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.

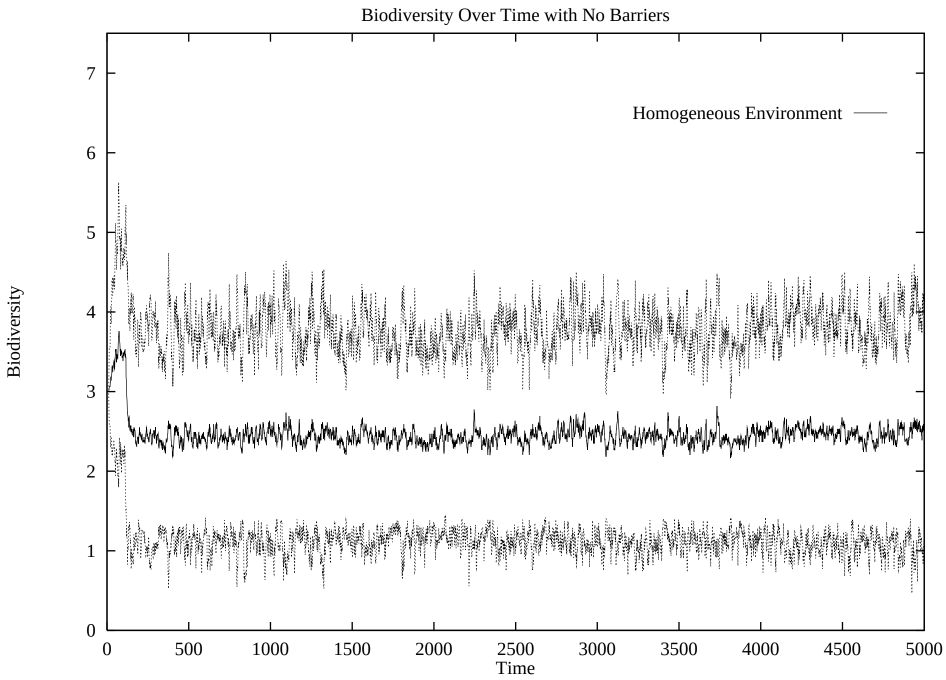


Figure 4-1: The average biodiversity (middle curve) for 50 runs of the model with no barriers under a homogeneous environment. The 95% confidence interval has also been plotted, surrounding the mean. The model was seeded with three species, a plant, an herbivore, and a carnivore, but is not able to sustain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

