Introduction

from *Evolutionary Dynamics of Interacting Ecosystems*

Aditya Mahadevan

August 29, 2025

1 Biophysics

The field of biology is in the midst of a period of transformation akin to the seismic shift experienced by physics a century ago. New experimental techniques—from advances in microscopy to genetic sequencing—allow us to describe biological processes in unprecedented levels of detail, and give us a window into systems that span scales from populations to single molecules. Many paradigm-shifting discoveries lie in the future if we are to understand the huge range of biological phenomena in nature: from cellular physiology to organismal development to cognition and collective behavior.

The study of biophysics—of which this thesis is a part—is an endeavor to approach the rich mysteries of biology with the quantitative reasoning which has been so successful in physics. Although there are—and will remain—physical phenomena that lack a satisfying theoretical explanation, the study of physics over the last few centuries has given humanity a large degree of both predictive power and control over the world around us, spanning scales from individual atoms to the emergent phases of matter that they constitute. Condensed matter and statistical physics have allowed us to greatly reduce the dimensionality of our descriptions of reality, identifying the small number of collective variables that describe large ensembles. For example, if we want to know the volume occupied by a particular number of molecules of gas, we need not perform a microscopic calculation involving the individual positions and velocities of each molecule: our phenomenological understanding of the system tells us that, for a large range of conditions, it suffices to know just two measurable quantities: the *temperature* and *pressure*. An understanding of such emergent descriptions—and their independence from microscopic details—has been one of the most far-reaching developments in twentieth-century physics [1].

However, a similar understanding of effective variables and emergent simplified descriptions in biological systems is lacking. As a result, even though we have a huge amount of control over certain model organisms—well-studied systems which have been invaluable to advancing the field—we generally lack an ability to engineer desired functions *de novo* or predict the response of a biological system to a novel perturbation. To compare the situation with that of physics, by now our understanding of solid and fluid mechanics is good enough to allow us—with the aid of computers—to test and modify aircraft designs before they are even built. Although it seems possible that we will reach a similar level of competence in, say, designing and iterating over protein structures that catalyze desired reactions, at present there is a gap in our understanding of and control over biology that must be closed before this scenario becomes a reality. If our experience in the physical sciences is generalizable to biology, it suggests that increased understanding of biological systems will come not only from more and better experiments, but from a theoretical synthesis that allows us to understand diverse phenomena under the umbrella of a relatively small number of widely-applicable principles.

The need for discovery and application of novel theoretical principles in biology is due to characteristics of biological systems that make them very different from more familiar physical systems. Because of the conditioning of living systems on a long history of evolution and the concomitant amplification of small differences through natural selection, behavior that we might think of as atypical becomes typical, and living systems can implement highly nontrivial mechanisms which *a priori* appear fine-tuned, and whose understanding requires stepping out of conventional physics paradigms. Study of such biological systems that exhibit qualitatively new phenomena has been fruitful for physics as well as for biology. For example, efforts to understand the physical laws governing learning and memory—essential biological processes—have spun off into hugely influential developments in neural networks and artificial intelligence, pointing toward emergent principles that lie beyond the purview of conservation laws and extremization principles which govern much of conventional physics.

There is a branch of biophysics that is concerned with the physical properties of biological systems:

for example the rheological properties of viral particles or the elasticity of single molecules. This area of study is essential for a detailed description of living systems, just as particle physics gives us the most detailed description we have for our physical world. However, biophysics of the kind addressed in this thesis engages with the challenge of understanding the function and self-organization that emerge from *many interacting agents*, whether they are cells, proteins, genes or organisms. In this way the approach herein is one of a statistical physicist, seeking a coarse-grained description that depends on microscopic details only through a few effective parameters, just as the physical equations for fluid flow describe equally well the dynamics of water, oil and honey—albeit with different values of the density and viscosity. The field of such "high-dimensional" biology is rife with challenges. For example, how do neurons talk to each other in order to sense and perform computations on their environment? How do cells divide and grow and coordinate with one another to robustly self-organize into an embryo from a single fertilized egg? Indeed, underlying all functional units—from gene regulatory networks to neural architectures—there is a question of how evolution gives rise to such robust function. In this thesis we will focus on the questions: how do evolutionary forces shape biodiversity in the natural world, and how does extensive biodiversity evolve and coexist?

2 Role of theory

Biology is ultimately rooted in experiment and observation. What, then, is the role of the kind of theory pursued in this thesis? Here we suggest that there are three main ways in which theory contributes to biology. (1) Theory allows us to identify the *right variables* in order to describe a system, often in a much lower-dimensional space than might be naively assumed. Just as temperature and pressure are useful descriptors of large ensembles of particles, are there analogous quantities in classes of biological systems? (2) Theoretical investigation of *simple models* which incorporate essential aspects of a biological system allows us to calibrate our expectations and thereby be honest about we do and do not understand in empirical data. For example, Hopfield's violated expectation for the error rate of DNA replication indicated the presence of active error correction that birthed an understanding of kinetic proofreading [2]. (3) At its best, theory works in tandem with experiment, guiding where and what kind of new data are needed. This synergy has a long and robust history in physics, and is increasingly possible in biophysics as well.

