mutants with reproductive genes that are neighbors of the parental mutant node. But since there are at most 2 new organisms, the chance of this is low. However, if the populous node has many organisms, then the 1 organism from the mutant node might engage in many mating events and so the probability of producing neighboring nodes would be higher.

Problem 1 *Given a mutation rate per mating m, a death rate per organism per time step d, a constant population size N, and starting with a homogeneous population in a single node, what are the expected number of components in the graph at equilibrium?*

The addition of the number of organisms represented by each node adds another dimension to the states of the Markov process. Each node in the graph could have from 1 to N organisms. This leads to a $N2^{2^n}$ state Markov process. If we assume a carrying capacity of the environment, then we may assume that N is constant. In the model, N is typically around $20,000$.

There are a number of critical elaborations to Problem 1:

Problem 2 *What are the expected number of components given different definitions of adjacency in the hypercube? What if nodes are connected when they have $\leq k$ bits different, can the hypercube results be generalized to these more complex topologies?*

Problem 3 *What are the expected number of components given positive and negative assortative mating? That is, given the tendency to mate with organisms in the same node (positive) or in neighboring nodes (negative).*

Problem 4 *What are the expected number of components given spatial structure with its concomitant limitations on the mating pool and migration between mating pools?*

A simplified version of these problems may be more tractable. Let us elide the underlying population dynamics and just assign birth and death rates to the nodes.

Problem 5 *Given a birth rate m for the probability that a new node will appear next to an extant neighbor, a death rate per node per time step d, and starting with a single node, what are the expected number of components in the graph at equilibrium?*

This can be explicitly computed for low n where n is the number of bits in the reproductive gene. Let us assume, in accordance with the model, that an extant node can give birth to neighbors and die in the same time step. However, a node that is born in a time step may not die or create new neighbors in that time step. Then the probability a reproductive graph in state s_1 makes a transition to state s_2 in a time step can be decomposed into the product of the individual dynamics for all the nodes and potential nodes in the hypercube. There are four possible cases for each node:

1. **A node is born:** The chance that a node appears in a time step is a function of the number of living neighbor nodes x:

$$Pr[\text{birth}] = 1 - (1 - b)^x \tag{A.1}$$

2. **A node is unborn:** The chance that a dead, or non-existent, node remains that way is just 1 minus the chance that it is born:

$$Pr[\text{not born}] = (1 - b)^x \tag{A.2}$$

3. **A node dies:** The chance that a node dies is just the node's probability of disappearing, d, times the probability that it is not replaced by a mutation from a neighboring node. That is d times the probability of no birth:

$$Pr[\text{death}] = d(1 - b)^x \tag{A.3}$$

4. **A node survives:** The chance that a living node survives into the next time step

is the chance that it either does not die or that it dies but is born again due to a mutation from a neighboring node:

$$Pr[\text{survival}] = (1 - d) + d\,(1 - (1 - b)^x) \tag{A.4}$$

The product of these probabilities over all 2^n nodes in the binary hypercube gives the state transition probability for the Markov process.

I have computed the $n = 2$ case of the Markov process merely to check if the abstraction is reasonable. A sign of a bug in the abstraction would be a high probability for an equilibrium at a single node graph. We know from running the model that the number of reproductive genotypes at any given time is about two orders of magnitude greater than the number of species.

The results for computing the Markov process with $n = 2$ depend on the relation of the birth probability b and the death probability d for the nodes. If the birth rate is relatively high, then the system is most likely to settle into the full 4-node graph. The 3-node graphs are the next most likely, and so on, except that there is a fair chance that the process gets stuck in the 0-node graph. The 0-node graph is an absorbing boundary on the process. If the death rate is relatively high, then, predictably, the process tends to get stuck in the 0-node state. In between these extremes, there tends to be a bimodal distribution in the probability for the different states. The 0 and 4-node states tend to collect most of the probability, with the other graphs garnering a smear of the remaining probabilities in between. In other words, the results for $n = 2$ show reasonable behavior.

Unfortunately, since the problem scales super-exponentially, even moderate values of n are not computable.

Appendix B

Derivations

B.1 Bit Differences in a Heterogeneous Environment

To generate a heterogeneous environment, I flipped 8 bits in the habitat bit pattern of every patch. A bit was selected by generating a random number with uniform probability from 1 to 32. The selected bit was flipped in the current patch. This was repeated 7 more times. The same bit could be flipped and then flipped back to its original value. Each patch was treated independently. The question then arises, what is the expected number of bits that differ between any two patches. This is important because it determines the expected disadvantage a plant would suffer if it was perfectly adapted to the habitat of its parent's patch but migrated to a new patch.

The best way to think of this is question is to model it with a birth-death Markov process (Drake 1967). First off, the difference between two patches, each with 8 bits flipped, is equivalent to the difference between the original state of a patch and its state after 16 bits have been flipped. If we start with all bits set to 0, as we did in the model, then we can represent the possible states of the patch's habitat by states in the Markov process labeled with the number of 1's that have been set. If the habitat pattern has i 1's in it, then the chance that the next bit flip adds another 1 is $(32 - i)/32$. And similarly, the chance that the bit chosen to be flipped is already a 1, and so is changed to a 0, is $i/32$. One end of this Markov chain is shown in Figure B-1.

