The Statistical Mechanics of Biodiversity

Andrew James Rominger

Biologists have searched for a simple set of universal governing mechanisms that dictate the dynamics of biodiversity since at least the time of Darwin. While much progress has been made in understanding system-specific processes and in documenting the context-dependent roles of such mechanisms as competition and facilitation, we still lack a universal governing rule set. The goal of understanding and predicting biodiversity dynamics comes at a critical moment when human systems are disrupting those very dynamics. In this thesis I approach the long-standing problem with the hypothesis that general patterns in biodiversity emerge from a combination of the statistical mechanics of large systems and the unique non-equilibrium dynamics imparted to biological systems by their evolutionary history. Statistical mechanics provides the key analytical approaches to abstracting the complex details of biodiversity into general macroscopic predictions that I show receive support from empirical data. However, key deviations from the simplest statistical mechanics of biodiversity reveal the key role of biological evolution in driving systems away from the idealized steady state predicted by statistical mechanics.

In Chapter 1 I expanded a branch of non-equilibrial statistical mechanics, known as super statistics, to explain previously unaccounted for wild fluctuations in the richness of taxa through the Phanerozoic marine invertebrate fossil record and show how this non-equilibrium is driven by clades’ punctuated exploration of their adaptive landscapes. This theory provides a novel explanation for deep time diversity dynamics invoking emergence of lineage-level traits as the drivers of complexity via the same mechanisms by which complexity emerges in large physical and social systems. In the context of fossil diversity I show how this complexity arises naturally from the uniquely biological mechanisms of punctuated adaptive radiation followed by long durations of niche conservatism and thus identify these mechanisms as sufficient and necessary to produce observed patterns in the fossil record. I test this theory using two seminal fossil datasets.

In Chapter 2 I use the chronosequence afforded by the Hawaiian Islands to capture evolutionary snapshots of arthropod communities at different ages and stages of assembly to understand how the history underlying an assemblage determine its contemporary biodiversity patterns. I apply static ecological theory of trophic networks based on statistical mechanics to these rapidly evolving ecosystems to highlight what about the evolutionary process drives communities away from statistical idealizations. I have tested static ecological theories of herbivory network structure and the distributions of abundances and metabolic rates across species. This study indicates that rapid assembly from immigration and speciation in young ecosystems and extinction in old ecosystems could drive observed patterns.

In Chapter 3 I highlight and explain the computational requirements to testing one statistical theory of biodiversity—the Maximum Entropy Theory of Ecology—with real data and make those test available in a stream-lined framework via the R package meteR that I authored.

I dedicate this thesis to my guides and sources of solace through the process of completing my Ph.D: to my mother Judy Rominger and my godmother Marsha Keener; to my partner Linden Schneider; and to the kolea *(Pluvialis fulva)* and giant koa *(Acacia koa)* of Hawaii.

I would like to first and foremost thank my co-Chairs, Profs. John Harte and Rosemary Gillespie; their dedication and support to fostering my research and advocating for my contribution to the scientific community has been the defining feature of my graduate experience. I would also like to thank my outside member Prof. Charles Marshall who’s scientific curiosity and practical advice have been critical in forging my path as a researcher.

This work would not have been possible without the support and teamwork of many friends and collaborators. I would like to thank especially Prof. Pablo Marquet, Prof. Miguel Fuentes, Prof. Daniel Gruner, Dr. Cory Merow, Prof. George Roderick, Jun Lim, Dr. Kari Goodman and Dr. Matthew Van Dam for their camaraderie and tireless collaboration. I would like to thank Prof. Manisha Anantharaman, Dr. Naim Darghouth, Dr. Danielle Christianson, Erica Newman, Dr. David Hembry, Jade Zhang and Meredith Jabis for their friendship and support.

This work is rooted in the diversity of nature, particularly the unique biota of Hawaii. I am grateful to those who have dedicated their lives to preserving and restoring that unique legacy, especially Pat Bily, Ed Misaki and Russell Kallstrom of The Nature Conservancy Hawaii, Rhonda Loh and Raina Kaholoa’a with the National Park Service, and Cynthia King and Betsy Gagne with the Hawaii Department of Land and Natural Resources.

I would also like to acknowledge my funders including the National Science Foundation Graduate Research Fellowship, NSF DEB 1241253, the Moore Foundation, the Department of Environmental Science, Policy and Management and the Walker Fund at UC Berkeley.

# Introduction

Biodiversity is a striking feature of our world. Diverse ecosystems provide humanity with irreplaceable services (Daily 1997) and yet predicting biodiversity’s dynamics—a key goal in order to better manage and preserve it—continues to elude ecologists and evolutionary biologists (McGill and Nekola 2010; Gould 1985). Over the past 550 million years biodiversity has fluctuated between periods of rapid diversification, such as the Cambrian explosion, and devastating extinctions, such as the end Permian extinction (Sepkoski 1992; Alroy et al. 2008). In the Modern, where we can directly sample living assemblages of species, surprising regularities are found in the distributions of population sizes, resource uses and trophic interactions across species (J. H. Brown 1995; Hubbell 2001; Nekola and Brown 2007; J. Harte 2011; Jordano, Bascompte, and Olesen 2003; Nuismer, Jordano, and Bascompte 2013; Berlow et al. 2009). Thus to predict biodiversity we must embrace both the extreme variability and contingency of its history as well as the persistent emergent similarities in macroscopic patterns across systems.

Statistical mechanics provides a powerful framework for predicting macroscopic generalities in systems composed of many constituent, potentially complex, parts (e.g. Albert and Barabási 2002; Sethna 2006). In this thesis I use the predictive power of statistical mechanics to understand the origin of pervasive patterns in biodiversity, such as the distribution of its fluctuations through time and the distribution of trophic interactions across species. I combine this approach with an explicit consideration of the evolutionary process leading to these biodiversity patterns, seeking to understand how biological evolution drives biological systems away from the steady state idealizations predicted by statistical mechanics. Thus I show using examples from the fossil record and rapidly evolving island biotas that general patterns in biodiversity emerge from a combination of the statistical mechanics of large systems and the unique non-equilibrium dynamics imparted to biological systems by their evolutionary history.

Biodiversity theory seeks to predict observed assemblages of species (Hubbell 2001). Historically, this pursuit has been dominated by “bottom-up” theories that attempt to account for all possible mechanistic interactions between species and their environments (Hubbell 2001; Haegeman and Loreau 2008; J. Harte 2011), which invariable depend on the specifics of the systems for which they are developed and cannot be generally applied across the breadth of biodiversity. The pursuit of simple universal theory was ignited by the work of MacArthur and Wilson (1967) whose quantitative description of island assemblages as a dynamic equilibrium of colonization and extinction paved the way for more nuanced theories such as the unified neutral theory of biodiversity (Hubbell 2001), predicting the species abundance distribution of any given local community based on the stationary distribution of a birth-death-immigration-speciation process.

Surprisingly, the patterns these theories seek to predict do not appear to be unique to biology (Nekola and Brown 2007), implying that the mechanisms leading to these outcomes should not be uniquely biological. Indeed, there is historical precedence in ecology to consider macroscopic properties of assemblages, such as the species abundance distribution, as the outcome of general statistical laws (Fisher, Corbet, and Williams 1943; Preston 1950; Preston 1962a; Preston 1962b). This “non-mechanistic” line of thinking has culminated in purely statistical theories of biodiversity, such as the maximum entropy theory of ecology (J. Harte 2011). Such theories parallel the coarse-graining of statistical mechanics, where the details of how micro-states interact (e.g. particles in the case of theoretical physics or populations in the case of biology) are averaged over in order to predict the behavior of the system as a whole.

The application of coarse-graining in ecology has met with considerable empirical success (e.g. Banavar and Maritan 2007; Pueyo, He, and Zillio 2007; Dewar and Porté 2008; J. Harte 2011), yet biology differs fundamentally from statistical physics because of the non-Markovian memory imparted to species and populations due to common inheritance of genomes and niche-constructed environments (Odling-Smee, Laland, and Feldman 2003) that are the result of natural selection. If biological systems are in steady state they have the chance to overcome the effect of evolutionary contingency, but far from equilibrium the importance of evolutionary history should be clearly seen in deviations between observed biodiversity patterns and statistical idealizations thereof.

Here I use two complementary approaches to understand the importance of evolutionary history through the lens of statistical mechanics. In the first I develop a dynamic non-equilibrium theory of changes in biodiversity by extending a branch of non-equilibrium statistical physics, known as super-statistics (Beck and Cohen 2003; Beck 2004), to demonstrate that the complex dynamic of diversity fluctuations through time arrises from the same non-equilibrium processes that can describe complexity in large physical (Beck 2004) and social systems (Fuentes 2009). In the second approach I test a static equilibrial theory of biodiversity (the maximum entropy theory of ecology J. Harte 2011) across rapidly evolving arthropod communities of different ages in the Hawaiian archipelago. This work combines population genetics with the geologic chronosequence to understand how the rate and mode (e.g. through adaptive differentiation) of evolutionary community assembly drives deviations from statistical steady sate. I develop open source software, available as an R package, implementing these statistical mechanical approaches to biodiversity theory to showcase the computation required to achieve these insights and to make that computation more accessible to other biodiversity scientists. Broadly testing this new form of theory will be critical for ultimately evaluating its utility in predicting and understanding the spectacular diversity of life that is so critical to the health of our planet and humanity.