Physics has historically been effective at isolating the key elements required to explain a phenomenon through the use of *toy models* which are highly simplified and sometimes even stylized models that nonetheless, at their best, capture something essential about a phenomenon. The style of the theory in this thesis—here applied to the evolutionary dynamics of microbial communities—is in this spirit: not seeking to connect directly with a particular experiment but rather more generally exploring the behavior of simple models as a means for calibrating our expectations and exploring the consequences of certain assumptions.

It seems timely to comment on the relevance of artificial intelligence to the goal of reducing the complexity of the description of biological systems. Increasingly sophisticated machine learning models have proven to be capable of distilling large amounts of data into apparently simple descriptions. Indeed, one can already answer many questions faster with artificial intelligence than with natural stupidity. All of science will certainly be touched by the advent of artificial intelligence. However, at least in the present, one of the contributions that human scientists can offer is a focus on the *questions* whose pursuit is likely to drive the field forward. It is perhaps likely that finding answers to such questions might be more effectively done with artificial intelligence than with our own brains, but for now there the niche for *asking insightful questions* has not yet been filled. As such, in this thesis we aim to emphasize a number of the broad open questions in the field of microbial ecology and evolution, with a bias towards those that require increased theoretical understanding of simple toy models.

3 Microbial evolution and ecology

One of the most rapid advancements in biology has been in DNA sequencing. The cost of sequencing has been decreasing super-exponentially over the last two decades. As a result, huge amounts of sequencing data are being produced daily around the world. Among many other insights, DNA sequencing has revealed to us the immense diversity and prevalence of microbial communities in nature. Across environments—from the soil to the oceans and lakes to our guts and skin and even the air we breathe—microbes (including, bacteria, fungi, viruses, and protists) are found almost everywhere [3, 4, 5, 6]. Furthermore, they always occur in *communities* comprised of multiple species, strains, and substrains. In particular, the diversity of the microbial world exhibits structure on all scales: from family-level differentiation all the way down to—now resolvable—differences in genetic content between individual cells. There are a number of striking examples of the scales scanned by this "fine-scale" diversity, which appears to be generic in microbial communities, from natural *Escherichia coli* populations to *Synechococcus* in the hotsprings of Yellowstone, to communities of the ocean-dwelling *Prochlorococcus marinus*, to marine viruses that prey upon bacteria [7, 8, 9, 10].

Understanding the structure of microbial communities is of vast importance—not only for its own sake, but also for the crucial role that microbes play in the health of our planet and of ourselves. Global nutrient cycling is driven largely by microbes, with major fluxes in carbon and nitrogen driven by microbial metabolism. Therefore understanding the response of the microbial ecosystems that perform these essential processes to environmental perturbations will enhance an understanding of the stability of the biosphere and therefore of the planet to the human-induced changes that are underway [11, 12]. In addition, as the COVID pandemic has demonstrated, the continual evolution of microbes is tightly linked with the health of our own species. Indeed, the interplay between microbial (in this case viral) evolution and our immune systems is responsible not only for acute events such as the emergence of SARS-CoV-2 but also for the perennial need for flu vaccinations [13, 14]. At the same time, bacteria are hugely important to humans: the human gut microbiome is known to influence our health, and we are just beginning to understand the way in which its changes and evolves over the course of our lifetime [15, 16].

There are a number of tools at our disposal for studying both ecological interactions and evolution within microbial communities. Since microbes are small and reproduce quickly, their evolution can be studied in the laboratory, where hundreds of generations—on the order of weeks for microbes growing under serial dilution—is already enough to see new mutations emerging and rising to high abundance [17]. Longer-term evolution experiments are also possible, and the longest such experiment, started by Richard Lenski, has been studying the evolutionary dynamics of *E. coli* in glucose limitation for now more than 30 years, which amounts to more than 75000 generations [18]. Large populations (a single mL of saturated lab media can host $\sim 10^7$ cells) mean that mutations are occurring all the time: to wit, a single person's gut is home to $\sim 10^{13}$ microbial cells which produce roughly 10^9 single mutations every day as they reproduce [16]. These large population sizes, in addition to being convenient for experimental studies, increase theoretical tractability. In the limit of large populations, intuition from the law of large numbers suggests that random fluctuations become less important and physiological details can hopefully be coarse-grained into effective parameters—similar to how viscosity and density emerge from the microscopic details of a fluid.

Empirical data on microbial biodiversity pose many questions that resist easy theoretical answers: (1) Why is there so much coexisting fine-scale diversity? Simple theoretical models predict that the amount of diversity that can stably coexist is proportional to the number of limiting resources in an environment [19], but this bound appears to be utterly violated by any and all natural microbial communities [20]. Closely related microbial strains compete more directly for the same resources, and therefore, at least in naive models, are expected to exclude one another more often than they coexist, but data show a wide spectrum of coexisting yet closely-related strains [8]. (2) How does biodiversity evolve? Though we know that ultimately evolution *must* be responsible for the emergence of the structure we see, theory is needed to understand whether expectations from simple models are consistent with this structure. (3) How does the mode of ecological dynamics affect the resulting evolutionary

dynamics? For example, does a community with predominantly predator-prey interactions evolve differently from a community governed by resource competition? (4) Why does evolution—measured through the continual emergence of successful mutations—apparently proceed for very long times without converging on a fitness maximum and slowing down? This phenomenon has been observed in laboratory experiments [21] as well as in nature [22], and has inspired the "Red Queen" hypothesis in biology [23]: in this work we endeavor to make this hypothesis of continual eco-evolutionary dynamics quantitative, with specific models and questions about the generality of the behavior that we observe.

