Proposal for Modeling the Evolution of Biodiversity

C.C. Maley

September 27, 1996

1 The Problem

A large class of outstanding problems in Evolutionary Biology concern the gap between our understanding of macroevolutionary dynamics like speciation, and microevolutionary dynamics like natural selection, migration, and genetic drift. Darwin[7] gave us the basis of a solid understanding of microevolutionary dynamics, but so far, macro-evolution lacks the analogous grounding. Computational simulations hold out the hope of examining the macro-evolutionary dynamics that emerge from an implementation of the micro-evolutionary dynamics. In the end, I hope to fill the gap between micro and macro-evolution.

There are two dramatic aspects of life as we know it; adaptations and diversity. Darwin explained the origin of adaptations, and his ideas are now being used in engineering to 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?

1.1 Diversification

There are three major obstacles to understanding diversification; the metric upon which diversity will be measured, the poor quality of data in the fossil record, and the impracticality of implementing controlled experiments.

Diversity might be measured either by morphological differences between organisms or by the number of different extant species. Unfortunately, morphological complexity and variation is difficult to quantify. Most work in the field has focused on species diversity, despite the fact that there does not exist a species definition that unambiguously partitions organisms into mutually exclusive groups. The problems with defining a species should not obscure the fact that most organisms would group into clusters if one were to plot them in some multidimensional character space. This is probably due to reproductive barriers between gene pools that prevent the mixing of genes from different species. In other words, it is reasonable to use the concept of a 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. In this work, I will employ an analogous implementation of a species based on reproductive barriers.

The fossil record is notorious for its poor quality of data. For our purposes, 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.

It is often impossible to distinguish between fossilized species. Since families are relatively arbitrary and historical groupings of species, there is a layer of noise between observational data and the 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 patch 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 the fossil record, subject to the above difficulties, as well as limited experimental studies. Darwin himself made extensive use of animal breeding to sharpen his understanding of evolution. A complementary 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. In this case, we are proposing to develop a model of the evolution of organisms in order to study the processes 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. 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 speciation and extinction, and is there some form of regulation on the numbers of extant species?

1.1.1 Speciation

There are a number of theories describing the process of speciation. It has been hard to distinguish between them because it is difficult to observe the process as it unfolds. Most of our evidence for speciation events is post hoc. Speciation theories tend to group into those that call upon geographical separation, e.g., allopatric speciation, and those that posit the creation of a species as a subpopulation within a larger population, e.g., sympatric speciation.

The best theory we have for the process of speciation focuses on geographical separation of populations. Allopatric 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. However, the observation of diversity of similar species across island chains tends to support the allopatric speciation model.

There is little evidence to support sympatric speciation. The most 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.

1.1.2 Extinction

Extinction, on the other hand, is thought to mainly occur through stochastic fluctuations in small populations. These fluctuations may be enhanced through environmental fluctuations, or simply through the stochastic variations of prey, predators, or other species in the food web.

1.1.3 Regulation of Diversity

If it turns out, as some have suggested, 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 growth. It is curious that this view is so widely held in the absence of data. To be fair, analyses of many groups of organisms, like North American Mammals, show relatively constant rates of diversity. Other work suggests correlations between origination (not necessarily at the species level) and extinction rates. However, recent work by Benton[2] suggests that if one looks across a large collection of organisms, species diversity has been increasing exponentially over time.

It is possible that both sides are right. For any given non-universal taxon, species diversity within that taxon approaches an equilibrium. This is supported by the species-area curve from island biogeography which proposes that species diversity of an island remains relatively constant. On the other hand, across all taxa, species diversity may continuously increase. Proponents of an exponential diversification curve point out that every new species provides a new form of resource for other species. In this way the number of potential niches may grow over time. 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 logarithmically, 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, we are left with at least two questions. What factors regulate diversity in non-universal taxa? And, what factors lift the ceiling on diversity?