# Punctuated non-equilibrium and niche conservatism explain biodiversity fluctuations through the Phanerozoic

Andrew J. Rominger, Miguel A. Fuentes and Pablo A. Marquet

In revision for *Proceedings of the Royal Society of London B*

**Abstract**

Fluctuations in biodiversity, both large and small, are pervasive through the fossil record, yet we do not understand the processes generating them. Here we use a novel extension of theory from non-equilibrium statistical physics to show that three universal properties of macroevolution, punctuated adaptive radiation, niche conservatism and resultant heterogeneity of diversification rates between taxa, are sufficient to explain previously unaccounted for biodiversity patterns throughout the Phanerozoic. Using this theory, known as super-statistics, we identify taxonomic orders as largely autonomous evolutionary units, each likely experiencing its own unique and conserved adaptive landscape. This indicates that while within-order diversification could be adequately explained by neutral processes, between order diversification is likely driven by major evolutionary innovations. Super-statistics, has been successfully used to explain other complex systems including driven turbulent flows and wild stock market fluctuations. Compared to other approaches that have used simple birth-death processes, equilibrial dynamics or non-linear theories from complexity science, super-statistics is superior in its ability to account for both small and extreme fluctuations in fossil diversity. Its success opens up new research directions to better understand the evolutionary processes leading adaptive landscapes to be conserved within orders and undergo punctuated innovations between orders.

## Background

Biodiversity has not remained constant nor followed a simple trajectory through geologic time Raup, Sepkoski Jr, and others (1982; Sepkoski 1984; Gilinsky 1994; Liow and Stenseth 2007; Alroy et al. 2008; Alroy 2010). Instead, it has been marked by fluctuations in the number of extant taxa, both positive in the case of net origination or negative in the case of net extinction. Major events, such as adaptive radiations and mass extinctions have received special attention (Benton 1995; Erwin 1998), but fluctuations of all sizes are ubiquitous (Sepkoski 1984; Alroy et al. 2008; Quental and Marshall 2013).

Several approaches have been taken to study the complex trajectory of paleo-biodiversity ranging from the hypothesis that biological systems self-organize to the brink of critical phase-transitions (Bak and Sneppen 1993; Solé et al. 1997) to invocations of non-linear environmental perturbations (Newman and Roberts 1995) and escalatory co-evolutionary interactions (Vermeij 1987). New data and analyses have not supported any of these hypotheses at the scale of the entire Phanerozoic marine invertebrate fauna (Kirchner and Weil 1998; Madin et al. 2006; Alroy et al. 2008). Other studies have modeled the mean trend in diversity as tracking a potentially evolving equilibrium (Sepkoski 1984; Alroy et al. 2008; Alroy 2010; Rabosky 2009) and yet ignore the potential role of stochasticity and non-equilibrium dynamics in producing observed patterns (Erwin 2012; Liow and Stenseth 2007; Quental and Marshall 2013). As such, we still lack a synthetic theory of evolving biodiversity through the fossil record. Here we use a simple model of evolution in an abstract niche space derived from universal non-equilibrium processes to predict, with great accuracy, the complex distribution of pervasive diversity fluctuations throughout the marine Phanerozoic.

Despite the heterogeneity of explanations of Phanerozoic biodiversity, consensus has emerged on three properties of macroevolution:

[*i)]*

gross ecological and life history attributes of clades (i.e. groups of related species descending from a common ancestor) are often maintained, a phenomenon known as niche conservatism (Roy et al. 2009; Hopkins, Simpson, and Kiessling 2014);

long periods of niche conservatism are interrupted by adaptive diversification and exploration of new ecological niche space (Eldredge and Gould 1972; Newman, Cohen, and Kipnis 1985; Hopkins, Simpson, and Kiessling 2014); and

as a consequence of the interaction between their life history characteristics and the dynamics of the environments they inhabit (Vrba 1983) different clades experience different rates of morphological evolution, speciation and extinction (Simpson 1953; Sepkoski 1984; Holman 1989; Gilinsky 1994).

Observed bursts of adaptive radiation leading to novel morphologies in the fossil record led Eldredge and Gould to their hypothesis of punctuated equilibrium (Eldredge and Gould 1972). Here we show that this punctuation is actually akin to the “super statistical theory” of non-equilibrium dynamics in statistical physics (Beck and Cohen 2003). Super-statistics (Beck and Cohen 2003) proposes that non-equilibrial systems can be decomposed into locally equilibrial sub-systems. The distribution of equilibria across sub-systems determines the dynamics of the complete system (Beck and Cohen 2003). When these sub-systems are superimposed the resulting system can no longer be described by a single equilibrial model. In the context of macroevolution we propose that a clade with conserved life history characteristics corresponds to a locally equilibrial sub-system. If a certain region of niche space can only contain a finite diversity of taxa (Simpson 1953; Gavrilets and Vose 2005; Rabosky 2009; Price et al. 2014) then diversity within clades should fluctuate stochastically about this equilibrium due to random origination and extinction. The magnitude of these macroevolutionary rates should be a function of the life history and ecological characteristics that define that region of niche space. Larval type (Jablonski 2008), body plan (Erwin 2012), body size (Harnik 2011), range size (Harnik 2011; Foote et al. 2008) and substrate preference (Hopkins, Simpson, and Kiessling 2014) have all been shown to influence such rates. Thus different regions of niche space, and the clades occupying them, will experience different magnitudes of stochastic fluctuation in diversity. As clades occasionally split to fill new regions of niche space their punctuated diversification determines the non-equilibrium nature of the entire biota. Here we show that these properties of macroevolution are sufficient to explain the complex fluctuations of marine invertebrate diversity through the Phanerozoic.

In statistical mechanics, local sub-systems can be defined by a simple statistical parameter often corresponding to inverse temperature. In the context of macroevolution we define as the inverse variance of a homogenous origination-extinction process, which will capture all relevant information about a clade’s diversification under such a process. The limit distribution of time averaged fluctuations in clade ’s diversity through time should be approximately Gaussian with variance (Appendix [sec:suppLimitDist]). We posit that just as in statistical systems, non-equilibrium dynamics can arise from the mixing of the dynamics of many locally equilibrial subsystems. For marine Phanerozoic diversity this corresponds to mixing the dynamics of many clades, all being described by their unique values. Three exemplar dynamics taken from a bias-corrected (see methods section) aggregation of the Paleobiology Database (PBDB) (Alroy et al. 2008) are shown in Figure [fig:pkf]. To predict the super-statistical behavior of the entire marine invertebrate Phanerozoic fauna we must integrate over all possible local equilibria that each clade could experience. The distribution of values describes the probability that a given clade, chosen at random, will occupy a region of niche space characterized by that inverse variance value. The form of this stationary distribution of values could shed interesting light on the biological processes that lead different clades to different equilibria, as discussed below. Figure [fig:pkf] shows the shape of this stationary distribution estimated from bias-corrected PBDB (Alroy et al. 2008) data.

To uncover the super-statistical nature of the marine invertebrate Phanerozoic fauna we analyze the distribution of diversity fluctuations using two canonical databased of fossil biodiversity, the PBDB (Alroy et al. 2008) and Sepkoski’s compendium (Sepkoski 1992) of fossil marine invertebrates (results from Sepkoski’s compendium are presented in Appendix [sec:suppSepk]). We filter PBDB data to include only well preserved marine invertebrates following previously published collection inclusion criteria (Alroy et al. 2008; Alroy 2010). We account for detection bias in the PBDB using an extension of the “three timer” correction (Alroy et al. 2008). “Three timer” correction accounts for the rate of failure to observe a genus, estimated by the number of times a gap occurs in its occurrence history. We extend this correction by also employing a new publication bias correction to help eliminate bias from preferential publication of novel taxa (see methods section). Results obtained from this correction strategy are similar to other published methods (Fig. [fig:supp3TPub]). Fluctuations within a clade are computed as the difference in standing diversity between two time intervals, or equivalently the number of originations minus extinctions in one interval.

Phanerozoic biodiversity can be deconstructed and grouped into clades, or sub-systems, in several different ways. Lacking a full phylogenetic hypothesis for all marine invertebrates we use taxonomic classifications to identify potential sub-systems. Taxa ideally represent groups of organisms that descend from a common ancestor and share similar ecologically and evolutionary relevant morphological traits (Mayr 1965; Erwin 2007). For Phanerozoic marine invertebrates, Holman (Holman 1989) has shown that variance in diversity dynamics is less between taxa belonging to the same order than taxa in different orders, indicating that the taxonomic level of orders is a likely candidate for sub-system delineation. To evaluate the optimal taxonomic level for sub-system designation, we test our superstatistical theory using taxonomic levels from order through class to phylum. Additionally we compare our results to randomized taxonomies to confirm that observed patterns are not an artifact of arbitrary classification but instead represent real biologically relevant differences between clades.