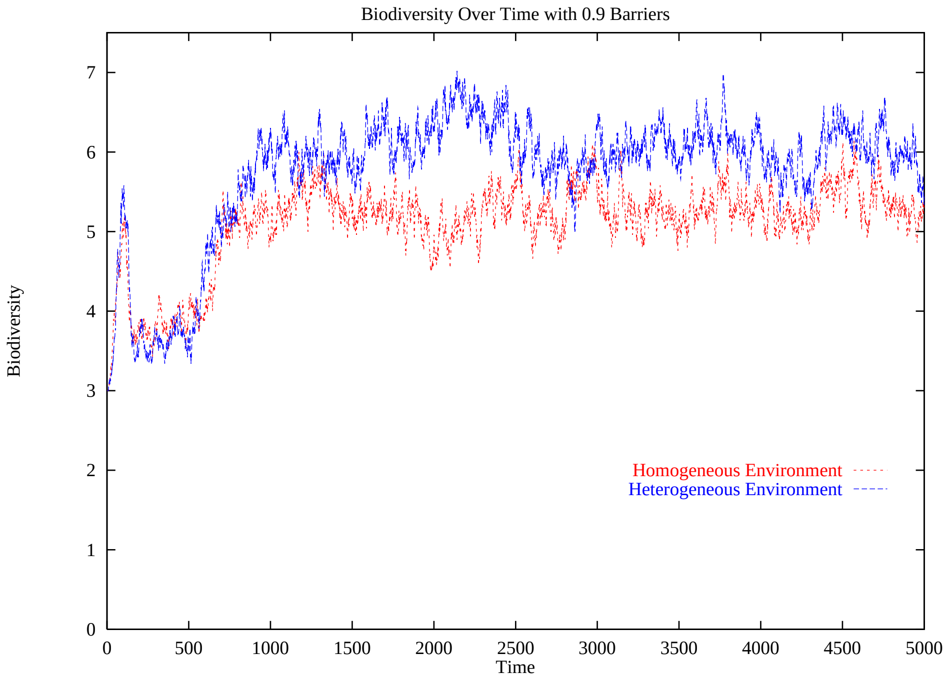


Figure 4-2: The average biodiversity for 50 runs of the model with 0.9 barriers under both a homogeneous and a heterogeneous environment. In the homogeneous environment case, there was no variation between patches in their habitats. The habitats in a heterogeneous environment differed in an average of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

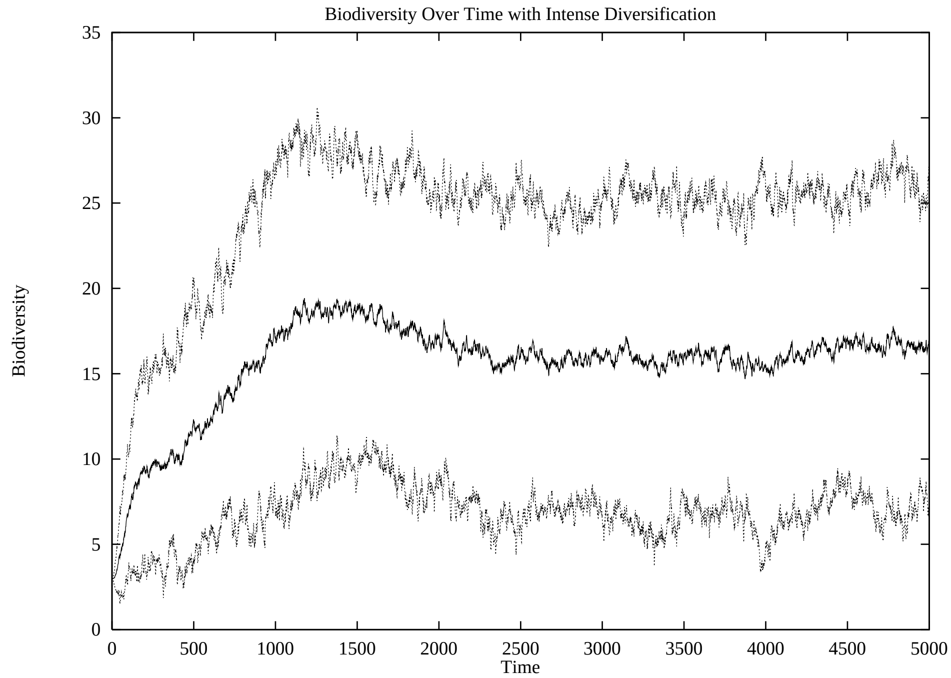


Figure 4-3: The average biodiversity (middle curve) for 50 runs of the model with parameters set so as to maximize animal species diversity. This means a heterogeneous environment, 0.95 migratory barriers, and positive assortative mating based on predation resistance chromosomes. The 95% confidence interval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