1.1.4 Factors Influencing Diversity

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 up diversity beyond its previous levels?[1]

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. These explanations are broadly classified into four categories. Diversification patterns may (1) simply be artifacts in the data. (2) Caused by physical factors in the environment. (3) Caused by biological factors in the environment. And finally, (4) the consequence of neutral phenomena. This last class can be understood as rejections of the logistic assumption of diversification. This opens up simpler explanations based on the hypothesis that speciation is more likely than extinction.

2 The Approach

Our approach to the problems of the diversification of life is to construct a model of the microevolutionary processes and then experiment on the model in order to understand how the macroevolutionary processes emerge.

2.1 Methodology

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

1. Isolate a question.
2. Identify the minimal requirements.
3. Build and Run a model from those requirements.
4. Validate 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. Those minimal requirements can then be built into a model. 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 help validate the model.

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 appear biologically implausible. This is a second, though weak, form of validation. 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.

This methodology has the benefit of providing useful results regardless of the biological plausibility of the results. 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 out all 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.

3 The Model

I have isolated three basic hypotheses that need to be tested. Under what conditions is speciation more likely than extinction at the microevolutionary level? If so, are there diversity dependent factors that influence speciation and or extinction rates? If there are, what factors influence these diversity dependent factors, thus raising (or lowering) the diversity equilibrium ceiling?

3.1 Requirements

A model of diversification must at the least contain the following elements:

1. Organisms
2. Reproduction
3. Mutation
4. Death
5. Species

Representations of organisms are the basis for all microevolutionary models. These organisms must be organized by some principle into species. Given the dominant reproductive concept of species, this requires the implementation of sexual reproduction and reproductive barriers. Furthermore, in order to implement allopatric speciation, topology and migration are required. Six of the factors driving diversification reviewed by Benton [1] can be tested with the addition of climate and specialization of species resource utilization. This introduces the concept of a niche to the model.

The addition of climate and niches would allow us to test:

1. Cladistic inevitability.
2. The "sum of untold numbers of microevolutionary processes"
3. Increased endemnicity (provinciality) as a result of global topological or climate changes.
4. Fragmentation of physical habitats (topography and climate), not necessarily increasing endemnicity (barriers).
5. Continuous change in physical environments (barriers and climates change.)
6. "Increase in overall adaptive space occupied by a clade."
7. Subdivision of niches.
8. Red Queen driven specialization

as explanations for the diversification of life.

3.2 Implementation

Theoretical biology traditionally garnered analytic power by clustering individual organisms into classes that could be treated as a single entity. That is, all organisms within a class were assumed to have identical states and so could be handled as a single unit. The simplest example of this is an exponential growth equation for a population.

$$dN = (\text{birth\_rate} - \text{death\_rate})N \tag{1}$$

In this case, all individuals in the population are effectively treated as identical and the entire population grows or shrinks as a unit. More sophisticated models segregate the population along dimensions of age, sex, or any other relevant characteristic for the model [4]. Unfortunately, these distribution models are often become intractable with the addition of biological complications. In particular, the addition of spatial structure with local interactions between organisms, a la cellular automata, cannot be handled by a distribution model. This is because the state of an organism depends on the states of its neighbors. This precludes clustering individuals into classes of effectively identical organisms. Thus, the state and evolution of such a system depends on the configuration of the interactions between individuals. Caswell and John draw this distinction between distribution models and configuration models [5]. Configuration models are often called "individual-based" models, but this falsely implies that distribution models are not based on the individuals. For this reason, I will adhere to the terminology introduced by Caswell and John.

We may now ask the question, can we adequately model diversification with a distribution model? Certainly, extinction may be modeled merely by maintaining a count of the surviving organisms in a species. Speciation is a different kettle of organisms. Given the reproductive definition of species, speciation can only happen through the evolution of a reproductive barrier. This might be cast as a traditional distribution model if we assume that a small subpopulation has acquired both a mutation that distinguishes it from the dominant population as well as some form of reproductive barrier to prevent mixing. Then the model would examine if the new species would expand in the presence of the current food web of species. In this way, the question is reduced to the question of the successful invasion of a mutant. This begs the question of why speciation might be more likely than extinction, although it might be a useful avenue of exploration for determining if qualities of food webs impose diversity dependent influences on invasion and extinction rates.