## Methods

### Paleobiology Database data download and filtering

Data were downloaded from the Paleobiology Database (PBDB; www.pbdb.org) on 28 May 2013. Collections were filtered using the same approach as Alroy (Alroy et al. 2008) to insure that only well preserved marine invertebrate occurrences were used in subsequent analysis resulting in 221202 genus occurrences. These were further filtered to exclude those occurrences without order-level taxonomy and those collections with age estimate resolutions outside the 10my default bins of the PBDB resulting in 189516 occurrences left for analysis. To avoid basic sampling concerns we excluded the earliest Cambrian and the Cenozoic.

To focus attention on the variance of fluctuations we center each clade’s fluctuation distribution. Because “equilibrium” in the statistical mechanical sense means a system undergoes coherent, concerted responses to perturbation the mean trend line is of less interest than deviations from it. We also note that most clades are already close to centered and so centering has little influence on clade-specific fluctuation distributions.

### Three-timer and publication bias correction

We use a new and flexible method, described below, to correct for known sampling biases in publication-based specimen databases (Alroy et al. 2008; Alroy 2010). We were motivated to use this method because rarefaction has been shown to under-perform compared to the more data-intensive shareholder quorum subsampling (SQS) method (Alroy 2010). However, subsampling cannot be applied to small orders (i.e. the majority) because SQS becomes increasingly unreliable as sample size decreases (Alroy 2010). We therefore develop a simple method based on first correcting for detection bias using the “three timer” correction (Alroy et al. 2008) in which the rate of failure to observe a genus is estimated by the number of times a gap occurs in the occurrence history of each genus. To eliminate further bias due to preferential publication of novel taxa we divide observed number of genera per order per time period by the expected number of genera given publications in that time period. The expected number is calculated by regressing the log-transformed number of genera on log-transformed number of publications. There is a weak trend toward higher diversity with more publications (Fig. [fig:divByPub]) meaning that the most important correction comes from the three timer correction.

Our new method effectively re-scales each genus occurrence from 0/1 (absent/present), to a weighted number continuously ranging between 0 and 1. This method achieves similar results to more computationally intensive sub-sampling procedures (Alroy et al. 2008; Alroy 2010). We directly compare our predicted time series of global genus diversity with results derived from SQS (Alroy 2010) and the raw data (Fig. [fig:supp3TPub]). Our method shows minor differences with the SQS prediction, However, these discrepancies do not have impact the distribution of fluctuations (Fig. [fig:supp3TPub]) and super-statistical analysis on uncorrected PBDB data (see section [sec:rawPBDB]) produces a similar result to the analysis on corrected PBDB data presented in the main text.

### Numerical methods

To fit our super-statistical prediction we use the method of least squares instead and maximum likelihood. When building the prediction for by calculating order-level Gaussian distributions and integrating over them, we use least squares to fit the variance term to each order. We do so because orders potentially show asymmetries in their distribution of fluctuations. Least squares is more flexible in fitting such distributions compared to maximum likelihood which will always estimate the empirical variance as the best-fitting parameters.

We also estimate directly from the raw data using maximum likelihood to compare the fit of our super-statistical prediction and that of a simple Gaussian distribution using AIC. To calculate a likelihood-based confidence interval on our prediction we bootstrapped the data, subsampling fluctuations with replacement from all orders combined.

## Results and Discussion

At the order level in both the sampling-corrected PBDB (Fig. [fig:pkf]) and Sepkoski’s compendium (Appendix [sec:suppSepk]), fluctuations in genus diversity are well described by a Gaussian distribution (Fig. [fig:pkf]). Gaussian fluctuations would result from a homogeneous origination-extinction process under the condition of independence between orders. Independence could result from neutral-like processes (Hubbell 2001), where the dynamics of one taxon are unaffected by those of another, or from dampening mechanisms that stabilize complex networks of interacting taxa (Brose, Berlow, and Martinez 2005). This is in direct contrast to the instability hypothesis underlying the self-organized criticality theory of paleo-biodiversity (Bak and Sneppen 1993; Solé et al. 1997).

We estimate the distribution of ’s simply as the maximum likelihood distribution describing the set of inverse variances for all orders. In both the PBDB and Sepkoski’s compendium, Phanerozoic marine invertebrate orders clearly follow a Gamma distribution in their values (Fig. [fig:pkf]). While multiple processes could lead to a Gamma stationary distribution (e.g. (Cox, Ingersoll Jr, and Ross 1985)), one interesting possibility is a mean-reversion process (Cox, Ingersoll Jr, and Ross 1985). Mean reversion could be a consequence of niche conservatism if values are associated with a clade’s physiological and life history traits, themselves evolving via Ornstein-Uhlenbeck-like exploration of an adaptive landscape (Cox, Ingersoll Jr, and Ross 1985; Butler and King 2004).

Using the observation of within order statistical equilibrium and Gamma-distributed parameters we can calculate, without further adjusting free parameters, the distributions of order-level fluctuations for the entire marine Phanerozoic, , as

where is the distribution of fluctuations within an order and is the stationary distribution of inverse variance in the magnitude of order-level fluctuations in diversity. This leads to a non-Gaussian, fat-tailed prediction for which matches both the PBDB and Sepkoski data closely (Fig. [fig:Px] and Appendix [sec:suppSepk]).

To quantitatively evaluate how well the super-statistical prediction matches the data we constructed a 95% confidence envelope from bootstrapped maximum likelihood estimates for . Observed fluctuations fall within this 95% confidence envelope (Fig. [fig:Px]), indicating that the data do not reject the super-statistical prediction. For further comparison, we fit a Gaussian distribution to the observed fluctuations, which corresponds to the equilibrium hypothesis that all orders conform to the same statistics. Using Akaike Information Criterion (AIC) we find that observed fluctuations are considerably better explained by the super-statistical prediction than by the Gaussian hypothesis (AIC = 11285.18). Thus, as expected under the super-statitical hypothesis, the fat tailed distribution of fluctuations arise from the superposition of independent normal statistics for fluctuations within orders.

Computing the distribution of fluctuations using classes instead of orders leads to a substantially poorer fit to the observed data (Fig. [fig:suppPBDBPxcls]). We quantify this shift with the Kolmogorov-Smirnov statistic, which changes from 0.041 in order to 0.062 in classes (Fig. [fig:dStat]). However, if super-statistical theory explains the data, this worsening fit with increasing taxonomic scale is expected as the different classes are not well defined subsystems in their fluctuation dynamics. Instead, classes aggregate increasingly disparate groups of organisms, and thus effectively mix their associated Gaussian fluctuations, meaning that one statistic should no longer be sufficient to describe class-level dynamics. Our analysis indicates that orders are evolutionarily coherent and independent entities, with all subsumed taxa sharing key ecological and evolutionary attributes allowing them to diversify concertedly and independently from other orders. Both the good fit at the order level and worsening fit at higher taxonomic levels is confirmed in Sepkoski’s compendium, which also allows analysis of phylum-level patterns (Fig. [fig:suppsepkPx]).

To further test the evolutionary coherence of orders we conducted a permutation experiment in which genera were randomly reassigned to orders while maintaining the number of genera in each order. For each permutation, we calculated the super-statistical prediction and the Kolmogorov-Smirnov statistic. The permutation simulates a null model in which common evolutionary history is stripped away (genera are placed in random orders) but the total number of observed genera per order is held constant. Repeating this permutation 500 times yields a null distribution of Kolmogorov-Smirnov statistics that is far separated from the observed value (Fig. [fig:dStat]) suggesting the good fit at the order level is not merely a statistical artifact of classification but carries important biological information.

## Conclusion

Our analysis makes no assumption that orders should correspond to super-statistical subsystems, but identifies them as the appropriate level for marine invertebrates. Holman (Holman 1989) has also shown that orders are “evolutionarily coherent” in that subtaxa within orders share common diversification dynamics. As we show, orders differ only in the variances of their diversity fluctuations (Fig. [fig:pkf]).

Our study is the first to demonstrate that complex patterns in the sequence of origination and extinction events in the fossil record are the result of a simple underlying process analogous to the statistical mechanisms by which complexity emerges in large physical (Beck 2004) and social systems (Fuentes 2009). We do so by identifying the biological scale at which clades conform to equilibrial dynamics, which could result from the process of niche conservatism. We then show that punctuated shifts to different equilibria between clades, a consequence of punctuated exploration of niche space by newly evolving clades, leads to a characteristically non-equilibrial distribution of diversity fluctuations when the marine Phanerozoic fauna is viewed as a whole macro-system.