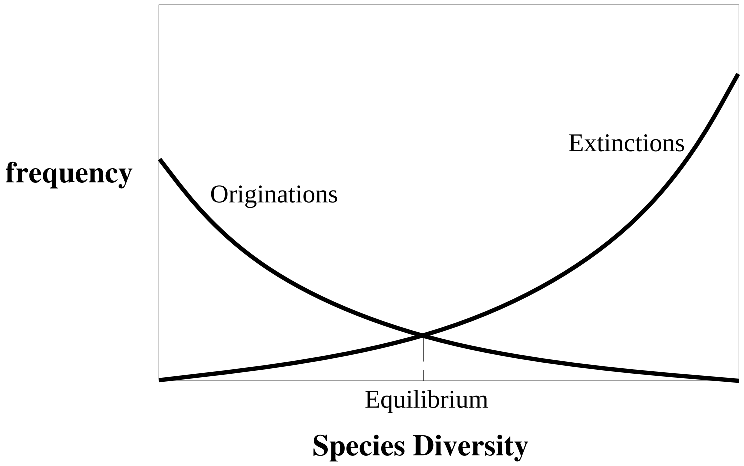


Figure 4-4: Species diversity equilibrium in island biogeography. The probability of a new species colonizing an island drops off with diversity because there is a higher chance that the species already exists on the island and competition makes it more difficult to establish a foothold. Similarly, extinctions 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.

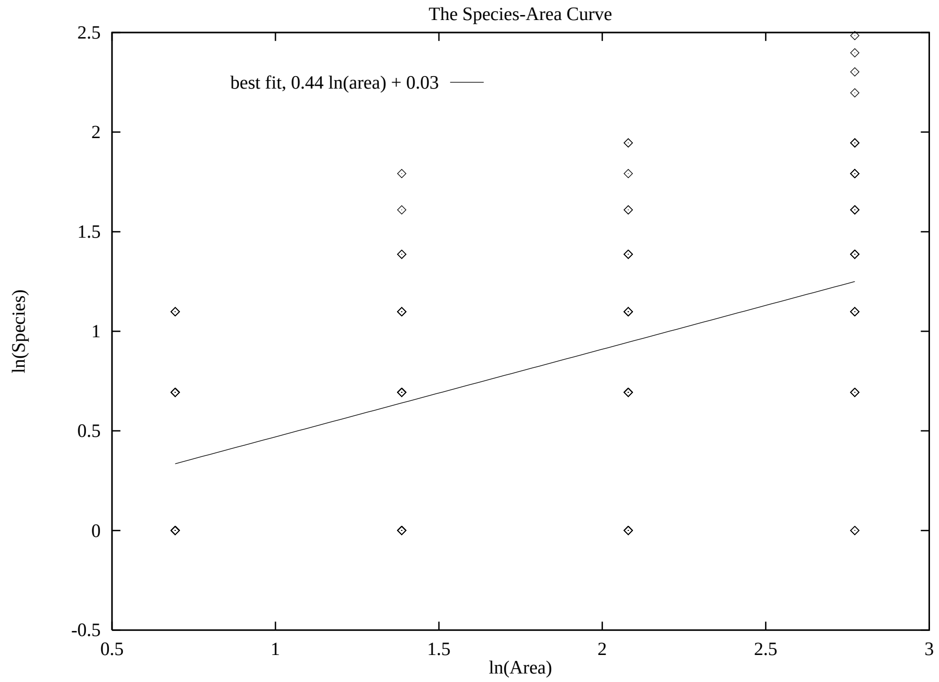


Figure 4-5: The species-area data. The number of patches (area) and the number of species have been scaled by their natural log. Most of the 50 data points for each area value overlap. The line of best fit is $0.44ln(\text{area})+0.03$. The coefficient, $z = 0.44$ is statistically significant ($p < 0.001$), 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.