To determine the probability of being in state i after n bit flips, we first construct a matrix M where the (i, j)-th entry indicates the probability of moving from state i to state j in one step. These are the values labeling the arrows in Figure B-1. Raising this matrix

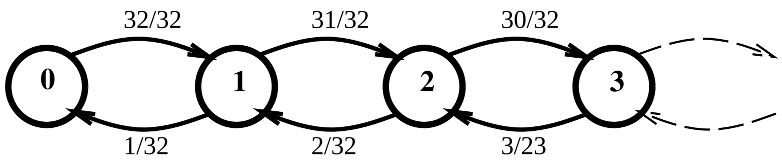


Figure B-1: A simple birth-death Markov process to model the number of bits set to 1 in a 32 bit habitat pattern for a patch. Only the first 4 out of 33 states are shown.

to the n-th power results in a matrix where the (i, j)-th entry is the probability of ending in state j if you started in state i and took n random steps. Since we know that we start in state 0, the zeroth row of the matrix gives the probabilities of being in each of the 33 possible states. Then to find the expected state we use the formula for the expectation of a probabilistic event:

$$E[\text{state}] = \sum_{j=0}^{32} jM(0, j) \tag{B.1}$$

For this problem $E[\text{state}] = 10.3$. On average, patches will differ in 10.3 bits under the heterogeneous environment condition. The progression of the number of bits set to 1, starting from all 0's, is shown in Figure B-2. There is an asymptote at 16 because a 32-bit string, when fully randomized by an infinite number of bit flippings, has on average half of its bits set to 1 and the other half to 0.

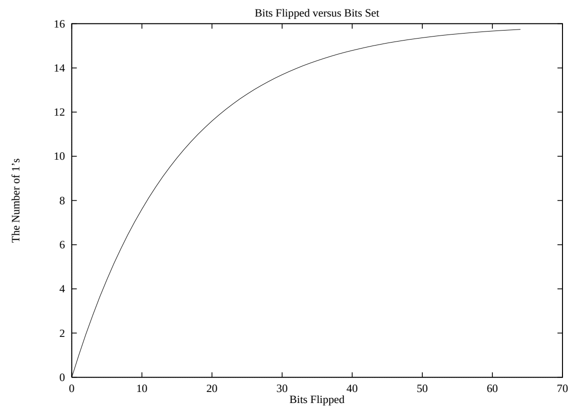


Figure B-2: The expected number of bits set to 1 after flipping a given number of bits. We start with all bits "cleared," set to 0.

B.2 Migration Rates and the Fixation Index

First let us derive a rough estimate of the number of migrants that might be exchanged between neighboring patches with 0.9 migratory barriers, in one generation. Once we have an estimate for the number of migrants, we can derive the fixation index. This will be a function of the migration rate. And finally, we will be able to calculate the fixation index by plugging in the rate of migration that we will have calculated.

B.2.1 Number of Migrants

What is the probability that a plant seed will land in a particular neighboring patch? First of all, the seed takes a random walk of 0 to 2 steps on the grid of patches. So 1/3 of the seeds will take exactly one step. With 0.9 migratory barriers, only 1 in 10 of these seeds will make it over the barrier. To be conservative, we will consider the worst case, when the patch of origin is not on the edge of the environment, and so it has 9 neighbors (including itself). In general, the plant populations were limited by the carrying capacity of their patches, so a new seed only had a 1 in 2 chance of replacing an endemic plant in the destination patch. The product of all these probabilities is:

$$\textit{origin} \overset{1/3}{\rightarrow} \textit{one step} \overset{1/10}{\rightarrow} \textit{over barrier} \overset{1/9}{\rightarrow} \textit{destination} \overset{1/2}{\rightarrow} \textit{take root} = \frac{1}{540} \tag{B.2}$$

Since we expect 1000 plants to reproduce in a generation, this results in about 2 migrants that take one step into a particular neighboring patch. Added to this are the seeds that attempt to take 2 steps and end up in that same neighboring patch. There are two cases for these 2-steppers. First, a seed might be blocked on the first step with probability 9/10, and then proceed in the next step, exactly as the one-steppers. Second, a seed might make it into the neighboring patch on the first step, but get trapped there on the second step with probability 9/10. Any two-stepper who first moves into a different patch and then moves into the destination patch would have to cross the migratory barriers twice. The chance of doing that is so small that they can be safely ignored. The total of the one-stepper and two-stepper migrants is:

$$1000\left(\frac{1}{540} + \frac{9}{10}\left(\frac{1}{540}\right) + \frac{9}{10}\left(\frac{1}{540}\right)\right) \approx 5 \tag{B.3}$$

B.2.2 The Fixation Index in a Haploid Population

The fixation index is generally calculated for a diploid population (Hartl & Clark 1997, p.194). That is, a population that has two copies of each chromosome, like humans. However, the formulas can easily be derived for a haploid population, with half the alleles of the analogous diploid population.