Our work highlights the importance of both niche conservatism and punctuated adaptive radiation in producing the statistical behavior of the Phanerozoic; our theory thus provides new motivation for identifying the eco-evolutionary causes of innovations between lineages and how those innovations are eventually conserved within lineages. Armed with an understanding of the statistical behavior of diversification we can go on to examine mechanisms underlying additional patterns in the mean trend of biodiversity through the Phanerozoic. In particular, clades have been shown to wax and wane systematically through time (Liow and Stenseth 2007; Quental and Marshall 2013), a pattern that we cannot explain with super-statistics alone.

Super-statistics could also be applied to other areas of evolution and macroecology. For example new phylogenetic models already consider heterogeneous rates of diversification (e.g. Rabosky 2006). The super-statistics of clades in adaptive landscapes could provide a means to build efficient models that jointly predict morphological change and diversification. This framework could also provide a new paradigm in modeling the distributions of diversity, abundance and resource use in non-neutral communities. Non-neutral models in ecology are criticized for their over-parameterization (Rosindell, Hubbell, and Etienne 2011), yet a persistent counter argument to neutral theory (Hubbell 2001) is the unrealistic assumption of ecological equivalency (Chave 2004) and poor prediction of real dynamics (Ricklefs 2006). If ecosystems are viewed as the super-position of many individualistically evolving clades, each exploiting the environment differently and thus obeying a different set of non-equivalent statistics, then diversity dynamics could be parsimoniously predicted while incorporating real biological information on ecological differences between taxa.

## Acknowledgments

We thank Charles Marshall, Joseph Felsenstein, John Harte, Rosemary Gillespie and John Alroy for helpful discussion. Michael Foote provided a digitized copy of Sepkoski’s compendium. AJR thanks funding sources Fulbright Chile and the National Science Foundation Graduate Research Fellowship Program; MAF thanks FONDECYT 1140278; PM thanks support from Grant PFB-023 (CONICYT) and ICM-P05-002).

## Figures

![](data:application/pdf;base64,)

The distributions of within-order fluctuations in genus diversity shown for the trajectories of three exemplar orders (A) and shown as an empirical cumulative density aggregated across all orders (B). To display all orders simultaneously we simply collapse their fluctuation distributions by dividing by their standard deviations. If orders conform to the Gaussian hypothesis their scaled fluctuations should fall along the cumulative density line of a normal N(0, 1) distribution, as shown in (B). In (C) the distribution of inverse variances across all orders matches very closely to a Gamma distribution (black line); exemplar orders are again highlighted.

![](data:application/pdf;base64,)

Distribution of fluctuations in genus diversity within orders of marine invertebrates in the Paleobiology Database (Alroy et al. 2008) after bias correction. The distribution is fat-tailed as compared to the maximum likelihood estimate of the normal distribution (blue line). At the order level the empirical distribution of fluctuations are well described by our super-statistical approach, both when computed from integrating over the distribution of observed variances (red line) and when fit via maximum likelihood (95% confidence interval; red shading).

![](data:application/pdf;base64,)

Distribution of Kolmogorov-Smirnov (KS) statistics from randomly permuting genera within orders (gray shading represents 95% confidence interval). Solid black line is observed KS statistic at the order level, while the dashed black line shows the observed KS statistic at the class level.

# Community assembly on isolated islands: macroecology meets evolution

Andrew J. Rominger, Kari R. Goodman, Jun Y. Lim, Ellie E. Armstrong, Lisa E. Becking, Gordon M Bennett, Michael S. Brewer, Darko D. Cotoras, Curtis P. Ewing, John Harte, Neo D. Martinez, Patrick M. O’Grady, Diana M. Percy, Donald K. Price, George K. Roderick, Kerry L. Shaw, Fernanda S. Valdovinos, Daneil S. Gruner and Rosemary G. Gillespie

Published online in *Global Ecology and Biogeography*, August 2015

**Abstract**

* Understanding how ecological and evolutionary processes together deter- mine patterns of biodiversity remains a central aim in biology. Guided by ecological theory, we use data from multiple arthropod lineages across the Hawaiian archi- pelago to explore the interplay between ecological (population dynamics, dispersal, trophic interactions) and evolutionary (genetic structuring, adaptation, speciation, extinction) processes. Our goal is to show how communities develop from the dynamic feedbacks that operate at different temporal and spatial scales.
* The Hawaiian islands (19–22N, 155–160W).
* We synthesize genetic data from selected arthropods across the Hawai- ian archipelago to determine the relative role of dispersal and in situ differentiation across the island chronosequence. From four sites on three high islands with geological ages ranging from Ma to 5 Ma, we also generate ecological metrics on plant–herbivore bipartite networks drawn from the literature. We compare the structure of these networks with predictions derived from the principle of maximum information entropy.
* From the perspective of the island chronosequence we show that species at lower trophic levels develop population genetic structure at smaller temporal and spatial scales than species at higher trophic levels. Network nestedness decreases while modularity increases with habitat age. Single-island endemics exhibit more specialization than broadly distributed species, but both show the least specialization in communities on middle-aged substrates. Plant–herbivore networks also show the least deviation from theoretical predictions in middle-aged communities.
* The application of ecological theory to island chronosequences can illuminate feedbacks between ecological and evolutionary processes in community assembly. We show how patterns of population genetic structure, decreasing network nestedness, increasing network modularity and increased specialization shift from early assembly driven by immigration, to in situ diversification after Myr. Herbivore-plant communities only transiently achieve statistical steady state during assembly, presumably due to incomplete assembly from dispersal in the early stages, and the increasing influence of island ontogeny on older islands.

## Introduction

Current biodiversity is a product of speciation, extinction and dispersal, contingent on the ecological interactions of organisms with their biotic and abiotic environment. The evolutionary history leading to the assembly of any given ecological community must in some way shape current ecological assemblages. However, because the processes of evolution and ecology occur on different temporal and spatial scales, disentangling the relative influence of local ecological mechanisms from historical evolutionary processes on patterns of community structure remains a central challenge (Ricklefs 2004).

The evolutionary processes of speciation and extinction are classically viewed as constraints on regional species pools, occurring in a manner largely removed from local ecology (Hubbell 2001; Cavender-Bares et al. 2009; Wiens, Pyron, and Moen 2011). Conversely, ecological mechanisms tend to be viewed as packing standing diversity into local communities through consumption, competition, facilitation and, more recently, neutral ecological drift (Hubbell 2001; Tilman 2004; Bascompte and Jordano 2007; Borer et al. 2014). While recent theoretical advances have provided greater insight into ecological drift (Hubbell 2001; Rosindell and Phillimore 2011), niche partitioning (Tilman 2004), competition, predation (Borer et al. 2014) and species interaction networks (Williams and Martinez 2000; Brose, Williams, and Martinez 2006), these insights typically do not contain realistic evolutionary assumptions (Ricklefs 2006) or ignore them entirely.

Insights into the genetic, biogeographic and selective mechanisms leading to diversification have also emerged based on inference from current patterns of species, genetic or phylogenetic diversity (e.g. Wiens, Pyron, and Moen 2011; Jetz et al. 2012). However, it is not possible to use current static patterns to infer the temporal dynamics of either the evolutionary mechanisms or their ecological consequences, nor can we understand what constitutes meaningful change in a system without a baseline for comparison. Here we show how testing idealized ecological theories—such as the unified neutral theory (Hubbell 2001) or the maximum entropy theory of ecology (J. Harte 2011)—on archipelagos composed of islands formed in a discrete geological sequence can help identify the shifting balance and feedback between fast-acting, local ‘ecological’ mechanisms, and longterm, large-scale evolutionary processes in determining ecological community structure. Islands having different ages of formation, along with discrete volcanoes within islands, provide the opportunity to study diversification of species and the assembly of communities in different stages. Ecological theory provides an idealized ‘null’ baseline against which to compare observed patterns.

### Hotspot oceanic archipelagos as model systems

Hotspot oceanic islands are opportune model systems for studying the interplay of local ecological mechanisms and the evolutionary drivers of biodiversity patterns. Due to their sequential formation as the tectonic plate moves over a volcanic hotspot, such island systems offer a range of spatial and temporal scales over which to analyze the outcomes of ecological and evolutionary processes (Warren et al. 2015). While many archipelagos around the world share these biotic and geological properties, the Hawaiian archipelago provides a particularly useful system for study because its linear geological chronology (Price and Clague 2002), ecosystem developmental trajectories (Vitousek 2004) and phylogeographic patterns of biodiversity are each well characterized (Wagner and Funk 1995). Moreover, studies of species diversity across the islands have revealed patterns that are non-uniform across the island chronosequence with marked differences among lineages (e.g. Gruner 2007; Gillespie 2009) that can be used to test for biologically meaningful differences among lineages that might drive their disparate diversification patterns.

