Title: Nothing in evolution makes sense (is predictable) except in the light of biology

Authors (order determined by random number generation):

Martha O. Burford Reiskind 1, Michael Moody 2, Dan Bolnick 3, Charles Hanifin 4, Caroline Farrior 5.

Affiliations:

1. **Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695, USA, mbreiski@ncsu.edu**
2. **Department of Biological Sciences, University of Texas at El Paso, El Paso, 79968 USA mlmoody@utep.edu**
3. **Department of Ecology and Evolutionary Biology, University of Connecticut, 75 North Eagleville Road, Storrs, CT, USA. daniel.bolnick@uconn.edu**
4. **Department of Biology, Utah State University Uintah Basin Campus, Vernal UT 84078 USA, charles.hanifin@usu.edu**
5. **Department of Integrative Biology, University of Texas at Austin, Austin TX 78712, USA. cfarrior@utexas.edu**

As the fundamental theory of biology, evolution has the potential to unite and integrate all of the disparate disciplines present in the modern biological sciences. However, questions about the predictability of evolutionary processes and outcomes are still unresolved at multiple levels of biological organization and time scales. Addressing these questions can provide a framework to scaffold forthcoming inquiry and identify focal areas for future work. There are clear reasons to believe that evolution can be predicted, and yet, there is experimental evidence and theory suggesting that evolution is fundamentally unpredictable (Orr 2005, Stern & Orgogozo 2009, Bolnick et al. 2018). There are examples that support the predictability of evolution at relatively short time scales and functional genes under similar selective pressures within populations (Ramiro et al. 2016, Monroe et al. 2016,

Hawkins et al. 2019). Patterns of parallel genomic evolution have also been recognized among closely related lineages (Renaut et al. 2014, Seehausen et al. 2014, Irwin et al. 2016, Ravinet et al. 2017). For example, Renaut et al. (2014) using genomic scans found repeatable patterns of divergence in large, shared genomic regions with similar genomic architecture, among three pairs of sunflower species and they related this to similar selective pressures. However, some of these same studies provide alternative examples of the unpredictability of evolution from single nucleotide polymorphisms (SNPs) to genomic architecture (e.g. Renaut et al. 2014, Ravinet et al. 2017). Other studies emphasize the unpredictability of evolution (Bolnick et al. 2018, Langerhans 2018, Sailer & Harms 2017, Fitzpatrick et al. 2014, Tegze et al. 2012, Takahashi et al. 2007). For instance, Langerhans (2018) found over 50% of the phenotypic variation in Bahama mosquito fish was not explained by predation across replicated blue holes with and without predators. At a molecular level, unpredictability may result from the fundamental dynamics of protein structure. Sailer & Harms (2017) showed that complex combinatorial effects of mutations on protein shape prevented prediction of evolutionary trajectories even after accounting for mutational effects and pairwise epistatic interactions. Ultimately, much of this work suggests that evolutionary predictability may only be possible at specific time scales or levels of biological organization, but not others.

We argue that given accelerated global forcing such as global climate change, land use changes, introduced species, and increased global pandemics this is a critical time to address whether we can make predictions about how species will evolve. At larger spatial scales (biomes) and time scales (decades to centuries), if we can relax knowledge of which specific individuals or species evolve, we may have greater predictive power to understand the effects of these global forcings. For example, at the ecosystem level, ecologists use evolutionary theory as a predictive tool in understanding the variation of plant strategies across landscapes and experimental treatments (Farrior et al. 2013, McNickle et al. 2016, Weng et al. 2017, Dybzinski et al. 2019). This means that while we may not be able to say which

tree will form a canopy, we can be confident that some tree will. Given the above-mentioned global threats, it is critical to understand whether communities will re-assemble according to simple processes of individual-based selection to generate corresponding changes in ecosystems. Predicting the evolutionary trajectories of populations could help us preemptively respond to human health concerns (tumors, disease, etc.), conservation efforts, agriculture and pest management, ecosystem management, and engineering of sustainable systems.