In a haploid population with N individuals, there are N alleles. We let m be the proportion of new alleles introduced into the population through migration. Thus, if we choose two alleles randomly from this population, we can ask what is the chance, F_t, that the two alleles in generation t are copies of the same allele in some previous generation? we assume that if either allele is from a migrant, then the two alleles cannot be identical by descent. There are two possible cases for identity by descent. In one case, neither allele is one of the new migrant alleles, $(1-m)^2$, but they both are copies from an allele in the previous generation. This is equivalent to randomly choosing the same allele twice from the previous generation. The probability that the second allele chosen comes from the same one as the first allele chose is $1/N$. In the second case, the two alleles are not migrant alleles and they don't come from the same allele in the previous generation, $(1-m)^2(1-1/N)$, but the two ancestral alleles from the previous generation are themselves identical by descent, with probability F_{t-1}. This leads to the recurrence relation:

$$F_t = (1-m)^2\frac{1}{N} + (1-m)^2\left(1-\frac{1}{N}\right)F_{t-1} \tag{B.4}$$

To solve this we look for the equilibrium value of $\hat{F}$ when $F_t = F_{t-1}$. We can get an elegant and reasonable approximation if we throw out all terms with m^2 and m/N. These terms are too small to matter.

$$\hat{F} \approx \frac{1}{N} + \left(1-2m-\frac{1}{N}\right)\hat{F} \tag{B.5}$$

$$\hat{F} \approx \frac{1}{1+2Nm} \tag{B.6}$$

Since Nm is exactly the number of new migrants in the population we can plug in the value $Nm = 5$ as calculated above to get $F \approx 0.09$.

B.2.3 Migration Rates

What is the proportion of migrants in a generation of new organisms in a patch? Note that migrants do not disappear off the edge of the environment. Since the number of migrants is conserved by the model, the average proportion of immigrants into a patch should equal the average proportion of emigrants leaving the patch. It is slightly easier to analyze the amount of emigration.

Consider a migratory barrier b which is assumed to be high ($1 - b \ll 1$). This lets us ignore the rare cases where a migrant crosses a barrier on both steps of its random walk. We can split up the population of new organisms into three equal classes: the stay-at-home types, the 1-steppers, and the 2-steppers, corresponding to random walks of 0, 1, and 2 steps. The stay-at-home types contribute nothing to the population of emigrants. The 1-steppers contribute $1 - b$ of their number, and the 2-steppers contribute $1 - b^2$ of their number. So the total proportion of emigrants m in the population of new organisms (not the parental organisms) is:

$$m = \frac{1}{3}(1 - b) + \frac{1}{3}(1 - b^2) \tag{B.7}$$

If $b = 0.95$ then $m = 0.049$, and if $b = 0.9$ then $m = 0.097$. So, for high migratory barriers, the proportion of emigrants, which is also the proportion of immigrants, is very close to the proportion that can get over the migratory barrier in one step $(1 - b)$.

References

Alvarez, L. W., Alvarez, W., Asaro, F., & Michel, H. V. (1980). Extraterrestrial cause for the Cretaceous-Tertiary extinction. *Science*, **208**, 1095–1108.

Anderson, R. M. & May, R. M. (1982). Coevolution of hosts and parasites. *Parasitology*, **85**, 411–426.

Andersson, M. (1994). *Sexual Selection.* Princeton, NJ: Princeton University Press.

Ausich, W. I. & Bottjer, D. J. (1985). Phanerozoic tiering in suspension-feeding communities on soft substrata: Implications for diversity. In: *Phanerozoic Diversity Patterns: Profiles in Macroevolution* (Valentine, J. W., ed), pp. 255–274. Princeton, NJ: Princeton University Press.

Axelrod, R. (1984). *The Evolution of Cooperation.* New York, NY: Basic Books.

Bambach, R. K. (1985). Classes and adaptive variety: The ecology of diversification in marine faunas through the Phanerozoic. In: *Phanerozoic Diversity Patterns: Profiles in Macroevolution* (Valentine, J. W., ed), pp. 191–253. Princeton, NJ: Princeton University Press.

Banzhaf, W., Nordin, P., Keller, R. E., & Francone, F. D. (1998). *Genetic Programming.* San Francisco, CA: Morgan Kaufmann Publishers, Inc.

Barry, D. (1998). Bad science. *The Boston Globe Magazine*, pp. 8–9.

Baum, D. A. & Donoghue, M. J. (1995). Choosing among alternative "phylogenetic" species concepts. *Systematic Botany*, **20**, 560–573.

Bedau, M. A., Ronneburg, F., & Zwick, M. (1992). Dynamics of diversity in an evolving population. In: *Parallel Problem Solving from Nature, 2* (Männer, R. & Manderick, B., eds), pp. 95–104. Amsterdam, The Netherlands: Elsevier Science Publishers.

Begon, M., Harper, J. L., & Townsend, C. R. (1990). *Ecology.* London: Blackwell Scientific Publications.

Belding, T. C. (1995). The distributed genetic algorithm revisited. In: *Proceedings of the Sixth International Conference on Genetic Algorithms* (Eshelman, L. J., ed), pp. 114–121, San Francisco, CA: Morgan Kaufmann.

Benton, M. J. (1990). The causes of the diversification of life. In: *Major Evolutionary Radiations* (Taylor, P. D. & Larwood, G. P., eds), pp. 409–430. Oxford, UK: Clarendon Press.

Benton, M. J. (1995). Diversification and extinction in the history of life. *Science*, **268**, 52–58.

Bernstein, H., Byerly, H. C., & Michod, R. E. (1985). Sex and the emergence of species. *Journal of Theoretical Biology*, **117**, 665–690.

Booker, L. B. (1985). Improving the performance of genetic algorithms in classifier systems. In: *Proceedings of the First International Conference on Genetic Algorithms* (Grefenstette, J. J., ed), pp. 80–92, Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Booth, G. (1997). Gecko: A continuous 2-d world for ecological modeling. *Artificial Life*, **3**, 147–164.

Boyajian, G. F. (1986). Phanerozoic trends in background extinction: Consequence of an aging fauna. *Geology*, **14**, 955–958.