### Development of genetic structure

High levels of dispersal and associated gene flow among localities limit the extent to which populations can diverge genetically. However, when gene flow is low, distinct populations in different localities are free to diverge through local selective pressures and drift, which can lead to diversification (Slatkin 1987) Thus, the magnitude of genetic connectivity among populations provides a measure of the relative importance of dispersal-driven assembly (dictated by processes removed from the local setting) in contrast to assembly by local (*in situ*) diversification in determining community composition. Using the chronosequence of the Hawaiian archipelago, we can analyze populations from multiple sets of taxa across trophic guilds occurring in geological contexts from young to old. We predict that dispersal-driven (ecological) processes will dominate in community assembly in young habitats, with the importance of in situ (evolutionary) processes increasing with habitat age. If evolutionary processes are not important, we predict that communities should reach a statistical steady state through ecological processes alone (J. Harte 2011). If, as we expect, evolutionary processes become increasingly important in community assembly over time, we would expect to find associated deviations from an ecological null model of community assembly, provided by idealized ecological theory. Differences in population structure among taxa or trophic groups could indicate whether sufficient time has passed along the chronosequence for the group of interest to experience significant evolutionary pressures.

### Macroecological metrics and idealized ecological theory

By their nature, unified theories of biodiversity (e.g. Hubbell 2001; J. Harte 2011) provide a simplified view of ecology, but deviations from theory can provide insights into which particular ecological patterns require additional biological mechanisms for their explanation (J. Harte 2011). The maximum entropy theory of ecology (METE; J. Harte 2011) in particular provides predictions of species abundance distributions, species–area relationships and metabolic rate and network linkage distributions for idealized ecological communities in which the behavior of a system is governed by a simple set of state variables. The principle of maximum information entropy (MaxEnt), from which the METE is derived, is an established inference procedure that has yielded accurate predictions of diverse patterns in fields as varied as thermodynamics (Jaynes, 1957), economics (Golan, Judge, and Miller 1996), forensics (Roussev 2010), imaging technologies (Gull and Newton 1986) and, more recently, ecology (e.g. Phillips, Anderson, and Schapire 2006a; Dewar and Porté 2008; J. Harte 2011). MaxEnt works by seeking the least-biased prediction of a distribution of interest (e.g. the distribution molecular velocities in the case of thermodynamics or of species abundances in the case of ecology) while constraining that prediction to be consistent with state variables describing the macroscopic attributes of the system (e.g. temperature or the total number of species and individuals). These are the most ignorant possible predictions about the system. Thus, studying the unique ecological conditions and evolutionary histories of real-world systems that deviate from the conditions predicted from maximizing information entropy can provide insights into the processes driving ecological systems away from the statistical steady state (J. Harte 2011).

Ecological networks are complex systems forming hierarchical structures to which the principle of MaxEnt has recently been applied (Williams 2010; J. Harte 2011) and are a prime study focus because networks of interacting species embody both the ecology of trophic links and evolutionary processes such as co-evolution (Bascompte and Jordano 2007; Donatti et al. 2011; Nuismer, Jordano, and Bascompte 2013; Thompson et al. 2013). Thus they present an opportune starting place to study ecological and evolutionary feedbacks. The distribution of linkages in ecological networks can test whether plant–animal interaction networks assemble neutrally or through deterministic processes such as co-evolution of traits involved in foraging (Vázquez et al. 2005). Analysis of other network metrics such as modularity (the degree to which species interact in semi-autonomous modules) and nestedness (the degree of asymmetry in interaction between specialists and generalists) can further illuminate the underlying eco-evolutionary processes driving patterns of species interactions (Bascompte and Jordano 2007; Donatti et al. 2011; Nuismer, Jordano, and Bascompte 2013). In nested networks, species with fewer interactions (i.e. more specialized species) will interact with a subset of the species with which generalists interact. In this way interaction nestedness is mathematically equivalent to island nestedness (in which islands that are less species rich are subsets of islands that are more species rich). However, we only consider network nestedness here.

To gain insights into community assembly as it happens, we propose an integrative framework that harnesses advances in both evolutionary and ecological theory, placed in the context of age-structured archipelagos. Mechanistically simplified ecological theories such as the METE (J. Harte 2011) can be used as powerful null models; deviations from theoretical expectations can flag biological phenomena that warrant further study. Here we demonstrate how community-level data from age-structured island systems, combined with population genetic and phylogenetic data, can test the extent to which the evolutionary histories behind such communities drive their deviation from theoretical expectations. We provide an initial test of this concept using a synthesis of published data on arthropod lineages in the Hawaiian islands. We provide metrics of ecological and evolutionary dynamics across communities from settings that range in geological age from 500 years to 5 Ma. We estimate taxon-specific timelines for the development of population genetic structure for both herbivores and predators and couple these results with macroecological measures of community structure, using predictions from statistical steady-state and ecological network theory to provide insights into changes in community structure over the extended timeframe provided by the island chronosequence.

## Methods

### Dispersal-driven processes to in situ differentiation across the island chronosequence

To evaluate the balance between regional immigration and the potential for local differentiation, we measured how molecular variation is partitioned among populations within species across locations of known substrate age on the islands of Hawaii and Maui (Fig. [fig:map]). We compiled published [DNA sequences, amplified fragment length polymorphism (AFLPs) and allozymes] and new data sets for a diversity of native Hawaiian arthropod groups that represent a spectrum of trophic levels (Table1). New sequences were included for sap-feeding Hemiptera group *Nesosydne* planthoppers [COI; data generated following the protocols in Goodman, Welter, and Roderick (2012); GenBank accession numbers: KT023113–KT023179] and *Trioza psyllids* [COI, cytB; data generated following protocols in Percy (2003); GenBank accession numbers: KR108061–KR108144]. Samples were from the focal sites described below for the ecological analysis, as well as from other locations across Hawaii and Maui. These data were used to provide an estimate of how arthropod populations have accumulated genetic population structure within the focal sites of different geological age.

We used analysis of molecular variance (AMOVA) to examine how genetic variation is partitioned at two scales of population structure: among sites within volcanoes and among volcanoes on both the island of Hawaii and the islands of the Maui Nui complex (Maui, Molokai, Lanai). All analyses of allozyme and DNA sequence data were performed in Arlequin v.3.5 (Excoffier and Lischer 2010) using the AMOVA procedure to compute , a measure of genetic variance, or, where possible, , an analogue that incorporates genetic sequence information. The *Laupala* AFLP data were analyzed using tfpga v.1.3 (Miller 1997), using the same hierarchical approach of comparing within and among volcanoes as described above. To provide a temporal framework for the population differentiation analysis we assembled divergence-dating information from the literature for as many of the taxa as possible.

To explicitly test the association between landscape age and the potential for in situ genetic divergence we analyzed how within-site varies with the geological age of volcanoes on the islands of Hawaii and Maui Nui. For each volcano we calculated or (Excoffier and Lischer 2010) for each taxon among sites within volcanoes. This analysis assumes that volcano age parallels habitat age, allowing more or less time for the presence of the populations.

### Ecological metrics across the island chronosequence

To investigate how ecological patterns change as communities age, we selected four focal sites across the chronosequence and island ages (two on the island of Hawaii, one on Maui and one on Kauai; Fig. [fig:map]) of approximately 12 km (each was defined as a point with a 2 km radius buffer). Focal sites were selected to have similar forest composition (dominated by *Metrosideros polymorpha*; Myrtaceae), elevation (1100–1400 m) and rainfall (mean annual precipitation 2000–3000 mm). We then constructed bipartite interaction networks between native herbivorous Hemiptera species and native plants at each of the study sites. Bipartite networks describe the topology of ecological interactions between two guilds of organisms (e.g. herbivores and their plant hosts). Quantitative information on the relative importance of interaction links can be incorporated into network analyses (Vázquez et al. 2009). However, currently available data are restricted to binary networks: those that describe the potential for interaction between any two species but not the relative frequency of that interaction to each species.

We compiled species lists of all native herbivorous Hemiptera for each focal site from published species accounts. Species accounts and other published sources were used to determine the presence, probable presence, or probable absence of each species at each of our four focal sites. A documented presence was defined as a known specimen collected at the focal site; a probable presence was defined as a species whose abiotic tolerances and known geographic range overlap with a focal site but no known specimen exists confirming its presence. Probable absence was assumed when the criteria for presence or for probable presence are not met. Two sets of species lists for each focal site were compiled: a conservative data set composed of only documented presence occurrences and a less conservative data set that also included probable presences.

Host plants for each species of Hemiptera were determined from published species accounts. Data on host plant use at each specific site were not available so we assumed that if a known host plant were present at a site it would eventually be used. Host plant occurrence in the focal sites was determined using distribution models for 1158 species of Hawaiian plants (Price 2012). Each focal site was spatially joined in a geographic information system with all coincident plant distribution models that fell within its boundaries. Two sets of resulting focal site-specific networks were constructed: one using the conservative data set of Hemiptera species presences and the other using the less conservative data set.