Here, we propose there is a continuum of predictability of evolution and that our ability to predict decays with time, from short (e.g. 1 generation) to longer time scales (e.g. 1,000,000 generations). We also recognize that predictability at the genome level can also be scale dependent from SNPs to aspects of genomic architecture (Lee & Coop 2019). Because we are not focusing on predicting the evolutionary history of life from first principles, to avoid confusion, we specify what we mean by prediction, and discuss the many factors holding us back and the multifarious problems we need to solve to improve our predictive power. We address this topic across different levels of biological organization from molecular, to populations, to ecosystems and different spatial scales from local to global scales. Finally, we argue that the only way we can address this question is from an integrated approach that encompasses the expertise and tools found in fields of molecular biology, genetics, developmental biology, evolutionary theory, ecology, phylogeny, paleontology, biomathematics, and that ultimately our predictive power in biology will come from an integrative biological perspective on evolution. To paraphrase the great evolutionary biologist of the mid-20th century, Theodosius Dobzhansky: *Predicting evolution is only possible in light of the rest of biology, while understanding biological processes and patterns is possible only in light of evolution.*

"I do not think it means what you think it means"

When discussing evolutionary prediction, we must first settle on a shared vision of what it means to predict. Although this seems simple at first glance, there are many biological levels of organization and various degrees of precision. As we build from simple statements about evolution, our definition of prediction becomes more nuanced and specific:

a) At the simplest level, we can predict that evolution will take place. This is rarely in doubt, and rarely stoppable except in the case of extinction. Here we are not taking the adaptionist view that given enough time, life would occupy mars, but because evolution is ubiquitous, predicting its existence is trivial and not especially helpful. While not all traits will evolve, the question is which ones do (see (c), below)?

b) A slightly more useful statement would be to say that, in a given environment, mean fitness should increase over time (as long as there is genetic variation for fitness). More specifically, the rate of increase in mean fitness is proportional to the genetic variance in fitness, as laid out in Fisher's Fundamental Theorem of Natural Selection (Fisher 1930, Shaw 2019). This statement is preferable to (a) because it specifies a metric (mean fitness), direction (increasing mean fitness to an adaptive landscape peak), and a rate. Yet, it lacks mechanistic detail and so has relatively little utility for applied problems. Note also that mean fitness can also increase through non-evolutionary means, via adaptive phenotypic plasticity, matching habitat choice, or niche construction (Edelaar & Bolnick 2019).

c) A more useful and interesting goal of evolutionary biology is to predict which of the vast array of traits are likely to evolve in response to a particular selective challenge (e.g., an environmental change). Not all traits will evolve, as many may be at equilibria (e.g., subject to stabilizing selection) or are neutral so their evolution is too slow to be relevant to the time scale in question. But, typically, at least some traits and genes are likely to be evolving at any point in

time; the question is merely which ones. Identifying these genes requires a detailed understanding of how traits generate functions, thus determining fitness, a topic to which we will return later.

d) If we can indeed identify the correct traits that will evolve over the relevant time scale (c), we might then aspire to a greater degree of predictive power: in what direction will the trait(s) change? Predicting directionality should generally be within our reach because it simply requires knowledge of the sign of the slope of the selection gradient(s), acting on the population mean(s) of the relevant trait(s).

e) Better still, can we make a quantitative prediction? By what amount (fold change, standard deviations, proportion, etc.) will the chosen trait(s) change over a specified unit of time? Such quantitative prediction is an achievable end goal for quantitative geneticists. A related goal is predicting changes in the full trait distribution (e.g., variance, kurtosis), but this is a harder problem. There is no simple equation, akin to the Breeder's Equation, for predicting the evolution of genetic covariances and higher moments.

f) The quantitative genetic approach often treats genetic, molecular, cellular, developmental mechanisms as a black box, focusing on emergent and readily observable traits such as size, shape, behavior, etc. A more complex goal is to predict evolution and action of finer-scale mechanistic traits that ultimately generate the traits of interest at the organismal level (which we might call upstream traits). Examples might include levels of gene expression, pathway activity, enzymatic activity or concentrations, developmental patterning, etc. We could also study these upstream traits to attempt to predict which ones will evolve (c), in what direction (d), and by how much (e). This research agenda lets us predict not just evolution of the obvious traits, but also a mechanistic biological explanation of how these trait changes are actuated by changes in gene expression, development, etc. Ultimately all the phenotypic traits we might

choose to study, arise from changes in the expression of genes (where, when, how much), their translation (speed, timing, splicing), and subsequent protein function (folding, active site properties, dynamics, transport, degradation, interactions). These all have their roots in the sequence, packaging, and epigenetic modification of DNA. Thus, many biologists feel that the ultimate question of evolutionary prediction is to anticipate the precise genetic changes underlying evolution. We can define distinct levels of predictive precision within this ultimate question (Box 1).