Brenchley, P. J. (1989). The late Ordovician extinction. In: *Mass Extinctions: Processes and Evidence* (Donovan, S. K., ed), pp. 104–132. New York: Columbia University Press.

Brown, J. L. (1997). A theory of mate choice based on heterozygosity. *Behavioral Ecology*, **8**(1), 60–65.

Bull, J. J., Badgett, M. R., Wichman, H. A., Huelsenbeck, J. P., Hillis, D. M., Gulati, A., & Molineux, I. J. (1997). Exceptional convergent evolution in a virus. *Genetics*, **147**, 1497–1507.

Bull, L. & Fogarty, T. C. (1996). Evolutionary computing in multi-agent environments: Speciation and symbiogenesis. In: *Parallel Problem Solving from Nature–PPSN IV* (Ebeling, W., Rechenberg, I., Schwefel, H.-P., & Voigt, H.-M., eds), p. 12, Berlin: Springer-Verlag.

Caswell, H. (1976). The validation problem. In: *Systems Analysis and Simulation in Ecology, vol. IV* (Patten, B. C., ed), pp. 313–325. New York, NY: Academic Press, Inc.

Caswell, H. (1989). *Matrix Population Models: Construction, Analysis, and Interpretation.* Sunderland, MA: Sinauer Associates.

Caswell, H. & John, A. M. (1992). From the individual to the population in demographic models. In: *Individual-Based Models and Approaches in Ecology* (DeAngelis, D. & Gross, L., eds), pp. 36–61, New York: Chapman and Hill.

Charnov, E. L. (1993). *Life History Invariants.* Oxford, UK: Oxford University Press.

Cohoon, J. P., Hegde, S. U., Martin, W. N., & Richards, D. (1987). Punctuated equilibria: A parallel genetic algorithm. In: *Proceedings of the Second International Conference on Genetic Algorithms* (Grefenstette, J. J., ed), pp. 148–154, Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Collins, R. & Jefferson, D. (1992). The evolution of sexual selection and female choice. In: *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life* (Varela, F. J. & Bourgine, P., eds), pp. 327–336, Cambridge, MA: MIT Press.

Colwell, R. K. (1992). Niche: A bifurcation in the conceptual lineage of the term. In: *Keywords in Evolutionary Biology* (Keller, E. F. & Lloyd, E. A., eds), pp. 241–248. Cambridge, UK: Harvard University Press.

Cormen, T. H., Leiserson, C. E., & Rivest, R. L. (1990). *Introduction to Algorithms.* Cambridge, MA: MIT Press.

Cornell, H. V. (1993). Unsaturated patterns in species assemblages: The role of regional processes in setting local species richness. In: *Species Diversity in Ecological Communities* (Ricklefs, R. E. & Schluter, D., eds), pp. 243–252. New York, NY: University of Chicago Press.

Cowen, R. (1990). *History of Life.* Oxford, UK: Blackwell Scientific Publications.

Cracraft, J. (1983). Species concepts and speciation analysis. *Current Ornithology*, **1**, 159–187.

Cracraft, J. (1985). Biological diversification and its causes. *Annals of the Missouri Botanical Garden*, **72**, 794–822.

Cracraft, J. (1989). Speciation and its ontology. In: *Speciation and its Consequences* (Otte, D. & Endler, J. A., eds), pp. 28–59. Sunderland, MA: Sinauer Associates.

Cracraft, J. (1990). The origin of evolutionary novelties: Pattern and process at different hierarchical levels. In: *Evolutionary Innovations* (Nitecki, M. H., ed), pp. 21–44. Chicago, IL: The University of Chicago Press.

Craighurst, R. & Martin, W. (1995). Enhancing GA performance through crossover prohibitions based on ancestry. In: *Proceedings of the Sixth International Conference on Genetic Algorithms* (Eshelman, L. J., ed), pp. 130–135, San Francisco, CA: Morgan Kaufmann.

Crawley, M. J. (1992). Population dynamics of natural enemies and their prey. In: *Natural Enemies: The Population Biology of Predators, Parasites, and Diseases* (Crawley, M. J., ed), pp. 40–89. Oxford, UK: Blackwell Scientific Publications.

Cronin, H. (1991). *The Ant and the Peacock*. Cambridge, UK: Cambridge University Press.

Dall, S. R. X. & Cuthill, C. (1997). The information costs of generalism. *Oikos*, **80**, 197–202.

Darwin, C. R. (1859). *On the Origin of Species*. London: John Murray.

de Olivera, A. K. & Codeiro, A. R. (1980). Adaptation of *drosophila willistoni* experimental populations to extreme ph medium. II. development of incipient reproductive isolation. *Heredity*, **44**, 123–130.

de Queiroz, K. (1998). The general lineage concept of species, species criteria, and the process of speciation. In: *Endless Forms: Species and Speciation* (Berlocher, S. & Howard, D. J., eds), Oxford, UK: Oxford University Press. in press.

Deb, K. & Goldberg, D. E. (1989). An investigation of niche and species formation in genetic function optimization. In: *Proceedings of the Third International Conference on Genetic Algorithms* (Schaffer, J. D., ed), pp. 42–50, San Mateo, CA: Morgan Kaufmann.

Drake, A. W. (1967). *Fundamentals of Applied Probability Theory*. New York, NY: McGraw-Hill, Inc.