We hypothesized that potentially complex evolutionary feedbacks contributing to community assembly should result in departures from the predicted ecological statistical steady state. We used the METE (Williams 2010; J. Harte 2011) to compute the statistical steady state for the distribution of the number of host plants used by each Hemiptera species (hereafter referred to as degree distribution). To evaluate how well the METE predicts the data we simulated METE-conforming communities having the same number of species and links as observed. We then calculated the log-likelihood of each simulated data set and compared the resultant distribution of log-likelihoods under the hypothesis that the METE is true with the observed log-likelihood. This comparison is identical in approach to a z-score test using a Monte Carlo simulation to estimate the sampling distribution of log-likelihoods. R scripts (v.3.1.1; R Development Core 2013a) used for METE estimation and Monte Carlo methods are available in Appendix [sec:meteCode]. To investigate how speciation may in part drive network patterns and deviations from those predicted by idealized ecological theory, we analyzed the number of links assigned to each Hemiptera species (the degree distribution) separately for single-island endemics (those species found on only one island and thus probably derived from in situ diversification) versus multi-island endemics (those species found on multiple islands). Although multiple processes can lead to a species being a single-island endemic (Whittaker, Triantis, and Ladle 2008), such taxa provide a proxy for how much speciation occurs within islands. To compare species degree distributions between single-island endemics and multi-island endemics across sites of different ages we conducted a generalized linear model with binomial error, treating site identity as a categorical predictor. Binomial errors effectively account for network size due to the bounded support of the binomial distribution.

To understand how other network properties change with age of the ecosystem substrate, we calculated two widely used descriptive network metrics across sites––nestedness and modularity. Nestedness describes the degree of asymmetry of species interactions connecting specialists and generalists (Bascompte and Jordano 2007; Ulrich, Almeida-Neto, and Gotelli 2009). We calculated nestedness using the NODF metric (Almeida-Neto et al. 2008) as implemented in the R package vegan (Oksanen et al. 2013) and modularity using a variety of algorithms implemented in the R package igraph (Csardi and Nepusz 2006). These metrics are not directly comparable across networks of different size and connectance (Ulrich, Almeida-Neto, and Gotelli 2009), so for each metric in each network we calculate z-scores using a null model that randomizes network structure while maintaining certain aggregate network properties (Ulrich et al., 2009). These z-scores are calculated as the difference between the observed network metric minus the mean of the null model divided by the null model standard deviation, or . Because z-scores can be highly sensitive to the choice of null model (Ulrich, Almeida-Neto, and Gotelli 2009) we implemented both a probabilistic null model (Bascompte and Jordano 2007) and a null model that strictly constrains the degree distributions of plants and herbivores (Ulrich, Almeida-Neto, and Gotelli 2009). The probabilistic null uses the frequency of interactions as the probability that a randomized link gets assigned to that cell in the interaction matrix (Bascompte and Jordano 2007); thus the probabilistic null constrains row and column sums in probability but not absolutely.

## Results

### Dispersal-driven processes to in situ differentiation across the island chronosequence

The AMOVA revealed significant genetic population structure from the smallest to the largest spatial scales examined, all within a very recent timeframe. For mitochondrial loci, statistically significant molecular variation partitioned among sites within volcanoes ranged from 0.037 to 0.92 and among volcanoes from 0 to 0.30. Corresponding variation at multilocus nuclear loci among sites within volcanoes ranged from 0.21 to 0.58 and among volcanoes from 0.04 to 0.34. Taxa in the lower trophic levels (herbivorous sap-feeding Hemiptera: planthoppers and psyllids) had as much or more molecular variation partitioned at the among-site, within-volcano level than the among-volcano level, while the predatory spiders were less structured at localities within volcanoes compared with among them (Table 1). The analysis of genetic population structure across the chronosequence of localities revealed a similar pattern. The herbivores show high genetic population structure among localities even on young volcanoes (Fig. [fig:popGen]). By contrast, predatory spiders exhibited little genetic population structure within sites on the same volcano; this was higher among volcanoes, with values increasing with age across the chronosequence.

The observed levels of genetic divergence have evolved rapidly in many cases. For example, for species from the island of Hawaii for which phylogenetic data provide divergence times, estimates of dates of species divergence range from 0.5–4 Ma, with additional within-species genetic divergence having developed subsequently (Table 1). That some of these estimates are older than the known age of the ‘Big Island’ suggests that genetic divergence pre-dates their colonization to Hawaii, or alternatively that estimates include sampling error. For the one species where population genetic data were used to estimate divergence times between populations, herbivorous *Nesosydne* planthoppers, it was determined that populations diverged as little as 2600 years ago (Goodman, Welter, and Roderick 2012 Table 1).

### Ecological metrics across the island chronosequence

The degree distribution of Hemiptera species varied across the chronosequence with both the youngest and oldest sites deviating most from the statistical steady-state maximum entropy predictions (Fig. [fig:degree]). In the intermediate-aged site of Kohala, deviations are not significantly different from the predictions of maximum entropy. The generalized linear model revealed significant differences between the degree distributions of single-island endemics (species whose distributions are restricted to only one island) versus archipelagic endemics that are found across multiple islands (Fig. [fig:degree]). Single-island endemics show significantly lower degree distributions overall (i.e. more specialization) compared with more generalist species found across multiple islands. Furthermore, single-island endemics use more host plant species on the intermediate-aged Maui site. The slightly younger Kohala shows increased generalization for both single-island endemics and archipelago endemics. However, when considering the degree distribution defined by trophic links to plant genera instead of plant species, the pattern of increased generalization holds for Kohala, but endemics on Maui no longer show a difference in their degree distributions from other island endemics. This change in pattern suggests that increased generality of Maui endemics may be driven by increased plant species diversity within genera on that island.

Network nestedness decreased with habitat age while modularity increased (Fig. [fig:netMet]). This trend was recovered in networks constructed from both more and less stringent geographic criteria (Fig. [figSupp:netCons]). Choice of null model changed the magnitude of modularity and the sign of nestedness z-scores; however, the relative pattern of decreasing nestedness and increasing modularity remained across the different null models used to standardize network metrics (Fig. [figSupp:netMetComp]). The patterns were also robust to sampling intensity, as demonstrated by a rarefaction analysis (Fig. [figSupp:rfy]).

## Discussion

### Development of genetic population structure at different trophic levels

The analysis of available genetic data presented here indicates that divergence is occurring within the islands at small spatial scales and over short time periods (Table [tab:genStat], Fig. [fig:popGen]). Furthermore, the scale of population structure varies with trophic position, with structure developing in sap-feeding herbivore lineages at smaller scales (and hence shorter timeframes in the context of the chronosequence) compared with detritivorous crickets and predatory spiders (Table [tab:genStat], Fig. [fig:popGen]). Structure within species may allow populations to take independent evolutionary trajectories, especially when aided by other evolutionary processes acting differentially across species geographic ranges. A variety of factors have been associated with the genetic divergence of populations and species in the lineages described here, including combinations of genetic drift associated with geographic isolation (Percy 2003; Mendelson and Shaw 2005; O’Grady et al. 2011; Goodman, Welter, and Roderick 2012), adaptation associated with competition, predation and mutualism (Gillespie 2004; Roderick and Percy 2008; Brewer et al. 2015) and sexual signaling (Mendelson and Shaw 2005; Percy and Kennedy 2006; Magnacca, Foote, and O’Grady 2008; Goodman et al. 2015).

The *Nesosydne* planthoppers provide evidence that some period of geographic isolation preceded the divergence of sexual signals (Goodman, Welter, and Roderick 2012; Goodman et al. 2015). Shifts in plant host use are also associated with diversification in this group (Roderick and Percy 2008). In a phylogenetic study of a radiation of sap-feeding *Nesophrosyne* (Cicadellidae) leafhoppers, species divergence was associated with host plant specialization between 1 and 5 Ma, but only with geography on the younger island (Bennett and O’Grady 2013). Our network analysis shows that specialization and modularity are more pronounced on Maui than on Hawaii (Figs. [fig:degree] and [fig:netMet]), consistent with the phylogenetic results from *Nesophrosyne*. Available dating analyses of other arthropod taxa indicate that population genetic structure can develop in much less than 1 Myr (Table 1), and suggest that landscape fragmentation processes (e.g. lava flows) may dominate the earliest stages of diversification across taxa in the Hawaiian islands. Other taxa at low trophic levels, such as the herbivorous *Trioza psyllids*, detritivorous *Laupala* crickets and fungivorous Drosophila, show similar signals of geographic isolation combined with ecological and sexual processes driving genetic divergence and diversification across sites as young as those on Hawaii (Percy 2003; Mendelson and Shaw 2005; Percy and Kennedy 2006; Magnacca, Foote, and O’Grady 2008; O’Grady et al. 2011). By contrast, spiders, which are predatory, develop genetic discontinuities at larger spatial and temporal scales with a strong signature of increasing structure with age of the chronosequence (Roderick et al. 2012 Table 1). Further work is needed to assess the generality of this pattern of slower genetic differentiation in predators compared with herbivores.