g) The preceding kinds of evolutionary predictions are all concerned with evolution that is occurring within a particular focal population (changes in trait distributions and genotype frequencies in a defined group of individuals). Yet, evolution is still more complex in that it occurs in many interacting species simultaneously. Far more challenging than the 'ultimate question' defined above, is to predict the course of evolution (for traits or genomes) of ecological communities of two or more interacting species. This is especially crucial (as we explore more below) because species interactions are a major driver of evolution for constituent species, whose evolution in turn alters their interactions that will lead to dynamic changes in each other's selection landscapes (so-called "eco-evo feedbacks").

Box 1. Distinct Levels of Predictive Precision in Molecular Evolution

i) Evolution will occur in a particular group of genes (e.g., gene ontology category, pathway, family of paralogs).
ii) Evolution will occur in a particular gene.
iii) Evolution of that gene will entail changes in particular motifs or properties (e.g., a shift in polarity or shape, or within a particular active site).
iv) Evolution will entail changes in frequency of particular genetic variants (e.g., single nucleotide polymorphisms [SNPs], indels, gene copy number, chromosomal rearrangements, etc.).
v) Predicting evolution of single loci is insufficient, because evolution is rarely a single locus process. For instance, initial adaptive changes might impose costs that require compensatory mutations after. Therefore, for true predictive power we should aspire to scale up goals g.i - g.iv to multiple genes, how they interact, and ultimately the whole genomic shebang (many loci, architecture, and epigenetics).

To summarize, we frequently use "predict evolution" as a convenient shorthand that encompasses a wide range of goals with varying degrees of precision, qualitative or quantitative, applied to various scales of organization (e.g. genes, genomes, species, communities) due to a range of mechanisms (e.g. selection, fitness, genetic architecture, species interactions). It is equally crucial that we clearly specify the time scale over which our prediction applies. Some of these kinds of predictions seem well within our reach at present, others seem like moonshots and may strain our current theory and technologies, or some may be fundamentally impossible.

There is no question that the more ambitious forms of evolutionary prediction are beyond our reach, and may always be so. But, even for these intractable problems, we believe it is important that we distinguish between two distinct views:

H1: Evolution is predictable, if we simply had the right models and data

H2: Evolution is fundamentally unpredictable, not simply because we lack sufficient knowledge

H1. Evolution is predictable; we just don't know how yet (models and data)

Our poor performance at correctly predicting evolutionary outcomes may result from incomplete models of evolutionary and biological processes or insufficient data (spanning molecular to organismal to ecosystem levels) to parameterize such models. We will need to make *conceptual progress* in all of the following areas to create a successful predictive model of evolution. Note that there already exists a large literature of evolutionary theory that provides key building blocks of such a model, drawing on both population genetics and quantitative genetics (e.g., the Breeder's Equation; Walsh & Lynch 2018). This literature has led to useful tools like SLiM (Haller & Messer 2019), that can carry out whole-genome forward-in-time, spatially explicit population genetic simulations with recombination, mutation, selection, migration, in short almost all the evolutionary processes we might wish to incorporate. Yet, even this powerful new tool excludes many processes that we know shape the direction of evolution,

including epistasis, genotype-phenotype mapping, plasticity, species interactions, population dynamics, and many more points detailed below. Then, when we have a satisfactory predictive-evolutionary model in hand, to apply this model to any real biological system we would require extensive, perhaps prohibitive empirical data to actually parameterize the model to generate the desired predictions. Thus, to predict evolution, we need both *conceptual progress* and *data*, which we detail below. Specifically, we need:

1) A better understanding of the process of mutation. This includes variation in mutation rates, more precise measures of the rates of transitions versus transversions, the distribution of mutational hotspots and cold spots within the genome, and frequency of insertions and deletions. Not to mention better understanding of differences among species, among individuals or germlines within species, and within genomes. Such data exist for a few species (e.g. Smeds et al. 2016, Ellegren 2003), but their precision can be improved and we need to determine if the data can be generalized to more species or even within species. For short-term evolutionary prediction, *de novo* mutation has negligible effect on evolution simply for lack of generational time (with some possible exceptions, e.g. Hawkins et al. 2019), so we can likely ignore these *de novo* mutations. Because mutation is fundamentally stochastic, we will generally not be able to predict when (or whether) a specific mutation at a specific locus will occur. But, over very long time scales, the law of large numbers might work in our favor to regain some aggregate statistical predictive power. For instance, in very large populations with large per-base mutation rates (e.g. HIV within a patient; Cuevas et al. 2015), one can reliably predict that every feasible mutation will occur within a short time span, so the stochastic nature of individual mutations ceases to matter.