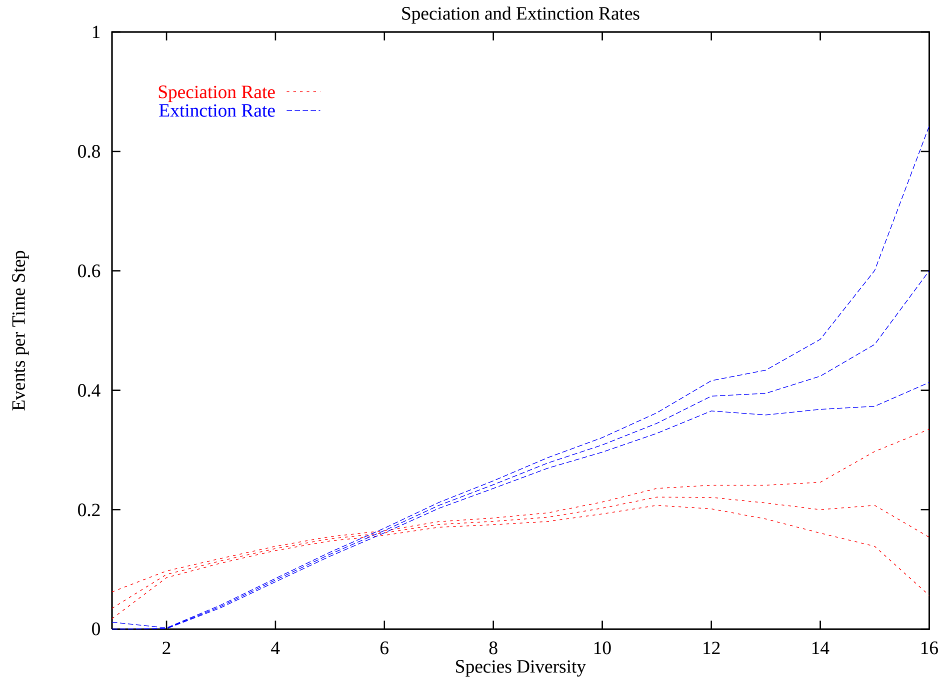


Figure 4-6: Speciation and extinction rates in the model as functions of diversity. This data comes from 50 runs the model with 0.9 migratory barriers in a homogeneous environment. The means for both rates are surrounded by curves representing the 95% confidence intervals for the means. In contrast to 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-

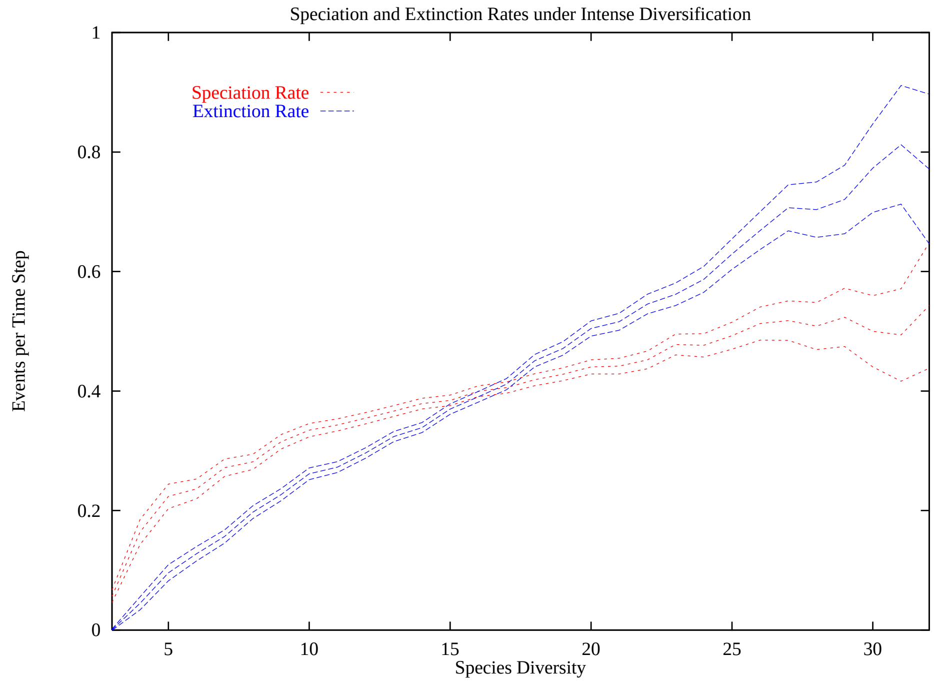


Figure 4-7: Speciation and extinction rates in the model with parameters set so as to maximize animal species diversity. The means for both rates are surrounded by curves representing the 95% confidence intervals for the means. The mean of 16.76 species observed over 50 runs, at the end of 5000 time 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:

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

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 ($a > c$).