### Macroecological metrics: network structure and steady state

Across the Hawaiian archipelago, nestedness appears to decrease generally with site age, and is highest on the geologically youngest volcano, Kilauea. High nestedness on Kilauea may arise with high immigration of new species with high probabilities to eat or be eaten by the generalist species already present at the site (Bascompte and Jordano 2007). However, despite high nestedness on Kilauea, and thus the potential for neutral colonization-driven assembly, this site did not conform to the statistical steady-state predication of the METE. The observed deviations from the METE at Kilauea appear to be largely driven by a surplus of singleton links (Fig. [fig:degree]), which may reflect a state of ‘incomplete’ assembly, possibly by lower species richness of the plant and herbivore biotas. Conversely, at Kohala, at intermediate age (150 ka), observations were not significantly different from the METE predictions. We posit that the reason why theoretical predictions fit Kohala so well is that the site has had sufficient time to undergo ecological succession and thus arrive at a statistical steady state, but is still too young to be affected by ecological specialization and rapid in situ diversification associated with host plants on older islands.

Interestingly, the communities on the older Maui and Kauai sites show strong deviations from the METE expectations (Fig. [fig:netMet]). The METE is agnostic about which mechanisms determine the values of the state variables that lead to its macroecological predictions (Harte, 2011). It does not account for the evolutionary history of biological systems. Thus, one possible explanation for the strong deviations from the METE expectations, compared with observations at our intermediate-aged site (Kohala), is that while the ages of Maui and Kauai are sufficient for evolutionary assembly driven by specialization and diversification on host plants, the older age of these islands may have led to range contractions and possibly extinction of plant species on the oldest island of Kauai (Whittaker et al., 2008).

Our results show decreased nestedness and increased modularity on Maui and Kauai. Co-evolution between interacting species should lead to greater modularity (Donatti et al. 2011; Nuismer, Jordano, and Bascompte 2013). However, the influence of certain network properties, such as nestedness, on stability is still unknown, and so theoretical predictions of how network properties should change over evolutionary time, generally, are lacking. Theoretical and empirical studies have suggested that nestedness may or may not promote stability (Allesina and Tang 2012; Suweis et al. 2014). Furthermore, almost all studies of food webs have focused primarily on single or short ecological time spans of network development that do not span as much evolutionary time as is included here (e.g. Albrecht, Riesen, and Schmid 2010). Food webs are dynamic emergent entities, with broad topological characteristics that may change dramatically over time (e.g. Yeakel et al. 2013). To our knowledge, our study represents the first to evaluate network topology over larger temporal scales, and we argue that age-structured landscapes such as the Hawaiian archipelago are promising for resolving long-standing debates on the causes and consequences of network properties such as nestedness.

We found that single-island endemics were always more specialized than multiple-island endemics. Although dietary breadth has been positively associated with geographic range size (Lewinsohn, Novotny, and Basset 2005), the direction of causality is unclear (Slatyer, Hirst, and Sexton 2013): while dietary breadth may allow some species to colonize other islands, it may also be driven by adaptation to exploit locally abundant hosts across a large range. Nevertheless, both scenarios are consistent with the hypothesis that in situ formation of single-island endemics may be the product of co-evolution and specialization. At the Kohala site, which showed the best fit to maximum entropy theory, single-island endemic and multiple-island endemic species alike showed increased generalization (i.e. a higher degree, or more links; Fig [fig:degree]), while at the youngest site of Kilauea, specialist single-island endemics may be limited by low plant diversity and thus appear more specialized (Fig. [fig:degree]). Conversely at the oldest site on Kauai, where plant diversity is high (Kitayama and Mueller-Dombois 1995), single-island endemics are again associated with decreased degree and thus genuine specialization (Fig. [fig:degree]). On Maui, single-island endemics show statistically significant increases in generalization, but this pattern disappears when analyzing the data at the resolution of plant genera, thus suggesting that Hemiptera species endemic to Maui may benefit from the diversification of plant species within genera.

### Future research

The data and analyses presented here describing insect and plant communities across a chronosequence of habitats in Hawaii generate testable hypotheses concerning the relative importance of ecological and evolutionary processes in community assembly. Our work to date suggests the overarching hypothesis that ecological processes dominate community assembly in younger environments, with evolutionary processes becoming increasingly important as communities age. We can also make predictions about the sequence of community assembly based on proposed mechanisms.

In younger communities we predict characteristics of ecological assembly, with species resembling random samples through immigration from regional source pools. Thus, metrics describing these communities will approach expectations of an ecological statistical steady state. An exception will be communities that are still undergoing the initial stages of primary succession, which will change rapidly through time and represent nonrandom samples of source pools. We also predict that these communities will exhibit a nested network structure, assuming new species will eat or be eaten by the generalist species already present in the community, as suggested by previous work on nestedness (Bascompte and Jordano 2007) and by our finding that widespread species tend to be generalists (Fig. [fig:netMet]). Following the same logic, in older communities we expect to see characteristics of evolutionary assembly, dominated by processes such as adaptive exploration of niche space, giving way to speciation. Thus, we predict increasing specialization and modularity with time (Bascompte and Jordano 2007; Donatti et al. 2011; Nuismer, Jordano, and Bascompte 2013) as reflected by age across the chronosequence.

#### Ecological data: assembly of species into communities

In order to build a more rigorous understanding of the assembly process in both younger and older communities, fine-grained sampling of all macroscopic arthropod taxa is needed from a large number of sites across the island chronosequence. This will allow an assessment of changes in overall species composition and diversity across all players in the time-calibrated landscape (Gruner 2007). Such data will allow us to test entire arthropod communities for deviations from METE predictions of statistical steady state (J. Harte 2011) across substrates of different ages. For example, predators, whose assemblages are likely to be more dominated by immigration and ecological assembly (Fig. [fig:popGen]), may never show strong deviations from METE predictions, whereas herbivores could show increasing deviation with age in agreement with the network results of this paper (Fig. [fig:degree]).

#### Evolutionary data: diversification within species

The current study demonstrates that taxa from different trophic guilds differ in the scale at which differentiation occurs and highlights the importance of fragmentation of the landscape in facilitating differentiation. Future work will be aimed at gathering data for additional focal taxa within this system, spanning different trophic levels. We will use these data to understand taxonomic and functional differences in the rate of differentiation, to assess the roles of genetic fusion and fission and the spatial scale over which they are important in fostering diversification (Gillespie and Roderick 2014), and to detail the relative rates of speciation and extinction across the island chronosequence.

## Conclusions

We have shown how a chronosequence can be used to understand biodiversity dynamics across an ecological–evolutionary continuum. Focusing on entire communities of arthropods in the Hawaiian islands allows us to incorporate predictions from idealized ecological theories to understand eco-evolutionary feedbacks and generate predictions about how entire communities develop over an extended time. Such an approach may prove fruitful for investigating the separate and interactive roles of ecological and evolutionary drivers of community assembly using age-structured systems as a simplified natural experiment, as exemplified by oceanic archipelagos.

We have demonstrated how taxa in the lower trophic levels developed genetic structure even in the youngest habitats of the observed chronosequence and at smaller spatial scales (Table [tab:genStat], Fig. [fig:popGen]). Thus, lower trophic levels are affected by in situ processes of diversification very early in the chronosequence, compared with higher trophic levels, though in situ processes become more important over time in the latter. Network nestedness decreased while modularity increased with age (Fig. [fig:netMet]), again indicating a possible shift from assembly driven by ex situ immigration early on to one based on in situ diversification, such as in co-diversification of insect herbivores with host plants (Bascompte and Jordano 2007; Donatti et al. 2011). That single-island endemics (probably the product of in situ diversification) show more specialization at older sites than more broadly distributed species (those taxa more likely to be initial colonists; Fig. [fig:degree]) also supports this hypothesis.

This study provides a framework for using chronologically arranged oceanic island systems to examine the interplay between evolutionary and ecological processes in shaping biodiversity. Our initial results provide a clear hypothesis that ecological processes dominate community assembly in younger environments, with evolutionary processes becoming more important as communities age. We demonstrate how this approach can provide insights into the development of communities over ecological–evolutionary time, and the dynamic feedbacks involved in assembly.

## Acknowledgements

We are indebted to many scientists and land managers in Hawaii who have provided access to the lands: Pat Bily (The Nature Conservancy of Hawaii), Melissa Dean, Christian Giardina, and Tabetha Block (Hawaii Experimental Tropical Forests), Betsy Gagne (Natural Area Reserve System), Lisa Hadway and Joey Mello (Department of Forestry and Wildlife Hilo), Cynthia King and Charmian Dang (Department of Land and Natural Resources) and Rhonda Loh (Hawaii Volcanoes National Park). We thank Robert Ricklefs and two anonymous referees for thoughtful commentary. We are very grateful to Guida Santos, Richard Field and Robert Ricklefs for inviting us to contribute to this Special Issue. The research was supported by the National Science Foundation DEB 1241253.