2) The ability to predict how a given genetic change (SNP, indel, etc) will produce a change in phenotype and function. This problem encompasses almost the entirety of biology, from

genetics to development, physiology, immunology, cell biology, biomechanics, etc., requiring understanding of protein function, gene networks, patterning, etc. In a sobering sense, detailed evolutionary prediction might need to wait until the rest of biology has finished their work. Or, might we get around this difficulty by reverse-engineering this process? If we know the selective pressure, might we anticipate the physiological, developmental, or protein structure changes required to adapt to this selection. From that inference, could we describe the suite of mutations that could achieve, individually or collectively, the protein or expression change? That is, would we do a better job of predicting evolution if we studied function ← genotype mapping, instead of genotype → function mapping? This reversed approach might be especially valuable in applied practice. If we seek to predict how a specific organism will evolve in response to a particular environmental change, the forward approach requires that we consider all possible mutations (or all possible genetic variants) in the genome, and how each of those changes might affect phenotype, function, and ultimately fitness. The reverse strategy could allow us to narrow the scope of problems to consider. For example, if we think of selection imposed by a novel pathogen, we might immediately narrow our focus onto immune genes, and in particular the most relevant specific genes (e.g., pattern recognition genes), and the specific motifs that need to change to successfully bind to and recognize the new pathogen. We thus focus on the computational problem for greater efficiency, though the problem remains daunting in practice and might miss other adaptive pathways in response to the same selective pressure (e.g. glucose aversion versus insecticide resistance in cockroaches; Wada-Katsumata et al. 2011). Encouragingly, theoretical models combined with bioinformatics based data seem to predict well the evolved distribution of biophysical properties of proteome (collection of all the proteins in a species) (Zou et al. 2014, Zeldovich et al. 2007).

3) The mapping of genotype to phenotype (2, above) is contingent on many factors including the environment (genotype by environment interactions (GxE)) and the effect of one gene's alleles on phenotype are also conditional on alleles at other gene(s) (epistasis). Therefore, genome wide allele frequency data, and covariance among alleles, needs to be known along with the epistatic effect of pairs of alleles at different genes. Three-way gene interactions and higher order epistasis make our brains hurt and are computationally challenging to model, but are exponentially more abundant than pairwise interactions (Kuzmin 2018). In practice, therefore, this goal requires understanding of GxE and epistatic modifications of the genotype→ phenotype mapping. One must therefore know genome-wide allele frequencies (for epistasis) and the range of environmental conditions that individuals experience.

4) A better understanding of the impact of genomic (genetic) architecture on the response of individual genes to selection. Taking a systematic approach, genomic regions where loci of high adaptive value are clustered, can be found in parallel within lineages and identified for multiple taxa (Yeaman 2013, Holliday et al. 2016, Raeymaekers et al. 2017). One mechanism is chromosome inversions that appear to suppress recombination in heterozygotes and may act as reservoirs of standing genetic variation (Morales et al. 2019). Then, in practice, we need an empirical description of how architecture varies within a focal population (e.g., inversion polymorphisms, and their effects on key genetic properties such as recombination and mutation rates). In other cases, strong selection pressure with a high fitness cost may result in rapid adaptation within a species that is similar across populations but vary in the genetic architecture or genomic regions (e.g. female choosiness against interspecific mating between species of *Aedes*, Burford Reiskind et al. 2018). In these cases, we need to know what combinations of genes are contributing to the phenotype.

5) Points 1-4 above are essentially the field of genetics, broadly defined. But, to predict evolution we need to go a step further and link phenotype to fitness. At the coarsest level this could entail statistical description of covariation between phenotype and fitness (e.g., for the Price equation; Queller 2017), or between genotype and fitness. Currently, we have the statistical tools for this approach. While this might suffice over very short time scales, we lack mechanistic understanding and therefore it will be hard to project further into the future. Projecting into the future requires a functional understanding of how traits affect fitness, drawing on biomechanics, behavior, ecology, etc. We lack the capacity, at present, to model how present-day traits, let alone traits that do not yet exist, generate variation in fitness.