We argued in section 1.1.1 that speciation tends to be driven by geographical isolation. Consequently, the model should include a representation of topology. In order to model migration and mating over a topology, we are forced to work with a configuration model where individuals interact through the 3M's: migration, mutation, and mating. All italicized terms below are parameters in the model. Unless otherwise noted, they represent integers.

Resource Utilization: In ecosystems, organisms provide resources for other organisms. That is, organisms eat each other. Furthermore, some organisms specialize on a small set of resources (e.g., Koala bears only eat eucalyptus leaves) while others are generalists and can subsist on a wide variety of resources (e.g., humans eat just about everything). To capture these dynamics and thereby set up a food web, we will need three bit pattern genes of equal length:

Phenotype: A bit pattern representing the structure and makeup of the organisms.

Prey Template: A bit pattern specifying the phenotype of an organism that can serve as food.

Generalism: A bit pattern mask specifying which bits in the prey template are not necessary for the selection of a prey organism.

The question of whether or not an organism can eat another organism is answered by applying the predator's generalism mask (e.g., 0011) to both the predator's prey template gene (e.g., 1010) and the prey's phenotype gene (e.g., 1011). The masking operation sets a bit in the target gene to 1 at every location that the mask has a 1. Bits at positions where the generalism mask has a 0 are left unmodified. The modified prey template gene (1011) is then compared to the modified prey's phenotype gene (1011) and if they match, then the predator may consume the prey organism.

The bottom of the food web is populated by autotrophs, organisms that synthesize their own food like plants. Instead of matching their prey template and generalism genes against a prey organism, autotrophs match them against the climate of their environment. The climate of a local is represented by a bit pattern (see below). A partial match results in the consumption of a fraction of the energy available from the environment. This energy is modeled by a particular number of meals granted to the autotroph.

Given this implementation alone, one might expect all lineages to evolve towards greater generalism. The fact that we see specialists in biology seems to imply that there is some tradeoff 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 genetic drift in genes for utilizing resources that aren't currently present and so are not under selective pressure for maintenance. In other words, specialists have no advantage over generalists. There is some evidence for this [**?**, **?**]. The real story is bound to be a lot more complicated. Thompson 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 [11]. 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, we will use the simplest assumption, that there is no tradeoff between specialism and generalism.

Mutation: At the time of conception of a new organism, every bit in the genes of the new organism will have a small probability of being flipped.

Death: Unlike taxes, death only comes in two forms. An organism may be eaten by a predator or it may starve to death. Starvation is modeled by requiring all organisms to eat at least one meal in the last *maximum-fast* time steps, where *maximum-fast* is a parameter of the model. Plants are the only organisms that may garner more than one meal in a time step.

Reproduction: Once *energy-conversion* number of meals have been consumed an reproduction attempt is triggered. If an organism can locate a viable mate in its vicinity, they will produce a new organism. The determination of which organisms are viable mates forms the representation of reproductive barriers. Initially, this will be implemented by prohibiting organisms from mating that have a hamming distance between their genes of greater than *species-radius*. This means that the relation of potential mating is not transitive. That is, organism A may mate with B, assuming a one bit difference, and organism B may mate with C, also assuming a one bit difference, but A may not be able to mate with C, assuming a more than one bit difference between their genomes. While this may create problems in classifying organisms into species, it reflects a similar phenomenon observed in biology. 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 the extremes of the gradient do not produce viable offspring.

There is evidence that reproductive barriers are not so much the result of the degree of difference between two organisms, but rather differences in the particular characters that influence mating. In other words, two populations may diverge over many characters, but if the characters that govern mating do not diverge, the two populations will remain one species. Thus, a second version of the model will be implemented with a species identification bit pattern gene. The hamming distance reproductive barrier will only be applied to this gene, thus allowing the other genes to vary within a species. The bits of the species identity gene are assumed to represent the set of characters influencing reproduction such as mating season, reproductive morphology, mate choice, etc. Under this implementation of reproductive barriers, individuals from opposite extremes of an environmental gradient would be allowed to mate as long as they shared similar species identity genes. It is conceivable that individuals in a species might differ radically in their prey preferences. However, due to the recombination of the parental genes, the hybrid offspring may not be viable. This constitutes an implicit or *de facto* reproductive barrier called a post-mating reproductive barrier.