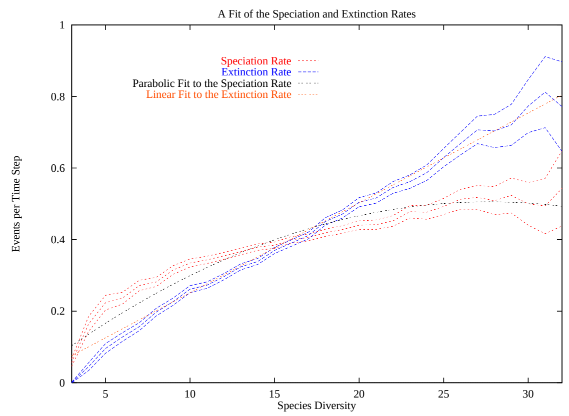


Figure 4-8: A parabola has been fit to the speciation rate curve and a line has been fit to the extinction rate curve from Figure 4-7. They intersect at an equilibrium of 17.27 species, not far from the observed average of 16.76 species observed at the end of 5000 time steps.

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.

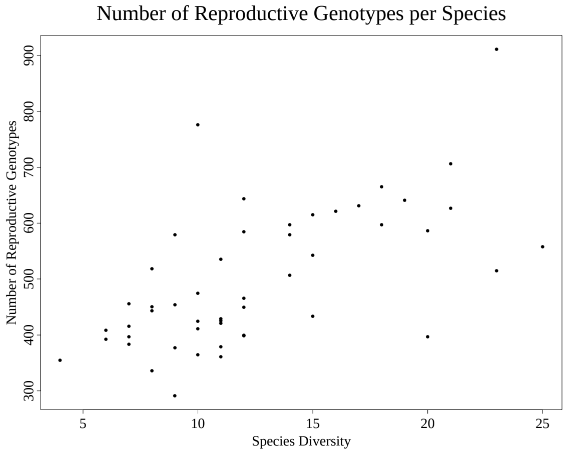


Figure 4-9: The number of reproductive genotypes per species under a homogeneous environment with 0.99 migratory barriers for 50 runs of the model. This parameter combination was chosen because it produced a wide range of diversity values.