6) Once we understand genetics (points 1-4), and selection (point 5), we can use existing tools of quantitative or population genetics to predict the course of evolution. This genetic knowledge can in some instances be simplified by omitting mechanistic detail and taking a quantitative genetic approach (e.g., the Breeder's equation), or population genetics for the rare case of simple single-gene traits. This is an established approach that works well over short time scales, but will break down over longer time scales because we lack a mechanistic model of how genetic variance-covariance matrices themselves evolve (we know that they do evolve), and how selective pressures will change. Thus, the more mechanistic approach in points 1-4 provides a potentially robust framework, but one that is harder to parameterize (if possible at all). Whether one takes a mechanistic, quantitative, or population genetics approach, the key is incorporating knowledge about the available genetic variation, how this affects fitness, and how response to the resulting selection is constrained.

7) The entire endeavor 1-6 above leads to our ability to predict evolution in response to known selective pressures. This works well when the present-day environment can be safely trusted to remain constant. Yet, selection on our focal species depends on abiotic conditions,

and biotic interactions, both of which change through time and must be forecast for evolutionary prediction. To do so, we must draw on fields ranging from meteorology (climate change being a major driver of evolution during the Anthropocene), to toxicology (from human pollution), to epidemiology and ecology more generally. We therefore need detailed data on the present-day state of multivariate environmental and ecological factors (e.g., species densities of predators, parasites, prey, competitors, mutualists), and the rules of how these change through time (e.g., how species interact to drive each other's changing population densities).

8) With models of changing selective pressures in hand (for example, the earth system models used by the IPCC for environmental changes, IPCC 2014), we then need to revisit the phenotype → function → fitness mapping (point 5, above) in the context of potential future environments (including possible future communities and interactions). This will allow us to predict how the fitness landscape will shift through time to favor different trait values, trait combinations, and genotypes at various points in the future.

9) Lastly, the evolution of our focal species is embedded within a community of other evolving species. Evolution of those other species can change the nature of their interactions with our focal species, thereby changing the fitness landscape even when population densities themselves have not changed. Conversely, evolution by our focal species can have reciprocal effects on the abundance, genotypes, phenotypes, and fitness of all the other species with which it interacts directly (or perhaps indirectly), driving evolution and ecological dynamics throughout the community. These community-wide changes then feedback to affect our focal species' fitness landscape. Although eco-evolutionary feedbacks are a popular topic of theoretical and empirical research today, we have not yet begun to grapple effectively with the community-wide web of evolutionary feedback loops. Over long time-scales, these diffuse co-

evolutionary feedback loops are likely to be essential for predicting evolution, but may never be sufficiently knowable.

The list of things we need to know, above, is sobering. And, it is likely we will never know enough to effectively predict evolution with high levels of precision and mechanism over long time scales and in natural environments. Yet, it is worth asking whether this is merely a practical constraint on our ability to gather sufficient data? Or, is this a fundamental problem that evolution is truly unpredictable regardless of our present knowledge? This leads us to hypothesis 2:

H2 evolution is not predictable, no matter how much we measure

Stephen Jay Gould famously argued in *Wonderful Life: The Burgess Shale and the Nature of Life* (1989) that evolution would not repeat itself; if we rewound the 'tape of life' and replayed it from the Cambrian, we would be unlikely to end up with anything like humans. In this spirit (and on a shorter time span), we posit that evolution is inherently unpredictable at the molecular and population level. This is due to multiple factors scaling from molecular to environmental. Selection is dependent on genetic mutation, which is random. Even if we know the genes and genetic pathways that should be important under a specific selective pressure, reliance on *de novo* genetic mutation makes it impossible to determine when/if evolution will be possible in any specific form. Therefore, while laboratory experiments can potentially determine what mutations could be adaptively beneficial we cannot predict how or where in the genome mutations will occur in nature. Thus, predictability of specific allelic distributions or morphological form is not going to be possible when evolution is mutation-limited. Genetic stochasticity (drift, recombination, allelic segregation) provide additional molecular level variation. We can describe probability distributions, but not make specific statements of what will happen. This is an important point because we normally assume that evolution is predictable over the short term but not long term. Yet, in the short to medium term, evolution relies on a small sample size

of new mutations, which are stochastic and so inherently not predictable. Over the very long term, there are enough opportunities for new mutations that perhaps our probabilistic tools regain predictive utility.