The main goal of this thesis is to explore various eco-evolutionary scenarios—in models as simple as seem reasonable—in order to develop our expectations for what happens when new variants are continually emerging in a complex ecologically interacting community. The models we study could therefore be viewed as null models for the process of microbial evolution. Deviations from the predictions of these models is a sign that our description is missing something important. As we shall see, even "simple" null models exhibit remarkably complex behavior that we have only been able to understand partially.

4 Review of models and results

The flavor of models that we study, inspired by similar toy models in statistical physics, derives from two concepts that pervade the style of this approach. The first is an assumption of *high dimensionality*. Given the vast complexity of biological systems, notions of phenotype (the properties of an organism), environment (the external factors that influence the organism), and even fitness (an organism's reproductive success as a function of both phenotype and environment) must be high-dimensional. Therefore intuition from oft-portrayed caricatures of evolution as gradient ascent in a low-dimensional fitness landscape can be misleading. It is useful instead to *start* from the perspective of large phenotypic or environmental dimension, and study these biologically-relevant regimes directly.

The second concept of which we will make extensive use is that of *approximating complexity as randomness*. This approach was introduced into physics by Eugene Wigner to study properties of heavy nuclei [24], and made its most famous appearance in biology through the work of Robert May [25], who used it to draw conclusions about the stability of diverse ecosystems. The idea is that, given the complexity of biological systems, instead of trying to resolve this complexity and parameterize all the various processes inherent in e.g. an interacting ecosystem, we could, if we are interested in coarse-grained *typical* properties, imagine replacing layers upon layers of biological complexity with constant-in-time randomness. In the case of ecosystems, this means replacing detailed molecular mechanisms for the interactions between strains—in the simplest case pairwise interactions—with a random pair of (possibly correlated) numbers for each pair of strains a and b, which quantify the effect of strain a on the growth rate of b and vice versa.

Along these lines, much of the work in this thesis can be framed in the context of a generalized Lotka-Volterra model for the ecological interactions between S different strains, with fractional population abundances $\{\nu_i\}$ obeying dynamical equations

$$\frac{d\nu_i}{dt} = \nu_i \left(s_i + \sum_{j=1}^{S} V_{ij}\nu_j - \Upsilon(t) \right) \tag{1}$$

where $\Upsilon = \sum_i \nu_i(s_i + \sum_j V_{ij}\nu_j)$. Here s_i is the intrinsic differential growth rate of strain i, and the V_{ij} are interaction strengths quantifying the effect of abundance j on the growth rate of strain i. Υ is a Lagrange multiplier that enforces $\sum_i \nu_i = 1$, and is of more technical than conceptual importance: as described in Ref. [26] it emerges naturally in models where the interactions between strains are similar to one another with a large mean part and small variations. Note that the dynamics of Equation 1 are deterministic for a given choice of the V matrix: this is an important simplification arising from the assumption of large population sizes, where birth-death stochasticity produces fluctuations that are neglible compared to the total number of individuals.

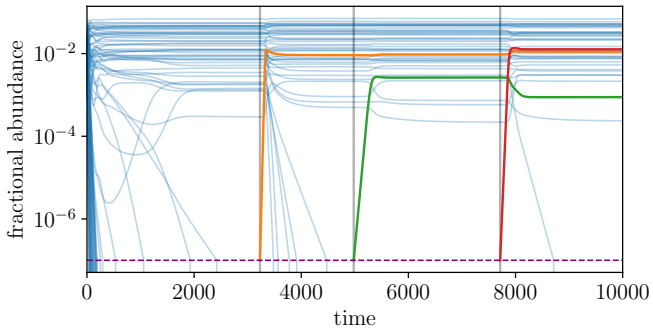

Figure 1: A simulation of the eco-evolutionary dynamics studied in Chapter 2. The ecological dynamics take the system to a stable fixed point, after some initial transient where a number of strains go extinct, here indicated by their abundances going below the dashed purple line. Once at this fixed point, a new strain (in yellow) is introduced, which perturbs the fixed point, shifting around the extant strains' abundances and triggering extinctions. This process continues, with new strains coming in at vertical gray lines and shown in colors other than blue. What is the long-term evolutionary outcome? Does the diversity increase, decrease or saturate? What are the properties of the resulting evolved community?

Previous studies focusing on the *ecological* dynamics of large complex communities have modeled V as a random matrix with independent entries, or with systematic correlations between entries across the diagonal. However, one of the questions we will ask is how the structure of V—and the resulting dynamics of the $\{\nu_i\}$—change when the community *evolves*, either by introduction of new strains or mutation of existing strains. In particular, it may be usefully pointed out that any interesting biological phenomenon is likely *not* simply random. Indeed, it is not all clear a priori what the correct ensemble from which to draw e.g. ecological interactions is, in order to capture features of real systems. To address this point, we have studied models of evolution where the mutational process combined with the ecological interactions yield an *ensemble* of systems as an output, which can then be studied *conditioned on having emerged from evolution*.