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

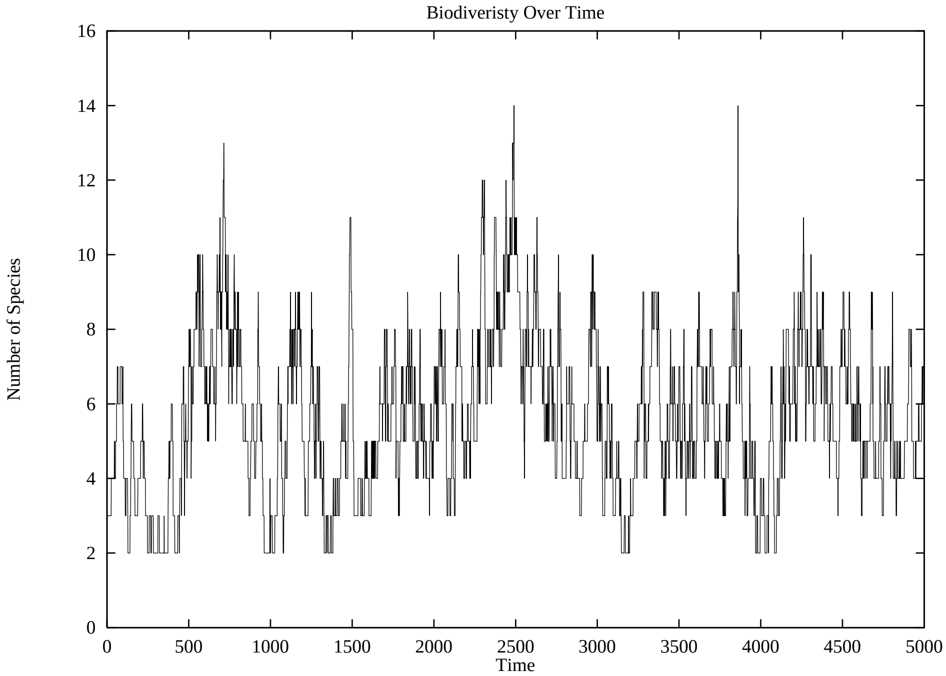


Figure 4-10: The number of species over time from a single run. This is one of the 50 curves that was averaged into Figure 4-2, with 0.9 barriers and environmental homogeneity. Notice that any horizontal lines are very short, indicating that the model did not sustain a particular diversity level for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

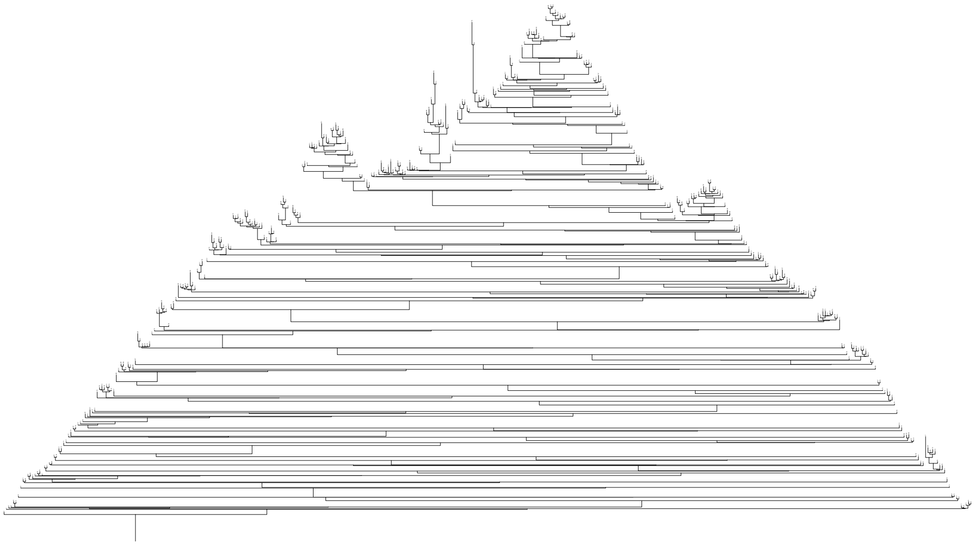


Figure 4-11: A plant phylogeny from a run with 0.9 barriers to migration and a homogeneous environment. Time flows up through the figure. Starting with one plant species at the root, the phylogeny shows many splitting events (the horizontal scale has no meaning), but only rarely do both new species survive 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).