If a pair of organisms are from the same species, they may mate. The new organism's genes are constructed by forming the parent's genes into two rings and then implementing cross-over between the two parental genomes. This avoids the bias towards separating genes at the ends of a linear genome. The new organism's genome then undergoes mutation where every bit has a *mutation-rate* chance of flipping. Finally, the number of accumulated meals of the original parent is reset to 0.

Migration: At birth, a new organism performs a random walk of length *dispersal* starting at the location of the parent. If there is space available in the location at the end of the walk, the new organism is placed there. Otherwise it dies. The random walk models the observed dynamic that the probability of dispersal over a distance falls off approximately quadratically with distance from the parent.

Topology: The random walk in the migration of a newborn organism is modified by the topological barrier value of each location. When an organism attempts to move from one location to a neighboring location, a random number is generated and compared to the barrier value of the neighboring location. If the random number is greater than the barrier value, the organism enters that location. However, if the random number is equal to or less than the barrier value, the organism remains in its current location. This process is repeated for each step of the random walk.

Climate: Each location has a bit pattern representing its climate. These bits are correlated with the barrier value of the location. This correlation simulates the relationship between altitude and climate. Furthermore, each location also has an *energy-modifier* value which may reduce the amount of energy available to autotrophs in that location. This allows the simulation of islands as areas with high *energy-modifiers* surrounded by lifeless areas with extremely low *energy-modifiers.*

One further computational limitation must be added. As the model stands, an unlimited number of plants could survive in any one locale. This is because, each plant is granted a certain amount of energy, regardless of the number of other plants in proximity. However, real plants are in constant competition for space and direct sunlight. So far, there has been no implementation of size because it is not an obvious aspect of the questions about diversification.[1] We might still avoid that complexity by putting a limit on the absolute number of plants in a location. The other organisms in a location will be limited by the amount of available food in the form of plants and other organisms.

[1]This is not to say that size is unimportant in biological interactions. In fact, there is a whole body of literature focusing on the importance of size in the life histories of individuals. Size is clearly critical for parasites and endosymbionts. However, initially there does not seem to be sufficient justification to introduce such complexity.

4 The Results

We are primarily interested in biodiversity. Our basic metric will be a count of the number of species extant at a given time. How do we identify species in a population of individuals? The hamming distance implementation of reproductive barriers can be used to structure the population into a graph. Each organism is a node in the graph, and each arc in the graph represents a the ability of the two organisms to mate. Our initial approach will be to identify as species the largest sub-graphs such that there is a path from every node in the sub-graph to every other node in the sub-graph. These sub-graphs represent distinct gene pools. Mutations may spread through a species sub-graph via reproduction, but cannot spread beyond.

4.1 Speciation vs. Extinction

The process of speciation differs depending on the implementation of reproductive barriers. Under the genomic difference implementation, speciation occurs when an organism dies and thereby severs the last potential reproductive link between surviving members of the ancestral species. The connected sub-graph of the species is thus split into two separate connected subgraphs, as diagrammed in Figure 1.

Alternatively, individuals may be allowed to mate even if they differ radically in their non-mating or species identity gene. If the bits of the species identity gene differ in more than one position, the organisms are unable to mate. In this way, there may be considerable genetic variation within a species. It is certainly possible in this case that recombination between radically different parental genes produces an organism is not viable in the environment of its parents. In any case, the process of speciation is similar to the process of speciation under the genomics difference implementation of reproductive barriers. Figure 2.

Under what conditions is speciation more likely than extinction? This can be tested by tracing species diversity over time under environmental conditions which will be systematically varied from experiment to experiment. Since speciation intimately involves geographical separation, we will examine speciation and extinction rates under a variety of landscapes from smooth to rugged. Furthermore, since extinction as well as genetic drift is enhanced by fluctuations in population size, speciation and extinction rates will also be investigated under a variety of climatic conditions from a fixed climate to a highly variable climate.