When a mutant is created or a new strain is added to the community, the V matrix acquires a new row and column which describe the interactions between the new strain and all the other members of the extant community. The composition of the extant community then determines whether this new strain can successfully invade the community or whether it goes extinct. In the long term, the result is a dynamics where the interactions between and the identities of extant strains change over time as the community evolves. In some ways this process is similar to the dynamics of learning in models of interconnected neurons: here the synaptic weights that govern interactions between neurons can be tuned so that the whole system computes a particular function. In the evolutionary dynamics there is not a clear analogue to the function computed by the neural net: instead the process of natural selection conditions on survival, since strains that go to zero abundance disappear forever from the community.

4.1 Gradual assembly in a well-mixed community interacting via resources

In Chapter 2 of this thesis we study the evolutionary dynamics of a number of related ecological models whose dynamics reach a stable fixed point on ecological timescales. These dynamics are illustrated in Figure 1. After the strain dynamics reach a fixed point, each subsequent addition of a new strain slightly changes the abundances at this fixed point. In the limit we study, strains are added at slow rate, only invading after the ecological dynamics have relaxed to their fixed point. True biological evolution would stipulate that a newly-added strain is closely related to an extant parent strain in the community, with the way it interacts with others tightly correlated with those of its parent. However

in this work we deal mostly with adding strains that are independent from (or only slightly correlated with) the rest of the community: for this reason the process we study is more akin to gradual assembly of a community than *bona fide* evolution.

We mainly study the case of interactions between strains that are mediated by resources: in this case the number of resources, D, sets the dimension of the phenotype of each strain, and defines a maximum rank of the effective interaction matrix between strains via $V = -FG^\intercal$ where F and G are $S \times D$ matrices, and the phenotype of a particular strain is the corresponding row of both F and G. D thereby sets a scale for the maximum diversity in the ecosystem in a manifestation of the competitive exclusion principle [27]. The interactions between strains are asymmetric, which amounts to a difference between the consumption rate of a resource and the growth rate from consuming that resource—in other words $F \neq G$—and is certainly biologically generic. Our main finding is that there is a robust dynamical phase of continual evolution in the eco-evolutionary dynamics of a range of models—a "Red Queen" phase—wherein new strains continue to invade the ecosystem and change the ecological fixed point, but evolution does not slow down indefinitely, instead proceeding at a nonzero rate even in the long-time limit. We also find that there is a qualitatively different dynamical phase in a model where D is effectively infinite: here, as the interactions between strains become more symmetric, there is a critical threshold above which one loses the continually-evolving statistical steady state and enters a regime of ever-slowing evolution, with the probability of invasion of new types consistently decreasing in time, and concomitantly with a few dominant strains rising to anomalously high abundance and turning over more and more slowly. We term this phase the "oligarch" phase but are not able to understand it quantitatively at the level of the Red Queen phase.

Importantly, our results persist in the presence of nonzero $\{s_i\}$ from Equation 1. Even when such unilateral fitness benefits exist, it does not pay to push arbitrarily far into the tail of the s distribution, due to the variance in the interaction-derived force overwhelming the variance in the intrinsic growth rates.

4.2 Continual evolution and diversification around a fitness peak

In Chapter 3 we start with an ecological model with interactions between strains determined through a shared D-dimensional environment, which in the case of Chapter 2 was the set of resources abundances. However now we take seriously the gradualness of evolution, by generating invaders to the ecosystem via small mutations of extant phenotypes. With the differential growth rate of a strain determined by both its phenotype and its environment, if strains do not affect their environment then the dynamics of evolution look like gradient ascent in a static fitness landscape. However, the ability of organisms to modify their own environment—for example by depleting the concentrations of available resources—changes the ensuing dynamics drastically. Now, as a population wanders through the fitness landscape, driven by selection pressures, it modifies the landscape, which in turn affects where it wanders next. Therefore, the idea of a static fitness landscape is replaced by a continually-modified fitness *snowscape* that is reshaped by the continually evolving population [28, 29].

We particularly analyze the eco-evolutionary dynamics in the vicinity of a *stationary point* of the evolution, where the evolutionary "force"—the gradient of the growth rate with respect to phenotype at constant environment—vanishes. We find a large number of possible behaviors around such a stationary point in phenotype space as a function of two parameters: the level of reciprocity between environmental impact on organism and organismal feedback on environment, and the overall strength of the organism's feedback on the environment.