Equally (or more) important is the ubiquity of epistatic interactions both within and among genes. There is now widespread evidence that the order in which substitutions occur has a dramatic impact on both the magnitude and sign of their phenotypic and fitness effects (Costanzo et al. 2010). It is possible the complexity of order effects coupled with the inherent randomness of the mutational process may render efforts to develop complete predictive models to overcome the genotype to phenotype process inherently impossible (as opposed to practically impossible) (Sailer & Harms 2017).

Environmental stochasticity is a major source of fundamental unpredictability: climate, and community dynamics. The environments in which some organisms or populations exist may prove so variable and/or unstable that a consistent model (or prediction) of environmental and gene by environment may not be possible as time scales increase. Chaotic dynamics in particular suggest that there are fundamentally unpredictable changes in conditions (as opposed to our theories being incomplete). However, time scale matters. Over short to medium time scales (years to decades), chaotic dynamics mean that we have no capacity to predict future environmental conditions that could impose selection on our focal organisms (dependent on population dynamics). Over very long time scales (centuries-millennia), chaotic systems can remain within stable attractors (lacking global catastrophic events), defining a field within which conditions are bounded. Likewise, stochastic processes such as weather might be unpredictable over short time scales (days to weeks) but follow predictable long-term trends (e.g., global warming over the coming centuries, or even cyclic dynamics such as Milankovich cycles). Considering these time scales and relaxing the need to know specific species or individuals, and given predictable long-term trends and the conserved structure of ecological guilds, we should be able to make some predictions (e.g., there will be herbivores, carnivores, dominate tree strategies, etc.).

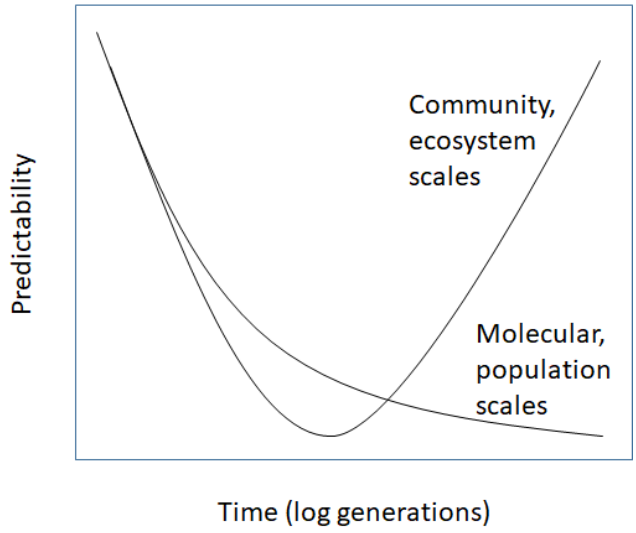

Figure 1. Proposed graphical description of predictability of evolution, while distinguishing two important sales, the molecular and population scales compared to the community and ecosystem scales

What about the future?

Our goal has been to evaluate the future prospects for predicting evolution. Is it predictable at all? If so, what information and modeling capacity do we need? In evaluating these questions, we also show the various ways of conceiving 'prediction' and how these depend on the time scale and biological level over which we wish to predict. In general, we develop the hypothesis that the predictability of evolution at the molecular and population levels decreases with the increasing time horizon (Figure 1). That is, we may not be able to predict the specific species that will overcome challenges, but if we can predict the environment in the future and have an understanding of the standing variation and traits under selection, we may be able to predict beyond a few generations. We are argue, that even distinguishing between H1 and H2 is useful and that the research and the data generated testing H2 will still provide important knowledge about the evolutionary process in general.

We also propose that the predictability of the results of evolutionary processes at the level of ecosystems may be high at long timescales (Figure 1). Over long time scales, evolution has some potential to be predictive of the functional groups that appear in ecosystems. And, if we know the environment in the future, we may be able to predict the specific traits of the species that make up the functional groups within ecosystems (e.g. canopy trees, predators, carnivores, etc.).

We believe there is a great deal to be learned through the intersection of ecological and evolutionary approaches to addressing many applied questions needing predictive biological solutions including: tumor growth, HIV spread, resistance to control approaches in pathogens and pests, vectors, energy, ecosystem functions, communities, species, computer viruses, memes. In order to make progress on developing the predictive power of biology in a world where the environment is rapidly changing and species are being lost from ecosystems, we must begin to integrate our understanding of the predictive power of evolutionary processes to generate ecosystem functions from whichever species happen to be present in the future.