Durrett, R. & Levin, S. A. (1994). The importance of being discrete (and spatial). *Theoretical Population Biology*, **46**, 363–394.

East, I. R. & Rowe, J. (1996). Effects of isolation in a distributed population genetic algorithm. In: *Parallel Problem Solving from Nature–PPSN IV* (Ebeling, W., Rechenberg, I., Schwefel, H.-P., & Voigt, H.-M., eds), pp. 408–419, Berlin: Springer-Verlag.

Ehrman, L. (1965). Direct observation of sexual isolation between allopatric and between sympatric strains of the different *drosophila paulistorum* races. *Evolution*, **19**, 459–464.

Eriksson, O. & Bremer, B. (1992). Pollenation systems, dispersal modes, life forms, and diversification rates in angiosperm families. *Evolution*, **46**, 258–266.

Erwin, D. H. (1994). The Permo-Triassic extinction. *Nature*, **367**, 231–236.

Erwin, D. H., Valentine, J. W., & Sepkoski, Jr., J. J. (1987). A comparative study of diversification events: The early Paleozoic versus the Mesozoic. *Evolution*, **41**(6), 1177–1186.

Eshelman, L. J. & Schaffer, J. D. (1991). Preventing premature convergence in genetic algorithms by preventing incest. In: *Proceedings of the Fourth International Conference on Genetic Algorithms* (Belew, R. K. & Booker, L. B., eds), pp. 115–122, San Mateo, CA: Morgan Kaufmann.

Feder, J. L., Stolz, U., Lewis, K. M., Perry, W., Roethele, J. B., & Rogers, A. (1997). The effects of winter length on the genetics of apple and hawthorn races of Rhagoletis pomonella (Diptera: Tephritidae). *Evolution*, **51**, 1862–1876.

Flessa, K. J. & Jablonski, D. (1985). Declining Phanerozoic background extinction rates: Effect of taxonomic structure? *Nature, London*, **313**, 216–218.

Flessa, K. W., Erben, K. R., Hallam, A., Hsü, K. J., Hüssner, H. M., Jablonski, D., Raup, D. M., Sepkoski, Jr., J. J., Soulé, M. E., Sousa, W., Stinnesbeck, W., & Vermeij, G. J. (1986). Causes and consequences of extinction. In: *Patterns and Processes in the History of Life* (Raup, D. M. & Jablonski, D., eds), pp. 235–257. Berlin: Springer-Verlag.

Foley, D. H., Beebe, N., Torres, E., & Saul, A. (1996). Misidentification of a phillippine malaria vector revealed by allozyme and ribosomal DNA markers. *American Journal of Tropical Medicine & Hygiene*, **54**, 46–48.

Fox, W. (1951). Relationships among the garter snakes of the *thamnophis elegans rassenkreis*. *University of California publications in zoology*, **50**, 485–530.

Fry, J. D. (1996). The evolution of host specialization: Are trade-offs overrated? *The American Naturalist*, **148**, S84–S107.

Futuyma, D. J. (1986). *Evolutionary Biology*. Sunderland, MA: Sinauer Associates.

Futuyma, D. J. & Peterson, S. C. (1985). Genetic variance in the use of resources by insects. *Annual Review of Entomology*, **30**, 217–238.

Gavrilets, S. (1997). Evolution and speciation on holey adaptive landscapes. *Trends in Ecology and Evolution*, **12**(8), 307–312.

Gavrilets, S. & Gravner, J. (1997). Percolation on the fitness hypercube and the evolution of reproductive isolation. *Journal of Theoretical Biology*, **184**, 51–64.

Giordano, R., Jackson, J. J., & Robertson, H. M. (1997). The role of *wolbachia* bacteria in reproductive incompatibilities and hybrid zones of *diabrotica* beetles and *gryllus* crickets. *Proceedings of the National Academy of Sciences*, **94**, 11439–11444.

Goldberg, D. E. (1989). *Genetic Algorithms: In Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley.

Goldberg, D. E. & Miller, T. E. (1990). Effects of different resource additions on species diversity in an annual plant community. *Ecology*, **71**, 213–225.

Gould, S. J. & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist program. *Proceedings of the Royal Society of London B*, **205**, 581–598.

Grimmett, G. (1989). *Percolation*. New York, NY: Springer-Verlag.

Haldane, J. B. S. (1927). On being the right size. In: *Possible Worlds and other essays* (Haldane, J. B. S., ed). London, UK: Chatto & Windus.

Hallam, A. (1990a). Biotic and abiotic factors in the evolution of early Mesozoic marine molluscs. In: *Causes of Evolution: A Palaeontological Perspective* (Ross, R. M. & Allmon, W. D., eds). Chicago, IL: The University of Chicago Press.

Hallam, A. (1990b). The end-Triassic mass extinction event. In: *Global Catastrophes in Earth History* (Sharpton, V. L. & Ward, P. D., eds), pp. 577–583. Boulder, CO: Geological Society of America.

Hamilton, W. D., Axelrod, R., & Tanese, R. (1990). Sexual reproduction as an adaptation to resist parasites (a review). *Proceedings of the National Academy of Sciences*, **87**, 3566–3573.

Hartl, D. L. & Clark, A. G. (1997). *Principles of Population Genetics*. Sunderland, MA: Sinauer Associates, Inc.

Heard, S. B. & Hauser, D. L. (1995). Key evolutionary innovations and their ecological mechanisms. *Historical Biology*, **10**, 151–173.

