Chapter 1

Introduction

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

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

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

1.1 The Problem of Species Diversity

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

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

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

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

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

1.2 What is a Species?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1.3 Three Levels of Answers to the Problem of Diversification

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

1.3.1 The Basic Processes of Diversification

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

Speciation

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

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

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

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

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

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

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

Extinction

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

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

1.3.2 Regulation of Diversity

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

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

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

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

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

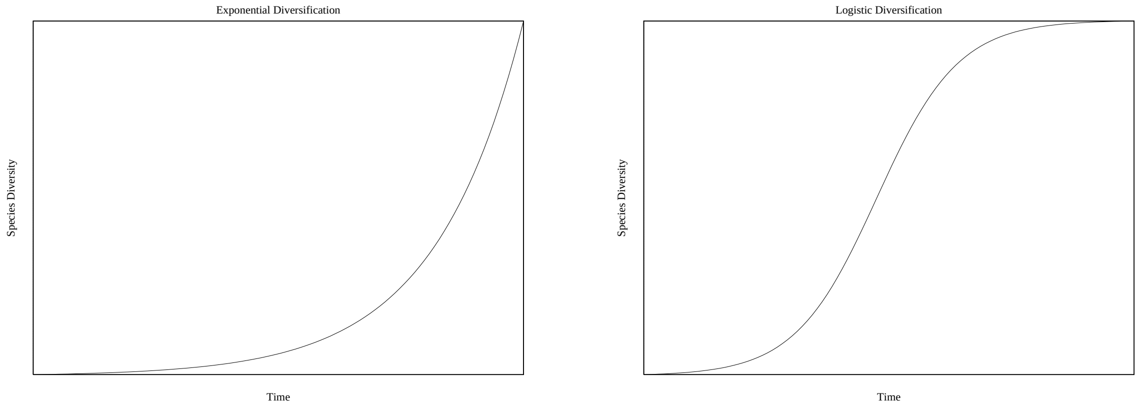

Figure 1-1: An exponential (left) and a logistic curve (right). Both have been proposed as a description of the pattern of species diversity over time. The answer to which is a more accurate description depends on whether life on earth has neared some form of maximum diversity as depicted by the logistic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

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

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

1.3.3 Factors Influencing Diversity Equilibria

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

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

Diversification as an Artifact

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

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

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

The Inevitability of Diversification

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

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

Diversification due to Physical Factors

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

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

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

Diversification due to Biological Factors

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

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

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

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

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

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

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

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

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

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

In Summary

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

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

1.4 The Role of Computational Models

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

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

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

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

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

1.4.1 Configuration Models

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

1.4.2 Methodology

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

1. Isolate a question.

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

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

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

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

1.4.3 Caveats

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

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

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

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

1.5 Previous Work

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

1.5.1 Microevolutionary Models

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

---

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

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

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

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

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

1.5.2 Macroevolutionary Models

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

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

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

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

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

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

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

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

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

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

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

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

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

1.6 Harbingers

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

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

And now I shall get on with it.