4.2 Diversity Dependent Factors

Assuming that speciation is generally more likely than extinction for relatively low levels of diversity, we will then investigate potential changes in speciation and extinction rates as diversity levels rise. Is there a diversity "carrying capacity" such that extinction rates rise and/or speciation rates decline with greater diversity? This can be separated from the previous investigations into the base speciation and extinction rates if those investigations are limited to relatively few species existing at one time.

Of course, there exists an a priori diversity ceiling. Given the amount of energy input into the model, let us assume that the plants are perfectly matched to their local climates. There is an upper limit to the number of plants in a location. The *energy-conversion* parameter

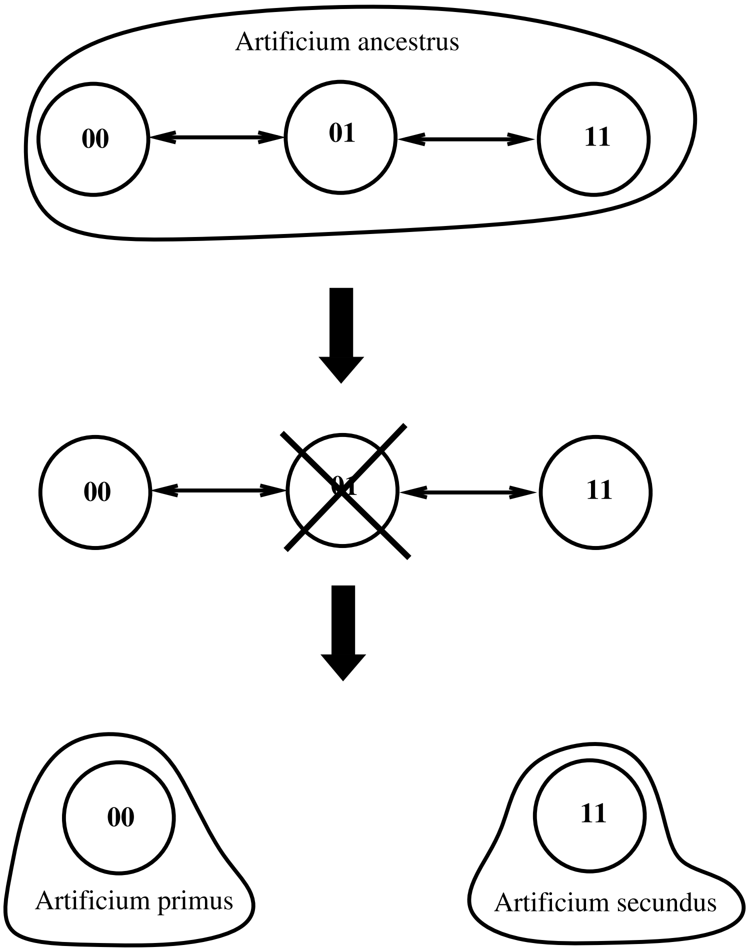

Figure 1: Speciation under the genomic difference implementation of reproductive barriers. Organisms are only allowed to mate with other organisms that have a genome with at most a one bit difference. Potential matings are represented by the double arrows between organisms. A species splits in two when an organism dies and thereby prevents two subpopulations from interbreeding. Time is shown as flowing down through the figure. The species *Artificium ancestrus* divides via a death into the two species *Artificium primus* and *Artificium secundus*.