Hennig, W. (1966). *Phylogenetic Systematics*. Urbana, IL: University of Illinois Press.

Herráiz, C. I., Merelo, J. J., Olmeda, S., & Prieto, A. (1997). Biodiversity through sexual selection. In: *Artificial Life V* (Langton, C. G. & Shimohara, K., eds), pp. 308–318, Cambridge, MA: MIT Press.

Hoffman, A. (1986). Neutral model of Phanerozoic diversification: Implications for macroevolution. *N. Jb. Geol. Palaont. Abh.*, **172**, 219–244.

Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems*. Cambridge, MA: MIT Press.

Holland, J. H. (1993). Echoing emergence: Objectives, rough definitions, and speculations for echo-class models. Technical Report 93-04-023, Santa Fe Institute.

Hraber, P. T. & Milne, B. T. (1997). Community assembly in a model ecosystem. *Ecological Modelling*.

Hunter, J. P. (1998). Key innovations and the ecology of macroevolution. *Trends in Ecology and Evolution*, **13**(1), 31–36.

Hurst, G. D. D. & Schilthuizen, M. (1998). Selfish genetic elements and speciation. *Heredity*, **80**, 2–8.

Hurst, L. D. (1991). The incidences and evolution of cytoplasmic male killers. *Proceedings of the Royal Society of London B*, **244**, 91–99.

Huston, M., DeAngelis, D., & Post, W. (1988). New computer models unify ecological theory. *Bioscience*, **38**(10), 682–691.

Huston, M. A. (1994). *Biological Diversity*. Cambridge, UK: Cambridge University Press.

Jablonski, D. (1989). The biology of mass extinctions: A palaeontological view. *Philosophical Transactions of the Royal Society of London B*, **325**, 357–368.

Jablonski, D. & Bottjer, D. J. (1990). The ecology of evolutionary innovation: The fossil record. In: *Evolutionary Innovations* (Nitecki, M. H., ed), pp. 253–288. Chicago, IL: The University of Chicago Press.

Jones, T., Hraber, P. T., & Forrest, S. (1997). The ecology of echo. *Artificial Life*, **3**, 165–190.

Judson, O. P. (1997). The rise of the individual-based model in ecology. *Trends in Ecology and Evolution*, **9**(1), 9–14.

Kauffman, S. A. (1993). *Origins of Order.* Oxford, UK: Oxford University Press.

Kirkpatrick, M. & Ryan, M. J. (1991). The evolution of mating preferences and the paradox of the lek. *Nature*, **350**, 33–38.

Knoll, A. H., Bambach, R. K., Canfield, D. E., & Grotzinger, J. P. (1996). Comparative earth history and late permian mass extinction. *Science*, **273**, 452–457.

Knuth, D. E. (1981). *The Art of Computer Programming, vol. 2.* Reading, MA: Addison-Wesley.

Krukonis, G. P. (1996). Evolutionary potential of generalists and specialists: An experimental study with viruses. In: *Bulletin of the Ecological Society of America*, volume 77 (3 SUPPL. PART 2), p. 245, Washington, DC: The Ecological Society of America.

Larson, A. (1989). The relationship between speciation and morphological evolution. In: *Speciation and its Consequences* (Otte, D. & Endler, J. A., eds), pp. 579–598. Sunderland, MA: Sinauer Associates.

Lenski, R. E. & Travisano, M. (1994). Dynamics of adaptation and diversification: A 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences*, **91**, 6808–6814.

Lindgren, K. & Nordahl, M. G. (1994). Artificial food webs. In: *Artificial Life III* (Langton, C. G., ed), pp. 73–104, Redwood, CA: Addison-Wesley.

Liou, L. L. & Price, T. D. (1993). Speciation by reinforcement of premating isolation. *Evolution*, **48**, 1451–1459.

MacArthur, R. H. & Wilson, E. O. (1967). *The Theory of Island Biogeography.* Princeton, NJ: Princeton University Press.

Maley, C. (1997). Mutation rates as adaptations. *Journal of Theoretical Biology*, **186**, 339–348.

Maley, C. C., Robeck, H., & Donoghue, M. (1997). Simulations of the impact of classification systems on diversity patterns. In: *The Geological Society of America 1997 Annual Meeting Abstracts* (Carlson, N., ed), p. 31, Boulder, CO: GSA.

Margulis, L. & Sagan, D. (1986). *Origins of Sex.* New Haven, CT: Yale University Press.

Marzluff, J. M. & Dial, K. P. (1991). Life history correlates of taxonomic diversity. *Ecology*, **72**, 428–439.

Mathiopoulos, K. D. & Lanzaro, G. C. (1995). Distribution of genetic diversity in relation to chromosomal inversions in the malaria mosquito *anopheles gambiae*. *Journal of Molecular Evolution*, **40**, 578–584.

May, R. M. & Watts, C. H. (1992). The dynamics of predator-prey and resource-harvester systems. In: *Natural Enemies: The Population Biology of Predators, Parasites, and Diseases* (Crawley, M. J., ed), pp. 431–457. Oxford, UK: Blackwell Scientific Publications.

Mayr, E. (1942). *Systematics and the Origin of Species.* New York: Columbia University Press.

Mayr, E. (1963). *Populations, Species, and Evolution*. Cambridge, MA: Harvard University Press.

Mayr, E. (1982a). *The growth of biological thought : diversity, evolution, and inheritance*. Cambridge, MA: Belknap Press.