Here we highlight the areas where research, both basic and applied, and computation modeling should focus to begin to unravel a fundamental question in biology, how and when is evolution predictable. While at the molecular and population level we are, in principle and given the correct information, able to predict change over a few generations, it is difficult to predict the evolutionary trajectories many generations in the future (e.g. 100 to 10,000,000 of generations), where stochastic processes may create novel phenotypes. We call this *predictive decay*. Above we outlined, in a stepwise fashion, the information needed to build predictive models of evolution. Yet, we ultimately need more *conceptual progress* and more *data*, including not only how to predict evolution given a status quo, but how we make predictions given global forcing of climate change, habitat loss, land conversion, and invasive species.

In the future, if we achieve a greater capacity to predict evolution over short and long time scale, we can apply this to address questions at different levels of biological organization. First, at the molecular level, a predictive framework of molecular evolution in an environmental/ecosystem context will allow us to confidently modify genes in organisms to recover species of conservation concern or fight diseases without potential catastrophic consequences of down chain effects. While, at the ecosystem level, once we refine our understanding of physiological constraints, environmental pressures, and trophic interactions, we can tackle issues of restoration ecology, ecosystem management, and predictions of changes in ecosystem function with climate change. With an understanding of the evolutionary mechanisms that will not change themselves with the change in environment and species loss that comes with climate change, we have some capacity to predict the directions of changing ecosystem function in the future.

References

Bolnick DI, Barrett RDH, Oke KB, Rennison DJ, Stuart YE (2018) (Non)Parallel Evolution. *Annu Rev Ecol Evol Syst,* 49: 303-330

Burford Reiskind MO, Labadie P, Bargielowski I, Lounibos LP, Reiskind MH (2018) Rapid evolution and the genomic consequences of selection against interspecific mating – *Molecular Ecology,* 27: 3641-3654

Costanzo M, Baryshnikova A, Bellay J et al. (2010) The genetic landscape of the cell. *Science,* 327:425-431

Cuevas JM, Geller R, Garijo R, Lopez-Aldeguer J, Sanjuan R (2015) Extremely high mutation rate of HIV-1 in vivo. *PLOS Biology,* doi:10.1371/journal.pbio.1002251

Dybzinski R, Kelvakis A, McCabe J, Panock S, Anuchitlertchon K, *et al.* (2019) How are nitrogen availability, fine-root mass, and nitrogen uptake related empirically? Implications for models and theory. *Global Change Biology*, 25(3):885-899

Edelaar P, Bolnick DI (2019) Appreciating the multiple processes increasing individual or population fitness. *TREE,* 24:435 -446

Ellegren H, Smith NGC, Webster MT (2003) Mutation rate variation in the mammalian genome. *Current Opinion in Genetics & Development,* 13: 562-568

Farrior CE, Dybzinski R, Levin SA, Pacala SW (2013) Competition for water and light in closed-canopy forests: a tractable model of carbon allocation with implications for carbon sinks. *Am Nat*, 181:314-330

Farrior CE, Tilman D, Dybzinski R, Reich PB, Pacala SW (2013) Resource limitation in a competitive context determines complex plant responses to experimental resource additions. *Ecology*, 94:2505-2517.

Fisher RA (1930) The Genetical Theory of Natural Selection. The Clarendon Press, Oxford, England.

Fitzpatrick SW, Torres-Dowdall J, Reznick DN, Ghalambor CK, Funk WC (2014) Parallelism isn't perfect: could disease and flooding drive a life-history anomaly in Trinidadian guppies? *Am. Nat.,* 183:290–300

Haller BC, Messer PW (2019) SLiM 3: Forward genetic simulations beyond the Wright–Fisher Model, *Molecular Biology and Evolution*, 36: 632–637 https://doi.org/10.1093/molbev/msy228

Hawkins NJ, Bass C, Dixon A, Neve P (2019) The evolutionary origins of pesticide resistance. *Biological Reviews*, *94*(1), 135-155.

Holliday JA, Zhou L, Bawa R, Zhang M, Oubida RW (2016) Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *P opulus trichocarpa*. *New Phytologist*, *209*(3), 1240-1251.