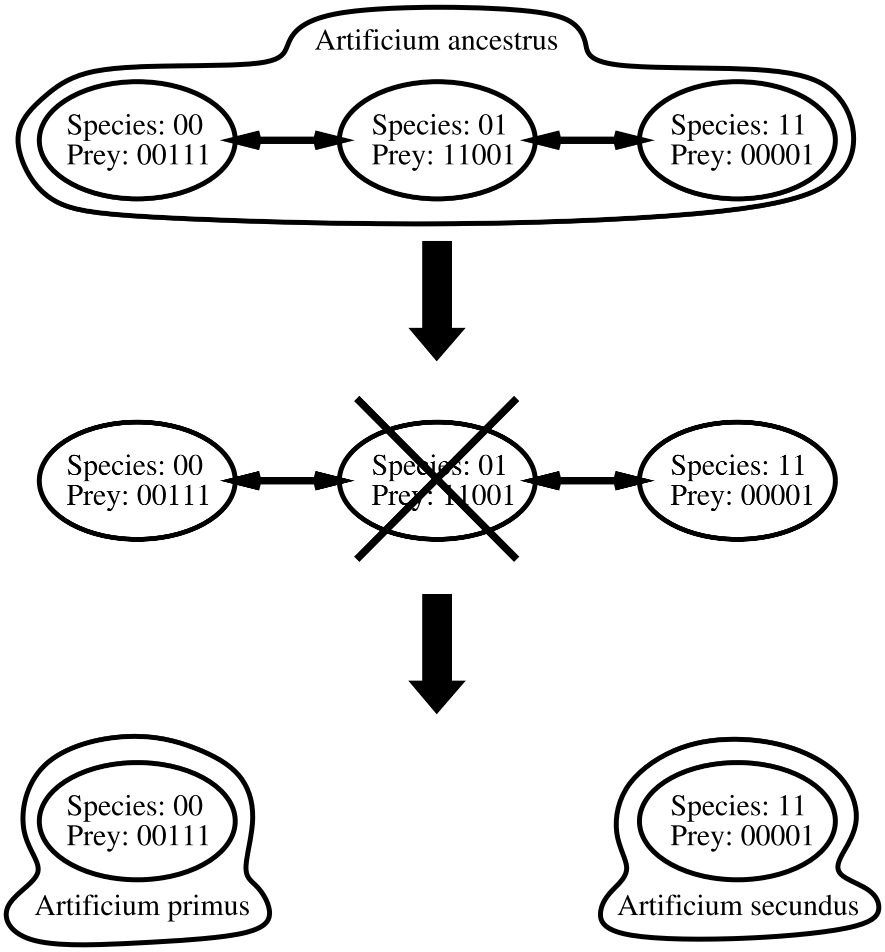

Figure 2: Speciation under the species identity gene implementation of reproductive barriers. Similar to the genomic difference implementation, organisms can only mate with organisms that differ in no more than one bit of the species identity gene. However, there are no restrictions imposed on the variation amongst the other genes of the mates.

means that there is energy loss as we go up the food web. This puts a bound on the number of non-plant organisms. Let this total number of organisms be called K, the carrying capacity of the environment. Since every organism must find a mate in order to reproduce, there is a maximum of K/2 reproducing species in the model. In actuality, the number of species will be lower since the death of one member of one of these two-organism species would spell the end of that species as a propagating entity. Given that organisms would be eating each other at a ferocious rate in this scenario, a lot of species would be reduced to a single member before it was able to reproduce. In other words, we know that locating a mate defines one diversity dependent factor. In actuality it is a density dependent factor, but given a finite K, diversity and population densities are not independent. A question remains, are there other diversity dependent factors that limit species diversity?

4.3 Raising the Ceiling

Answering the question of what factors limit species diversity is tantamount to asking what factors might raise or lower the diversity equilibrium level. Assuming that the factors are dynamic aspects of the environments and do not act as a fixed ceiling on diversity. The obvious candidates are topology and climate[13, 6]. Furthermore, the effects of changes in gene flow can be investigated through changes in the migratory random walk distance.

Benton[1] argues for subdivision of niches as well as an increase in overall adaptive space as the major factors driving diversification. The effects of the subdivision of niche space will be investigated by comparing the condition where all the bits of the *generalism* gene may mutate against the condition where a subset of the bits are fixed as 1's, thereby forcing a minimal level of generalism on all the organisms. The effect of increasing adaptive space can by similarly investigated. A subset of the bits in all the genes will be fixed as 0's as an initial condition. We might then examine the effects of allowing organisms to begin to flip bits in this region. In effect, the mutations in the previously fixed zone would act like evolutionary innovations, exploring new adaptive space.