Mayr, E. (1982b). Speciation and macroevolution. *Evolution*, **36**(6), 1119–1132.

McGhee, Jr., G. R. (1989). The Frasnian-Famennian extinction event. In: *Mass Extinctions: Processes and Evidence* (Donovan, S. K., ed), pp. 133–151. New York: Columbia University Press.

McPeek, M. A. (1996). Trade-offs, food web structure, and the coexistence of habitat specialists and generalists. *The American Naturalist*, **148**, S124–S138.

Mishler, B. D. & Theriot, E. (1998). Monophyly, apomorphy, and phyologenetic species concepts. In: *Species Concepts and Phylogenetic Theory: A Debate* (Wheeler, Q. D. & Meier, R., eds). New York, NY: Columbia University Press.

Mitchell, M. (1996). *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press.

Mooers, A. O. & Heard, S. B. (1997). Inferring evolutionary process from phyologenetic tree shape. *The Quarterly Review of Biology*, **72**, 31–54.

Oliver, C. G. (1972). Genetic and pheontypic differentiation and geographic distance in four species of *lepidoptera*. *Evolution*, **26**, 221–241.

Orr, H. A. (1995). The population genetics of speciation: The evolution of hybrid incompatibilities. *Genetics*, **139**, 1805–1813.

Paterson, A. M. & Gray, R. D. (1997). Diversification functions and the rate of taxonomic evolution. In: *Host-Parasite Evolution* (Clayton, D. H. & Moore, J., eds), pp. 236–250. Oxford, UK: Oxford University Press.

Paterson, H. E. H. (1985). The recognition concept of species. In: *Species and Speciation* (Vrba, E. S., ed), pp. 21–29. Pretoria: Transvaal Museum Monograph No. 4.

Raup, D. M. (1972). Taxonomic diversity during the Phanerozoic. *Science*, **177**, 1065–1071.

Raup, D. M. (1991). *Extinction: Bad Genes or Bad Luck?* New York, NY: W. W. Norton and Co.

Raup, D. M. & Boyajian, G. E. (1988). Patterns of generic extinction in the fossil record. *Paleobiology*, **14**, 109–125.

Raup, D. M., Gould, S. J., Schopf, T. J. M., & Simberloff, D. S. (1973). Stochastic models of phylogeny and the evolution of diversity. *The Journal of Geology*, **81**, 525–542.

Raup, D. M. & Sepkoski, Jr., J. J. (1982). Mass extinctions in the marine fossil record. *Science*, **215**, 1501–1503.

Ray, T. S. (1992). An approach to the synthesis of life. In: *Artificial Life II* (Langton, C. G., Taylor, C., Farmer, J. D., & Rasmussen, S., eds), pp. 371–408, Reading, MA: Addison-Wesley.

Rice, W. R. (1985). Disruptive selection on habitat preference and the evolution of reproductive isolation: An exploratory experiment. *Evolution*, **39**, 645–656.

Rice, W. R. & Hostert, E. E. (1993). Laboratory experiments on speciation: What have we learned in 40 years. *Evolution*, **47**, 1637–1653.

Ricklefs, R. E. & Schluter, D. (1993). Species diversity: Regional and historical influences. In: *Species Diversity in Ecological Communities* (Ricklefs, R. E. & Schluter, D., eds), pp. 350–363. New York, NY: University of Chicago Press.

Robeck, H., Maley, C., & Donoghue, M. (1998). Simulations of the impact of classification systems on diversity patterns. Ms. in preparation.

Rosensweig, M. L. (1995). *Species Diversity in Space and Time.* Cambridge, UK: Cambridge University Press.

Rousset, F. & Solignac, M. (1995). Evolution of single and double *wolbachia* symbioses during speciation in the *drosophila simulans* complex. *Proceedings of the National Academy of Sciences*, **92**, 6389–6393.

Roy, K. & Foote, M. (1997). Morphological approaches to measuring biodiversity. *Trends in Ecology and Evolution*, **12**, 277–281.

Sanderson, M. J. & Donoghue, M. J. (1994). Shifts in the diversification rate with the origin of angiosperms. *Science*, **264**, 1590–1593.

Sanderson, M. J. & Donoghue, M. J. (1996). Reconstructing shifts in diversification rates on phylogenetic trees. *Trends in Ecology and Evolution*, **11**, 15–20.

Saruwatari, T., Toquenaga, Y., & Hoshino, T. (1994). Biodiversity through sexual selection. In: *Artificial Life IV* (Brooks, R. A. & Maes, P., eds), pp. 424–429, Cambridge, MA: MIT Press.

Schluter, D. (1996). Ecological causes of adaptive radiation. *The American Naturalist*, **148**, S40–S64.

Schmitz, O. J. & Booth, G. (1997). Modeling food web complexity. *Evolutionary Ecology.*

Sepkoski, Jr., J. J. (1984). A kinetic model of Phanerozoic taxonomic diversity. III. post-paleozoic families and mass extinctions. *Paleobiology*, **10**, 246–267.

Sepkoski, Jr., J. J., Bambach, R. K., Raup, D. M., & Valentine, J. W. (1981). Phanerozoic marine diversity and the fossil record. *Nature, London*, **293**, 435–437.

Sepkoski, Jr., J. J. & Kendrick, D. C. (1993). Numerical experiments with model monophyletic and paraphyletic taxa. *Paleobiology*, **19**, 168–184.