For small environmental feedback the population converges toward a peak in the fitness snowscape, not affecting its environment enough to engender continual wandering but rather following gradient ascent, with the probability of successful mutations decreasing with proximity to the peak. The diversity here can be either extensive in the environmental dimension or $\mathcal{O}(1)$, depending on the value of the reciprocity between organism and environment. For larger feedback, we find two Red Queen steady states of continual evolution, one with $\mathcal{O}(1)$ diversity, and the other with $\mathcal{O}(D)$ coexisting strains. In this diverse Red Queen steady state, the extant phenotypes generically form a "cloud" in

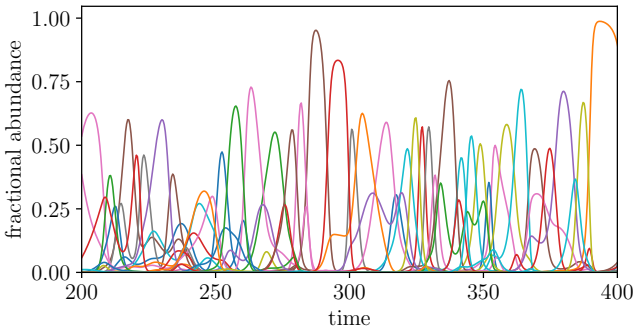

Figure 2: A simulation of the boom-bust dynamics of Equation 1 with negative cross-diagonal correlations in V. Here the strain dynamics are shown on a *single island*: similar desynchronized chaotic dynamics across different coupled islands prevent the range of chaotic fluctuations (on a log scale) from diverging, since local extinctions on a single island can be repopulated by migration events from blooms on other islands. As evident when shown on a linear scale, only a handful of strains are abundant at each moment in time.

phenotype space, where the shape of this cloud depends once again on the properties of the fitness landscape. This cloud of diversity has nontrivial phylogenetic relatedness whose properties we explore numerically. Our model therefore gives a plausible scenario for the evolution of fine-scale diversity, with the cloud of closely-related but coexisting phenotypes somewhat reminiscent of the many closely related types of microbes found in nature. Particularly intriguing is the possibility of diversity evolving across multiple scales, with the emergence of diverged clouds—species—comprised of coexisting closely-related strains, which could perhaps arise for forms of the fitness landscape with multiple stationary points. The model also alllows us to understand conditions for when the stationary point goes unstable to evolution, which can happen due to environmental feedback even when the static landscape has a stable fitness peak. We additionally observe a regime of evolutionary branching, wherein a strain splits into two species which are pushed in different phenotypic directions: this phenomenon has been seen in previous work [30] and suggests a possible mechanism for how a single initially-clonal population could diversify into multiple species, each of which might contain fine-scale strain structure and sit at a different stationary point of the fitness snowscape.

While the model described above can be formulated in the framework of Equation 1 with an interaction matrix V endowed with structure from organismal phenotypes, and the general fitness s_i a function of the phenotype of strain i, the conclusions we draw are more general than this specific Lotka Volterra framework, since they arise from expanding the evolutionary dynamics in the neighborhood of a stationary point: such an analysis should therefore describe a range of models close a "fitness peak."

4.3 Gradual assembly in a spatiotemporally chaotic ecological community

In Chapter 4 we again explore the dynamics of repeated addition of strains into an ecosystem, but now with very different ecological dynamics. Instead of dynamics that reach a fixed point, we consider chaotic ecological dynamics, which arise from Equation 1 when V is taken to have all elements $\mathcal{O}(1)$, and *negative* correlations between its cross-diagonal elements (see Figure 2). This form of the interactions is natural for considering fine-scale diversity, as a strain's self-interaction is of comparable magnitude to its interaction with any other strain. The chaos that ensues results in the abundances of each strain undergoing boom-bust dynamics. Generically this chaos has an unbounded range of fluctuations in log-space, and so leads to uncontrolled extinctions. However the chaos can be maintained for longer times in the presence of *spatial structure* which takes the form of a number of islands, with small migration between each pair of islands. The desynchronization of the chaos ensures that even when

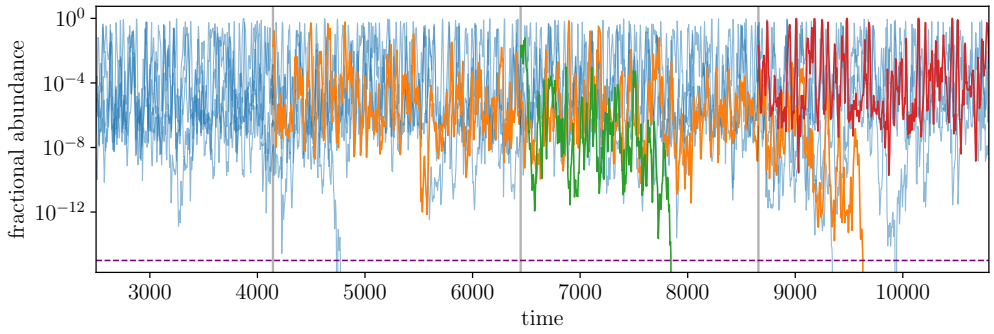

Figure 3: A simulation of the dynamics of gradual assembly (evolution with independent invaders) in the spatiotemporally chaotic phase. Only a subset of strains are shown (on a single island) for visualization purposes. New strains (shown in colors other than blue) are introduced sequentially—triggering extinctions and shifting mean abundances of other strains—with the chaotic ecological dynamics having time to relax to a statistical steady state between invasions (denoted by vertical gray lines). What is the long-term outcome of such a process?

strains go locally extinct, they do not go globally extinct, and subsequent blooms to large abundance will repopulate local extinctions, resulting in a chaotic state that lasts exponentially long in the number of islands [26].