Vermeij[14] and Van Valen[12] have each suggested that an evolutionary arms race of specialization might ensue between coevolving species. This requires a deeper analysis of the effects of specialization. Since the model includes both food web and phylogenetic information, it should be possible to determine if specialization in one lineage tends to stimulate specialization in species with direct links to the speciating lineage in the food web.

5 Validation

A model is both an simplification and an abstraction from reality. Thus, every model suffers from the problem of validation. Have the essentials of the real system been captured by the model? Does it have any predictive power? Given the time and spatial scale of an evolutionary and ecological model, we are unlikely to be able to test the predictive power of the model. However, the behavior of the model can be readily matched to other theoretical models that have been validated by field and laboratory experiment. The theoretical predictions, which are based on and generally confirmed by observation and experiment focus primarily on the ecological level. This is in part due to the difficulty in observing evolutionary effects.

What follows is a discussion of a set of theoretical predictions that should be testable in the model. I fully anticipate that the model may not fit all of the following predictions. In this case, we must ask why the model is not conforming to our understanding of reality. The process of answering that question should give us insights into unstated constraints in the theories. Thus, the necessary adjustments to the model can be usefully discussed in the dissertation.

5.1 Island Biogeography

Island Biogeography is the study of the interaction of biology with the geographical constraints of island clusters. Many of the results from this field have been generalized to habitat islands. This works particularly well for high altitude habitats where mountain tops effectively form islands for the species specialized to that climate.

Gleason was an early ecologist famous for arguing that similar and even near by environments contain different assemblages of species in differing abundances[8, 9, as referenced in [3]]. We should thus expect to see different food webs, or at least different abundances of species in environment locations with similar climates.

In 1967, MacArthur and Wilson proposed their "equilibrium theory of island biogeography."[10] They observed that diversity levels remain relatively constant on islands and proposed that this was due to the interaction between decreasing colonization rates and increasing extinction rates as the number of species on an island increases. When an island is relatively empty, colonists have an easy time establishing a foothold. Furthermore, without much competition from other species, even a feeble competitor can make a living. However, as diversity increases, life gets harder. There are fewer unexploited resources for newcomers to use and competition becomes fierce. At some point, these forces balance out. The specific equilibrium level of species (S) depends on the area of the island (A) as well as other unexplained, or residual factors. The relationship follows a power law.

$$S = cA^z \tag{2}$$

The exponent, z generally ranges from 0.15 to 0.35. For example, birds in the New Guinea islands have $z = 0.22$ while ants in Melanesia have $z = 0.30$. The constant c is scaled to the number of species on an island of area $A = 1$. Both area and species diversity are easily observed and manipulated in the model. It should be trivial to check the fit of the model to the "species-area curve."

Additional observations from island biogeography include the fact that islands distant from the mainland, or source of colonizers, tend to have fewer species than islands near to the mainland. This could be readily tested in the model.

Finally, islands are biologically interesting in part because they seem to be hotbeds of evolution. The geographical isolation inherent in an island chain is the perfect setting for allopatric speciation. If a species invades an island only rarely, the next invasion may find that the previous colonizers of that species have drifted so far as to have become a new species. Thus, multiple invasions of the same species can lead to multiple species.

5.2 Predation

A basic result of a simple Lotka-Volterra predator-prey model of two difference equations is that the two species will oscillate.

$$\begin{aligned} \text{Predator}_{t+1} &= d\text{Predator}_t + c(\text{Prey}_t)(\text{Predator}_t) \quad (3) \\ \text{Prey}_{t+1} &= b\text{Prey}_t - a(\text{Prey}_t)(\text{Predator}_t) \quad (4) \end{aligned}$$

Here, $0 < d < 1$, modeling the death rate in the predator population. Similarly, $b > 1$ models the intrinsic rate of increase in the prey population. The constants c and a model the frequency and effects of encounters between predators and prey on those respective populations. The oscillations are in part due to the time lag introduced by a difference equation, as compared to a differential equation. However, when we are modeling the effects of prey consumption on the predator species, a time lag seems reasonable. An ecological experiment will be performed with one prey species and one predator species in the model. Their population dynamics will be checked against the analytical model.