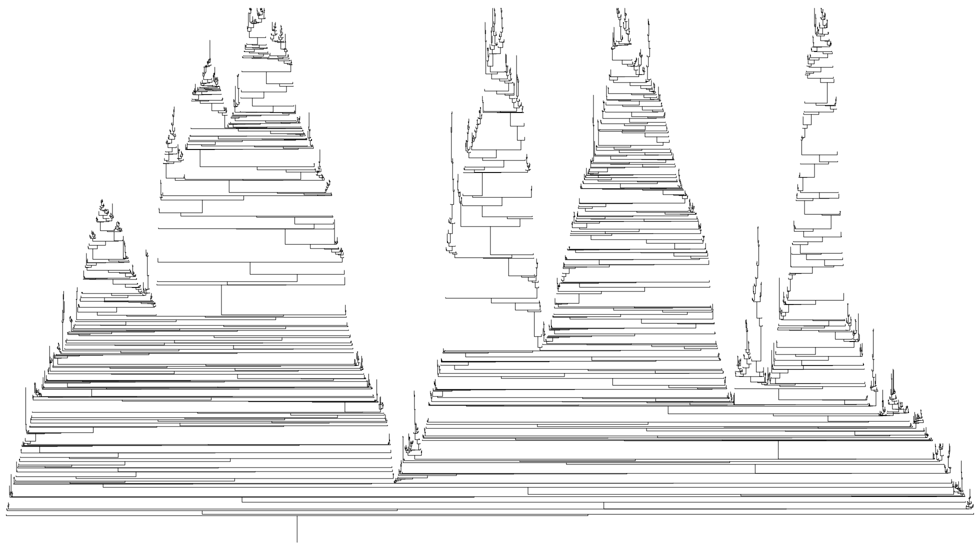


Figure 4-12: A plant phylogeny with maximally diversifying settings for the model. This run had 0.99 barriers to migration, a heterogeneous environment, and positive assortative mating (the tendency to mate with similar organisms) based on predation resistance characters. The Christmas tree shape seen 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

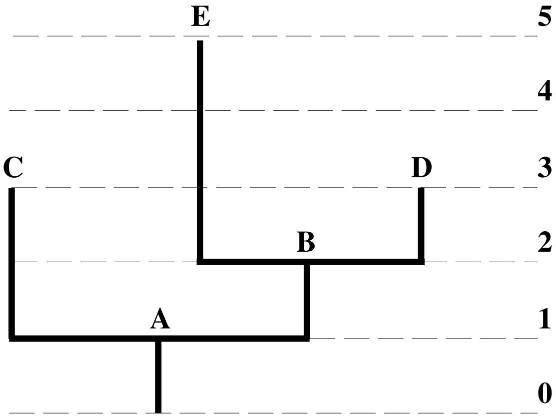
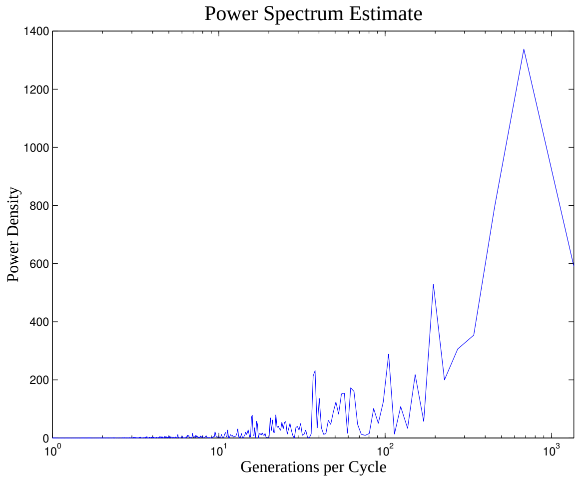


Figure 4-13: The problem of species durations. Is species E 5 time steps old or 3? Did species C and D survive for the same amount of time or did species C live for twice as long as D? To trace a species' origin all the way back to the root of a tree implies that as time progresses the mean species life 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.



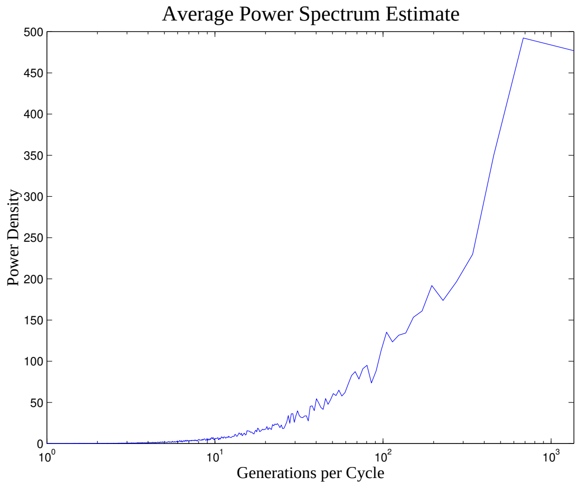


Figure 4-14: An individual (left graph) and the average (right graph) power spectra for diversity over time. The horizontal axis has been scaled logarithmically. The data comes from 50 runs of the model with 0.9 barriers and a homogeneous habitat. The left graph represents the power spectra for the run 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.