Adding strains one-by-one to this chaotic ecosystem (as in Figure 3) allows the gradual buildup of fine-scale diversity. Unlike in the previously-discussed stable models where the scale of the number of coexisting strains is set by the number of resources D, here there are no resources or niches, and as such there is no scale for the diversity: as a result, given a sufficiently large initial community, one can build up to arbitrarily high diversity with gradual addition of independent strains. The eco-evolutionary phase here is not of the same Red Queen type as observed in the models with a stable fixed point, but is instead one where the diversity continues to increase instead of saturating, and evolution continues forever. If there is some distribution of intrinsic growth rates s_i in the population, then evolution slows down as the population pushes into the tail of this distribution, but diversity is still unbounded if the tail of the s distribution decays fast enough.

4.4 Evolution in a spatiotemporally chaotic model of bacteria and phage

In Chapter 5 we explore evolution in yet another ecological model: this one also spatiotemporally chaotic but now in a more realistic context with interaction structure endowed by phenotypes, which were missing from Chapter 4. The model is inspired by interacting populations of bacteria and the viruses that infect them—known as bacteriophage. In such a model there are two types of strains, in contrast to the Lotka-Volterra model of Equation 1. One group of strains (within a single niche) constitutes the bacterial population, and the other group constitutes the phage population. Denoting the bacterial and phage strain abundances by b_i and p_x respectively, the population abundances of L_b different bacterial strains—all in a single niche and therefore interacting in the same way with each other—and L_p different phage strains obey

$$\frac{db_i}{dt} = b_i \left(\Upsilon^{(b)}(t) - d_i - \sum_{y=1}^{P} F_{iy} p_y \right) \tag{2}$$

$$\frac{dp_x}{dt} = p_x \left(s_x + \sum_{j=1}^{B} G_{jx} b_j - \Upsilon^{(p)}(t) \right). \tag{3}$$

where $\Upsilon^{(b)}(t) = \sum_i b_i(d_i + \sum_y F_{iy} b_i p_y)$, and $\Upsilon^{(p)}(t) = \sum_x p_x(s_x + \sum_j G_{jx} b_j p_x)$, which fix $\sum_i b_i = \sum_x p_x = 1$. Once again, the resulting boom-bust chaos can be stabilized by spatial structure with migration between different islands. The essential difference between this model and the Lotka-Volterra model of Chapter 4 is that the matrices F and G can inherit structure from D-dimensional phenotype vectors possessed by each bacterial and phage strain, which roughly correspond respectively to the receptor and tail proteins that must bind sufficiently well for a bacterial strain to be susceptible to infection by a given phage.

The presence of finite-dimensional phenotypes—as in Chapters 2 and 3—sets a scale for the diversity and we find that, upon repeated addition of strains to the ecosystem, there is a continually evolving Red Queen phase where the diversity saturates but the composition of the community is constantly turning over. This phase is an example of a Red Queen eco-evolutionary phase on top of a spatiotemporally chaotic ecological background, and therefore adds to our understanding of the various possibilities emerging from a combination of ecology and evolution. The presence of phenotypes allows us to introduce small-effect mutations, which correspond to the biologically-relevant regime of gradual evolution. We conclude with a number of open questions about whether a Red Queen phase exists in the limit of small mutations, as well as questions about how the interaction kernel between phenotypes affects the diversity that can be sustained in such communities.

5 Lessons from simple models

The works in this thesis explore a number of scenarios for the evolution of diversity—including fine-scale diversity—in simple models. One of the broad lessons from these studies is the fact that even simple models display a number of qualitatively different phases: from Red Queen continual evolution both with and without chaotic ecological dynamics (Chapters 2, 3, and 5), to an evolutionary dynamics which slows down and converges to a fitness peak (Chapter 3), to the slowing-down oligarch phase with stable ecological dynamics (Chapter 2), to the unlimited diversification (with a concomitant slowdown in the presence of general fitnesses) seen in the chaotic Lotka-Volterra model of Chapter 4. In many cases, we see that *evolution is not an optimization process*. In particular, with complicated enough ecology, which is biologically generic, evolution *cannot* be seen as gradient ascent on a landscape, due to the fact that traversing the landscape modifies the environment and therefore the landscape itself. In this way, the "Red Queen hypothesis" [23] appears to apply quite generally—though not universally—for reasons that we can begin to quantify.

6 Future theoretical directions

There are a number of theoretical directions that remain unexplored at this confluence between ecology and evolution. One wide-open question is how finite-dimensional spatial structure impacts the eco-evolutionary dynamics of communities. The spatial structure invoked to stabilize the chaos in Chapters 4 and 5 has migration between every pair of islands, giving us a kind of spatial mean-field. However, introducing more structure into the migration network is a necessary future step, either through a graph embedded in finite dimensions or a sparse network.

Another interesting direction concerns the impact of *horizontal gene transfer* on eco-evolutionary dynamics. Thus far we have studied models in which new strains emerge either independently from the community composition or as small mutations on top of a parent strain. How does sex, which shuffles genes between different individuals to produce offspring, impact the rate and dynamics of evolution, particularly in the presence of ecology? It is known that recombination can play a large role in shaping the genetic diversity of microbial populations [9, 31, 10, 32]: therefore studying its impact in simple models with explicit phenotypes is needed to interpret its signature in data.

All the work herein assumes that evolution occurs on timescales much slower than ecology. In microbial communities, given the large population sizes and concomitant high rate of mutation production,

this separation of timescales is not guaranteed to be valid [16]. Therefore more work is needed on understanding the dynamics of multiple mutations that occur and compete in an ecologically structured population.