IPCC, 2014: *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Irwin DE, Alcaide M, Delmore KE, Irwin JH, Owens GL (2016) Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Molecular Ecology*, 25(18), 4488-4507.

Kuzmin E, VanderSluis B, Wang W, *et al.* (2018) Systematic analysis of complex genetic interactions. *Science,* 360 (1-9)

Langerhans RB (2018) Prediction and parallelism of multitrait evolution. *J of Hered,* 59-70

Lee KM, Coop G (2019) Population genomics perspectives on convergent adaptation. *Philosophical Transactions of the Royal Society B*, *374*(1777), 20180236.

McNickle GG, Gonzalez-Meler MA, Lynch DJ, Baltzer JL, Brown JS (2016) The world's biomes and primary production as a triple tragedy of the commons foraging game played among plants. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1842), 20161993.

Monroe JG., McGovern C, Lasky JR, Grogan K, Beck J, McKay JK (2016) Adaptation to warmer climates by parallel functional evolution of CBF genes in Arabidopsis thaliana. *Molecular ecology*, *25*(15), 3632-3644.

Morales HE, Faria R, Johannesson K, Larsson T, Panova M, Westram AM, Butlin RK (2019) Genomic architecture of parallel ecological divergence: beyond a single environmental contrast. *Science Advances*, *5*(12), eaav9963.

Orr HA (2005) The probability of parallel evolution. *Evolution* 58: 216-220

Queller DC (2017) Fundamental Theorems of Evolution. *Am Nat* 189: 345-353.

Raeymaekers JA, Chaturvedi A, Hablützel PI, Verdonck I, Hellemans B, *et al.* (2017) Adaptive and non-adaptive divergence in a common landscape. *Nature communications*, *8*(1), 267.

Ramiro RS, Costa H, Gordo I (2016) Macrophage adaptation leads to parallel evolution of genetically diverse *Escherichia coli* small-colony variants with increased fitness in vivo and antibiotic collateral sensitivity. *Evol App,* 9: 994-1004

Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, *et al.* (2017) Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J of Evolutionary Biology,* 30: 1450-1477

Renaut S, Owens GL, Rieseberg LH (2014) Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology,* 23: 311-324.

Sailer ZR, Harms MJ (2017) Molecular ensembles make evolution unpredictable. *PNAS,* 114: 11938-11943

Shaw RG (2019) From the past to the future: Considering the value and limits of evolutionary prediction. *Am Nat*, 193: 1-10

Seehausen O, Butlin RK, Keller I, Wagner CE, *et al.* (2014) Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176-192.

Smeds L, Qvarnstrom A, Ellegren H (2016) Direct estimate of the rate of germline mutation in a bird. *Genome Research,* doi/10.1101/gr.204669.116.

Stern DL, Orgogozo V (2009) Is genetic evolution predictable? *Science,* 323.5915: 746-751

Takahashi K, Kohno T, Matsumoto S, Nakanishi Y, Arai Y, *et al.* (2007) Clonal and parallel evolution of primary lung cancers and their metastases revealed by molecular dissection of cancer cells. *Hum Cancer Biol*, 13:111–20

Tegze B, Sz´all´asi Z, Haltrich I, P´enzv´alt ´o Z, T´ oth Z, *et al.* (2012) Parallel evolution under chemotherapy pressure in 29 breast cancer cell lines results in dissimilar mechanisms of resistance. *PloS One,* 7:e30804

Wada-Katsumata A, Silverman J, Schal C (2011) Differential inputs from chemosensory appendages mediate feeding responses to glucose in wild-type and glucose-averse German cockroaches, *Blatella germanica*. *Chemical Senses,* 36: 589-600

Walsh B, Lynch M (2018) Evolution and Selection of Quantitative Traits. *Oxford University Press*, New York, Oxford.

Weng E, Farrior CE, Dybzinski R, Pacala SW (2017) Predicting vegetation type through physiological and environmental interactions with leaf traits: evergreen and deciduous forests in an earth system modeling framework. *Global Change Biology*, 23(6):2482-2498.

Yeaman S (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci. *PNAS*, 110: E1743-E1751

Zeldovich KB, Chen P, Shakhnovich EI (2007) Protein stability imposes limits on organism complexity and speed of molecular evolution. *PNAS,* 104:16152-16157)

Zou T, Williams N, Ozkan SB, Ghosh K (2014) Proteome folding kinetics is limited by protein halflife. *PloS One*, 9.e112701