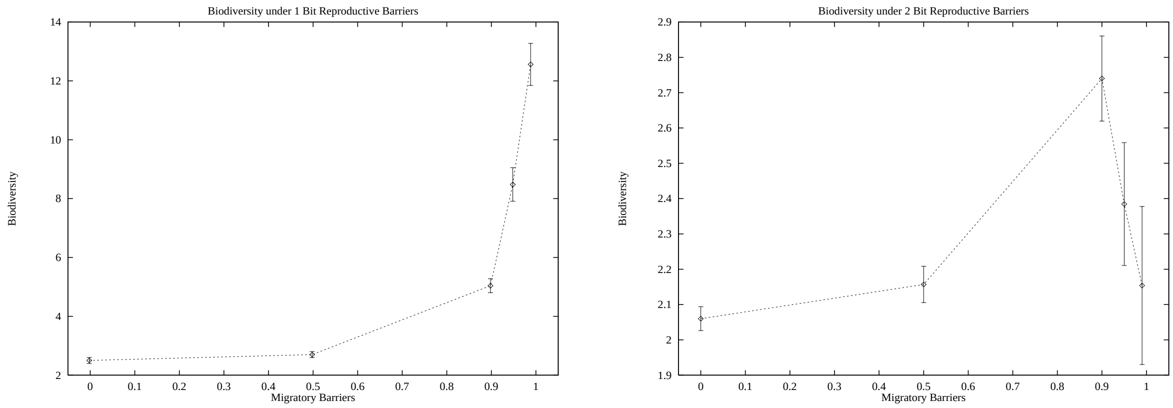


Figure 4-15: Diversity under 2 bit reproductive barriers. The left graph shows the effect of migratory barriers on diversity (with standard error bars) when organisms with at most 1 bit different in their reproductive chromosomes are allowed to mate. The right graph shows the same effect when mating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

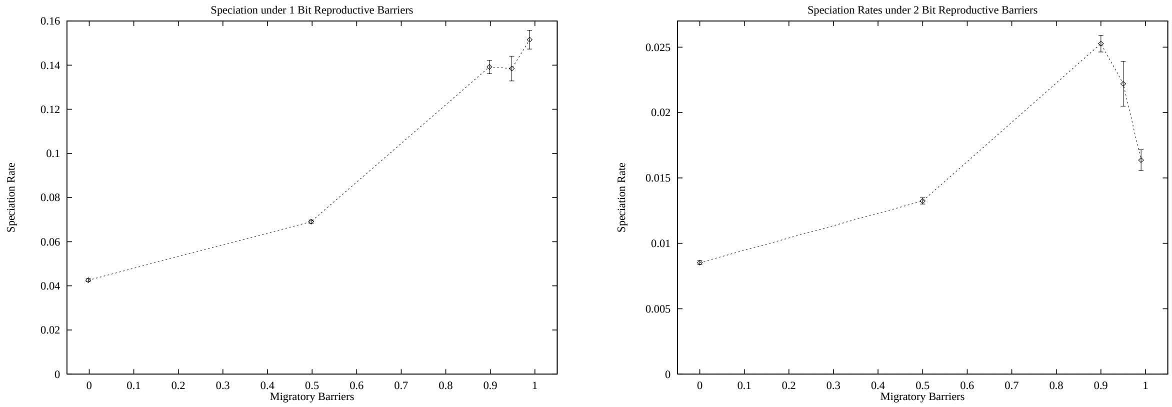


Figure 4-16: Speciation under 2 bit reproductive barriers. The left graph shows the effect of migratory barriers on speciation rates (with standard error bars) when organisms with at most 1 bit different in their reproductive chromosomes are allowed to mate. The right graph shows the same effect when 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?