Finally, exploring the dynamics of *sparsely interacting* ecosystems is of great interest and clear experimental relevance, since measured interactions matrices—for example in communities of bacteria and phage—are demonstrably sparse [33]. What fraction of the phenomenology described here remains true with sparse interaction matrices governing the population dynamics? In particular, with sparse interactions changing through evolution, what are the resulting structures that emerge in interacting ecosystems, and how do they compare with empirical findings [34]?

7 Future empirical directions

We conclude this introduction with a list of experimental directions of interest, biased by the theoretical directions explored here, with the hope to spark more of a dialogue between experimental and theoretical approaches to the understanding of microbial diversity. There are a number of datasets—some of which already exist—whose analysis would allow us to assess the viability of the scenarios we have explored. (1) Spatially-resolved timeseries of microbial strain abundances [35] allow us to assess turnover of biodiversity across space, which is related to the spatiotemporally chaotic scenario of biodiversity maintenance. Does one observe different dynamics at different levels of diversity? It is known that family-level abundances can look quite stable over time with dynamics at finer scales looking much more erratic [36]. Is this true across space as well? (2) Can the process of ecological diversification be observed in real time? Though Richard Lenski's long term evolution experiment, designed to be as simple as possible, still exhibited diversification, it is seen to occur on much shorter timescales in more complex environments [37]. Can we use the power of sequencing to understand in more detail how microbial strains diversify and coexist on observable timescales? (3) Synthetic microbial communities offer us a large degree of control in both gradual (bottom-up) and sudden (top-down) assembly of a large number of strains, and are increasingly being used to test theoretical predictions [38]. Can we leverage this control to test expectations from simple models of assembly described here, which generically result in a constantly-turning-over extant community, with new strains always coming in and driving others extinct, but with roughly constant coexisting diversity? (4) Just as single-cell genomics of bacteria have allowed us to look at fine-scale diversity in populations where isolates cannot be grown in the lab [9, 39], can we better understand the fine-scale diversity of *viruses*, which is known to exist [10], through single-virus genomics [40]? What scenarios of viral-host coevolution are ruled out or supported by empirical data?

References

[1] Kenneth G Wilson. The renormalization group and critical phenomena. *Reviews of Modern Physics*, 55(3):583, 1983.

[2] John J Hopfield. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences*, 71(10):4135–4139, 1974.

[3] Shinichi Sunagawa, Luis Pedro Coelho, Samuel Chaffron, Jens Roat Kultima, Karine Labadie, Guillem Salazar, Bardya Djahanschiri, Georg Zeller, Daniel R Mende, Adriana Alberti, et al. Structure and function of the global ocean microbiome. *Science*, 348(6237):1261359, 2015.

[4] A framework for human microbiome research. *Nature*, 486(7402):215–221, 2012.

[5] Noah Fierer. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology*, 15(10):579–590, 2017.

[6] Eoin L Brodie, Todd Z DeSantis, Jordan P Moberg Parker, Ingrid X Zubietta, Yvette M Piceno, and Gary L Andersen. Urban aerosols harbor diverse and dynamic bacterial populations. *Proceedings of the National Academy of Sciences*, 104(1):299–304, 2007.

[7] Thomas Sakoparnig, Chris Field, and Erik van Nimwegen. Whole genome phylogenies reflect the distributions of recombination rates for many bacterial species. *Elife*, 10:e65366, 2021.

[8] Nadav Kashtan, Sara E Roggensack, Sébastien Rodrigue, Jessie W Thompson, Steven J Biller, Allison Coe, Huiming Ding, Pekka Marttinen, Rex R Malmstrom, Roman Stocker, et al. Single-cell genomics reveals hundreds of coexisting subpopulations in wild prochlorococcus. *Science*, 344(6182):416–420, 2014.

[9] Gabriel Birzu, Harihara Subrahmaniam Muralidharan, Danielle Goudeau, Rex R Malmstrom, Daniel S Fisher, and Devaki Bhaya. Hybridization breaks species barriers in long-term coevolution of a cyanobacterial population. *bioRxiv*, 2023.

[10] Marcia F Marston and Christopher G Amrich. Recombination and microdiversity in coastal marine cyanophages. *Environmental Microbiology*, 11(11):2893–2903, 2009.

[11] Kevin R Arrigo. Marine microorganisms and global nutrient cycles. *Nature*, 437(7057):349–355, 2005.

[12] James M Tiedje, Mary Ann Bruns, Arturo Casadevall, Craig S Criddle, Emiley Eloe-Fadrosh, David M Karl, Nguyen K Nguyen, and Jizhong Zhou. Microbes and climate change: a research prospectus for the future. *MBio*, 13(3):e00800–22, 2022.

[13] Megan E Frederickson and Aspen T Reese. Microbial ecology and evolution are essential for understanding pandemics. *Mbio*, 12(5):10–1128, 2021.

[14] Marta Łuksza and Michael Lässig. A predictive fitness model for influenza. *Nature*, 507(7490):57–61, 2014.