Sheehan, P. M. (1991). Patterns of synecology during the Phanerozoic. In: *The Unity of Evolutionary Biology* (Dudley, E. C., ed), pp. 103–118. Portland, OR: Dioscorides Press.

Shoemaker, D. D. & Ross, K. G. (1996). Effects of social organization on gene flow in the fire ant Solenopsis invicta. *Nature*, **383**, 613–616.

Signor, P. W. (1982). Species richness in the Phanerozoic: Compensating for sampling bias. *Geology*, **10**, 625–628.

Simpson, G. G. (1961). *Principles of Animal Taxonomy.* New York: Columbia University Press.

Slowinski, J. B. & Guyer, C. (1990). Testing the stochasticity of patterns of organismal diversity: An improved null model. *American Naturalist*, **134**, 907–921.

Slowinski, J. B. & Guyer, C. (1991). Comparisons of observed phylogenetic topologies with null expectations among three monophyletic lineages. *Evolution*, **45**, 340–350.

Slowinski, J. B. & Guyer, C. (1993). Testing whether certain traits have caused amplified diversification. *The American Naturalist*, **142**, 1019–1024.

Subbarao, S. K. & Sharma, V. P. (1995). Anopheline species complexes and malaria control. *Indian Journal of Medical Research*, **106**, 164–173.

Tanese, R. (1989). Distributed genetic algorithms. In: *Proceedings of the Third International Conference on Genetic Algorithms* (Schaffer, J. D., ed), San Mateo, CA: Morgan Kaufmann.

Templeton, A. R. (1989). The meaning of species and speciation: A genetic perspective. In: *Speciation and its Consequences* (Otte, D. & Endler, J. A., eds), pp. 3–27. Sunderland, MA: Sinauer Associates.

Thompson, J. N. (1994). *The Coevolutionary Process.* Chicago, IL: University of Chicago Press.

Tilman, D. (1987). Secondary succession and the pattern of plant dominance along experimental nitrogen gradients. *Ecological Monographs*, **57**, 189–214.

Todd, P. M. & Miller, G. F. (1991). On the sympatric origin of species: Mercurial mating in the quicksilver model. In: *Proceedings of the Fourth International Conference on Genetic Algorithms* (Belew, R. K. & Booker, L. B., eds), pp. 547–554, San Mateo, CA: Morgan Kaufmann.

Todd, P. M. & Miller, G. F. (1997). Biodiversity through sexual selection. In: *Artificial Life V* (Langton, C. G. & Shimohara, K., eds), pp. 289–299, Cambridge, MA: MIT Press.

Valentine, J. W. (1980). Determinants of diversity in higher taxonomic categories. *Paleobiology*, **6**, 444–450.

Valentine, J. W. (1985). Biotic diversity and clade diversity. In: *Phanerozoic Diversity Patterns: Profiles in Macroevolution* (Valentine, J. W., ed), pp. 419–424. Princeton, NJ: Princeton University Press.

Valentine, J. W. & Moores, E. M. (1972). Global tectonics and the fossil record. *Journal of Geology*, **80**, 167–184.

Valentine, J. W. & Walker, T. D. (1986). Diversity trends within a model taxonomic hierarchy. *Physica D*, **22**, 31–42.

Van Valen, L. (1976). Ecological species, multispecies, and oaks. *Taxon*, **25**, 223–239.

Van Valen, L. M. (1973). A new evolutionary law. *Evolutionary Theory*, **1**, 1–30.

Van Valen, L. M. (1984). A resetting of Phanerozoic community evolution. *Nature, London*, **307**, 50–52.

Vermeij, G. J. (1987). *Evolution and Escalation.* Princeton, NJ: Princeton University Press.

Walker, T. D. (1985). Diversification functions and the rate of taxonomic evolution. In: *Phanerozoic Diversity Patterns: Profiles in Macroevolution* (Valentine, J. W., ed), pp. 311–334. Princeton, NJ: Princeton University Press.

Walker, T. D. & Valentine, J. W. (1984). Equilibrium models of evolutionary species diveristy and the number of empty niches. *The American Naturalist*, **124**(6), 887–899.

Werren, J. H. (1997). Biology of *wolbachia*. *Annual Review of Entomology*, **42**, 587–609.

White, M. J. D. (1978). *Modes of Speciation.* San Francisco, CA: W. H. Freeman and co.

Wiley, E. O. (1978). The evolutionary species concept reconsidered. *Systematic Zoology*, **27**, 17–26.

Wimsatt, W. C. (1987). False models as means to truer theories. In: *Neutral Models in Biology* (Nitecki, M. H., ed), pp. 23–55. Princeton, NJ: Oxford University Press.

About the Author

The author currently resides in Somerville, MA, with his two cats and a close friend, although he will be pulling up roots soon to take up a post-doctoral position at the University of New Mexico with Prof. Stephanie Forrest studying the evolution of diseases and epidemiology models. He received a B.A. in both psychology and computer science from Oberlin College in 1991. In the summer of 1991, he enjoyed a Summer Student Fellowship working the lab of Hal Caswell in the biology department of the Woods Hole Oceanographic Institute. Following that, he took up a Marshall scholarship at New College, University of Oxford, and received an M.Sc. in Zoology in 1993. His tenure at MIT in the EECS department has been considerably eased through the award of a National Defense Science and Engineering Graduate Fellowship. The author has published on papers on DNA computation, the coevolution of mutation rates, the impact of classification on diversity patterns, the evolution of virulence in parasites, as well as Tom Ray's Tierra model.