Any factor that reduces interspecific competition can have an effect on diversity. This applies both to predation of competing prey species as well as disturbance in the environment. Theory and evidence both suggest that intermediate levels of predation or disturbance will foster higher levels of diversity than either extreme of predation or disturbance. A predator acting in this way, to reduce competition and thereby allow a higher diversity of prey species is called a "keystone predator." The effects on diversity of the introduction of a predator will be checked in the model. Similarly, varying degrees of disturbance may be imposed on the simulated organisms and correlated to the resulting species diversity.

5.3 Evolution

Finally, some general predictions from evolution should apply to the model. Species that have a better match to their environment should drive the less fit species to extinction all other things being equal. Unexploited resources in the environment should eventually be utilized by new species. And, as discussed above in relation to island chains, geographically isolated populations should facilitate speciation events. All of these predictions may be checked in the model.

5.4 Data

The most powerful, and unfortunately, most difficult form of validation would proceed from a demonstration of the model's predictive power. Ideally, we would like to find a data set describing speciation in a food web and the relevant environmental conditions for those events. Then we might model the initial conditions of the data and check if the model provides an accurate simulation of historical reality. Unfortunately, most paleontological studies focus on a single clade, and so fail to account for the dynamics of a food web. Furthermore, it is often difficult to determine the environmental conditions at the time of fossilization. In fact, it is difficult to gather data on which species were extant at any one time. It remains to be seen what data we may find for comparison.

6 Conclusions

Conclusions should emerge from the model in at least three distinct aspects. First, the process of validation will show that I have created a reasonable tool for investigating evolutionary and ecological dynamics. Second, the process of developing the model will illuminate some of the hidden constraints and assumptions in our understanding of the dynamics of ecosystems and evolution. Third, and most importantly, the model will provide evidence as to the answers of our three basic questions. What makes speciation more likely than extinction? Are there diversity dependent factors that modify speciation and extinction rates? How might these factors have changed the diversity equilibrium levels over time? The last question includes a set of questions about the relationships between these factors. Are some diversity dependent factors more important than others? How do they interact? And finally, we will evaluate whether or not they are reasonable explanations for the observed diversification of life over evolutionary history.

References

[1] M. J. Benton. The causes of the diversification of life. In P. D. Taylor and G. P. Larwood, editors, *Major Evolutionary Radiations*, pages 409–430. Clarendon Press, Oxford, UK, 1990.

[2] M. J. Benton. Diversification and extinction in the history of life. *Science*, 268:52–58, 1995.

[3] J. H. Brown. *Macroecology*. Chicago University Press, Chicago, IL, 1995.

[4] H. Caswell. *Matrix Population Models: Construction, Analysis, and Interpretation*. Sinauer Associates, Sunderland, MA, 1989.

[5] H. Caswell and A. M. John. From the individual to the population in demographic models. In D. DeAngelis and L. Gross, editors, *Individual-Based Models and Approaches in Ecology*, pages 36–61, New York, 1992. Chapman and Hill.

[6] J. Cracraft. Biological diversification and its causes. *Annals of the Missouri Botanical Garden*, 72:794–822, 1985.

[7] C. R. Darwin. *On the Origin of Species*. John Murray, London, 1859.

[8] H. A. Gleason. The structure and development of the plant association. *Bulletin for the Turrey Botanical Club*, 44:463–481, 1917.

[9] H. A. Gleason. The individualistic concept of plant association. *Bulletin for the Turrey Botanical Club*, 53:7–26, 1926.

[10] R. H. MacArthur and E. O. Wilson. *The Theory of Island Biogeography*. Princeton University Press, Princeton, NJ, 1967.

[11] J. N. Thompson. *The Coevolutionary Process.* University of Chicago Press, Chicago, IL, 1994.

[12] L. M. Van Valen. A new evolutionary law. *Evolutionary Theory*, 1:1–30, 1973.

[13] J. W. Valentine. Determinants of diversity in higher taxonomic categories. *Paleobiology*, 6:444–450, 1980.

[14] G. J. Vermeij. *Evolution and Escalation.* Princeton University Press, Princeton, NJ, 1987.