[15] Nandita R Garud, Benjamin H Good, Oskar Hallatschek, and Katherine S Pollard. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. *PLoS biology*, 17(1):e3000102, 2019.

[16] Shijie Zhao, Tami D Lieberman, Mathilde Poyet, Kathryn M Kauffman, Sean M Gibbons, Mathieu Groussin, Ramnik J Xavier, and Eric J Alm. Adaptive evolution within gut microbiomes of healthy people. *Cell host & microbe*, 25(5):656–667, 2019.

[17] Sasha F Levy, Jamie R Blundell, Sandeep Venkataram, Dmitri A Petrov, Daniel S Fisher, and Gavin Sherlock. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, 519(7542):181–186, 2015.

[18] Richard E Lenski. Revisiting the design of the long-term evolution experiment with escherichia coli. *Journal of Molecular Evolution*, 91(3):241–253, 2023.

[19] Robert MacArthur. Species packing and competitive equilibrium for many species. *Theoretical Population Biology*, 1(1):1–11, 1970.

[20] G. E. Hutchinson. The Paradox of the Plankton. *The American Naturalist*, 95(882):137–145, 1961.

[21] Benjamin H Good, Michael J McDonald, Jeffrey E Barrick, Richard E Lenski, and Michael M Desai. The dynamics of molecular evolution over 60,000 generations. *Nature*, 551(7678):45–50, 2017.

[22] J Cesar Ignacio-Espinoza, Nathan A Ahlgren, and Jed A Fuhrman. Long-term stability and red queen-like strain dynamics in marine viruses. *Nature Microbiology*, 5(2):265–271, 2020.

[23] Leigh Van Valen. A new evolutionary law. *Evolutionary Theory*, 1:1, 1973.

[24] Eugene P Wigner. Random matrices in physics. *SIAM review*, 9(1):1–23, 1967.

[25] Robert M May. Will a large complex system be stable? *Nature*, 238(5364):413–414, 1972.

[26] Michael T Pearce, Atish Agarwala, and Daniel S Fisher. Stabilization of extensive fine-scale diversity by ecologically driven spatiotemporal chaos. *Proceedings of the National Academy of Sciences*, 117(25):14572–14583, 2020.

[27] Robert A Armstrong and Richard McGehee. Competitive exclusion. *The American Naturalist*, 115(2):151–170, 1980.

[28] Doebeli Michael and Ispolatov Iaroslav. Chaos and unpredictability in evolution. *Evolution*, 68(5):1365–1373, 2014.

[29] Daniel S Fisher. Inevitability of red queen evolution driven by organismic complexity and simple feedback via environmental modification. *bioRxiv*, pages 2021–09, 2021.

[30] Michael Doebeli and Ulf Dieckmann. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *The american naturalist*, 156(S4):S77–S101, 2000.

[31] Michael J. Rosen, Michelle Davison, Devaki Bhaya, and Daniel S. Fisher. Fine-scale diversity and extensive recombination in a quasisexual bacterial population occupying a broad niche. *Science*, 348(6238):1019–1023, 2015.

[32] Zhiru Liu and Benjamin H Good. Dynamics of bacterial recombination in the human gut microbiome. *Plos Biology*, 22(2):e3002472, 2024.

[33] Kathryn M Kauffman, William K Chang, Julia M Brown, Fatima A Hussain, Joy Yang, Martin F Polz, and Libusha Kelly. Resolving the structure of phage–bacteria interactions in the context of natural diversity. *Nature communications*, 13(1):372, 2022.

[34] Cesar O Flores, Justin R Meyer, Sergi Valverde, Lauren Farr, and Joshua S Weitz. Statistical structure of host–phage interactions. *Proceedings of the National Academy of Sciences*, 108(28):E288–E297, 2011.

[35] Christian Santos-Medellín, Katerina Estera-Molina, Mengting Yuan, Jennifer Pett-Ridge, Mary K Firestone, and Joanne B Emerson. Spatial turnover of soil viral populations and genotypes overlain by cohesive responses to moisture in grasslands. *Proceedings of the National Academy of Sciences*, 119(45):e2209132119, 2022.

[36] Antonio M Martin-Platero, Brian Cleary, Kathryn Kauffman, Sarah P Preheim, Dennis J McGillicuddy, Eric J Alm, and Martin F Polz. High resolution time series reveals cohesive but short-lived communities in coastal plankton. *Nature Communications*, 9(1):1–11, 2018.

[37] Paul B Rainey and Michael Travisano. Adaptive radiation in a heterogeneous environment. *Nature*, 394(6688):69–72, 1998.

[38] Jiliang Hu, Daniel R Amor, Matthieu Barbier, Guy Bunin, and Jeff Gore. Emergent phases of ecological diversity and dynamics mapped in microcosms. *Science*, 378(6615):85–89, 2022.

[39] Wenshan Zheng, Shijie Zhao, Yehang Yin, Huidan Zhang, David M Needham, Ethan D Evans, Chengzhen L Dai, Peter J Lu, Eric J Alm, and David A Weitz. High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. *Science*, 376(6597):eabm1483, 2022.

[40] Joaquín Martínez Martínez, Francisco Martinez-Hernandez, and Manuel Martinez-Garcia. Single-virus genomics and beyond. *Nature Reviews Microbiology*, 18(12):705–716, 2020.