## Figures

![](data:application/pdf;base64,)

Map of substrate age (millions of years, My) of the islands of Kauai, Maui and Hawaii. Colours correspond to substrate age from young (light) to old (dark). Focal sites are shown as black circles (on Hawaii, Kohala is in the north, Kilauea in the south) while sampling sites for genetic data are represented by grey circles.

![](data:application/pdf;base64,)

Population genetic structure ( for all taxa except *Laupala* for which we used ) among sites within volcanoes with volcano age for insects and spiders. Calculations were based on mitochondrial DNA only (see Table 1 for details). The plant-feeding groups, specifically the sap-feeding Hemiptera, show higher genetic structure among sites on young volcanoes relative to older volcanoes, whereas detritivores (crickets), fungivores (*Drosophila*) and in particular predators (spiders) show little structure on young volcanoes. For spiders, substantial structure develops only later in the chronosequence, for example on Maui at approximately 1 Ma. Numbers refer to different species: 1, *Nesosydne chambersi*; 2, *Nesosydne raillardiae*; 3, *Nesosydne bridwelli*; 4, *Trioza* HB; 5, *Trioza* HC; 6, *Drosophila sproati*; 7, *Laupala cerasina*; 8, *Tetragnatha anuenue*; 9, *Tetragnatha brevignatha*; 10, *Tetragnatha quasimodo*; 11, *Theridion grallator*.

![](data:application/pdf;base64,)

Patterns in degree distributions across sites, comparing archipelago-wide endemics (cosmopolitans) with single-island endemic (Endemics) taxa. The top panels show that networks deviate most from the predictions of the maximum entropy theory of ecology on the youngest and oldest sites. Inset figures show the distribution of simulated mean squared errors; if the vertical/red line falls within the grey region (95% confidence interval) the data are not significantly different from the predictions of maximum entropy theory. All sites except Kohala deviate from the predications. The bottom panel shows the number of links for endemics versus cosmopolitans. Endemics show lower linkage overall, but significantly increase on the intermediate-aged site Maui (highlighted with dotted box). Kohala shows increased linkage overall (highlighted with a solid box).

![](data:application/pdf;base64,)

Trends in network metric nestedness and modularity through time. Nestedness decreases while modularity increases. Error bars represent 95% confidence intervals from a null model simulation.

## Tables

[!ht] ![](data:application/pdf;base64,)

# meteR: An R package for testing the Maximum Entropy Theory of Ecology

Andrew J. Rominger and Cory Merow

In revision for *Methods in Ecology and Evolution*

**Abstract**

1. Macroecological patterns appear to follow consistent forms across a range of natural systems, however the origin of their regularity remains obscured. The Maximum Entropy Theory of Ecology (METE) predicts macroecological patterns of abundance, metabolic rates and their distribution within communities and across space using an information theoretic approach. METE’s success in predicting empirical patterns demands that we further press the theory’s predictions to determine how (or whether) predictability depends on attributes of the system and the temporal, spatial and biological scales at which we study it.
2. METE predicts multiple macroecolgical metrics using statistical idealizations from information theory; thus confronting METE with data represents a strong test of the underlying biological mechanisms that could drive real communities away from statistical idealizations. METE has remained somewhat inaccessible due to its highly mathematical nature and a lack of software for model construction/evaluation. To remedy this, we have developed an R package implementation of METE.
3. Our open source (GNU General Public License v2) R package, meteR (version 1.0; cran.r-project.org/web/packages/meteR), (1) directly calculates all of METE’s predictions from a variety of data formats; (2) automatically handles approximations and other technical details; (3) provides high-level plotting and model comparison functions to explore and interrogate models.
4. With these tools in hand, ecologists can more readily test the predictions of METE for their data sets. By facilitating tests of METE, we expect that a better understanding of its strengths and limitations will emerge. A better understanding of the strengths and limitations of METE will offer insight into how biological mechanisms and statistical constraints combine to drive macroecological patterns.

## Background on the Maximum Entropy Theory of Ecology

Macroecology (J.H. Brown 1995) seeks to predict patterns in the distribution of individuals within species, across body sizes, and over space. These patterns can vary with spatial, temporal and taxonomic scale which makes their regularities challenging to detect. Macroecological patterns can be described quantitatively in the form of well-defined metrics such as species abundance distributions (SAD), body size distributions (e.g. distribution of metabolic rates [=power] across individuals, or the individual power distributions; IPD), and species-area relationships (SAR). Macroecological theory attempts to predict the mathematical form of these metrics from combinations of explicit mechanisms and statistical assumptions. Harte and colleagues have developed a Maximum Entropy Theory of Ecology (METE) grounded in information theory, which predicts the form of most of the macroecology metrics found in the literature, needing very limited empirical data as input, and with no adjustable parameters (Harte et al. 2008; Harte, Smith, and Storch 2009; John Harte 2011).

Maximizing information entropy (MaxEnt) is a general inference procedure for solving a class of problems involving inference of the least biased mathematical form of a probability distribution (e.g., the species abundance distribution) given some prior knowledge that can be expressed in the form of constraints on that distribution (e.g., the mean abundance across all species). One special case of MaxEnt familiar to many ecologists is its use in machine learning and species distribution modeling (SDM; Phillips, Anderson, and Schapire 2006b), though there are a variety of applications in ecology (Shipley, Vile, and Garnier 2006; R. Williams 2010) and other sciences (e.g. Skilling 1988; Jaynes 2003; Banavar, Maritan, and Volkov 2010). The conceptual goal of the MaxEnt approach in macroecology is to build predictions that are not sensitive to potentially arbitrary model parameter choices and represent a statistical idealization of a community that is near steady state. The solution to the MaxEnt problem entails finding the form of a distribution, (e.g. the SAD), that maximizes the information entropy function: under the constraints imposed by prior knowledge on (e.g. mean abundance). Maximization is carried out by the method of Lagrange multipliers. The resulting distribution is the function that is as smooth as possible given the constraints, and thus reflects no information other than the prior constraints (Jaynes 1957; Jaynes 1982).

The development of METE parallels the method used to derive the laws of classical equilibrium statistical mechanics and thermodynamics by Jaynes (1957), in which constraints arise from knowledge of the state variables of the system: for example, the total energy, volume, and number of molecules in a container of gas. In METE, the state variables used to predict the metrics of macroecology are the area, , of an ecosystem at any spatial scale at which census data exist, the total number of species , censused in that area, the total number of individuals, , across the species, and the total rate of metabolic energy use, , by the individuals. With the constraints that arise from the ratios of these state variables, the maximum information entropy condition predicts the mathematical forms of macroecological metrics (Table [tab:metrics]). Two distributions are at the core of the theory. The first is a joint distribution (the *Ecosystem Structure Function*; ESF) over abundance, , and metabolic rate, : . is the probability that a species randomly selected from the census has abundance , and an individual randomly selected from that species has metabolic energy requirement in the interval . The second core distribution is the *Spatial Structure Function* (SSF; mirroring the ESF). If a species has individuals in an area , then is the probability that it has individuals in an area within . From these two core distributions many of the metrics commonly studied in macroecology can be derived: for example the species abundance distribution arrises by integrating over and the distribution of metabolic rates (or body size under metabolic scaling theory (Brown et al. 2004) arrises by summing over . The species area relationship can be derived by combining the species abundance distribution with at multiple scales. These derivations are detailed in John Harte (2011) and Table [tab:metrics] provides a summary.

### Stronger tests of METE

Tests of METE have been successful in a wide range of systems (Harte et al. 2008; Harte, Smith, and Storch 2009; John Harte 2011; White, Thibault, and Xiao 2012; Newman et al. 2014), however a great deal of further testing is needed to establish METE’s generality, strengths and weaknesses. Existing tests have focused on SADs or SARs while tests of metabolic rate distributions are rare (but see John Harte 2011; Newman et al. 2014; Xiao, McGlinn, and White 2015). However, the most illuminating cases are perhaps those where METE fails (John Harte 2011), as these will demonstrate what attributes of a community drive it away from the simple steady state into alternate, more complex or transient states. Additionally, macroecological predictions other than the SAD and SAR have received relatively little attention, and their generality and correlates of their successes and failures are unknown. Two categories of failed predictions stand out. The first was originally considered a success: the theory predicts an inverse relationship between metabolic rate (body size) and abundance, called ‘energy equivalence’ or the Damuth rule (Damuth 1981). Yet considerable data analysis reveals numerous exceptions to this prediction (White et al. 2007). Second, and more fundamentally, the theory fails to accurately predict empirical patterns in ecosystems undergoing relatively rapid change (John Harte 2011; Newman et al. 2014). Examples are rapidly diversifying habitats on newly formed islands, or ecosystems recovering from recent disturbance and undergoing relatively rapid succession, as for example in the aftermath of fire. Because the deviation between the patterns predicted by theory and empirical data from disturbed systems appears to itself follow a systematic pattern (Rominger *et al.*, in prep.; Newman et al. 2014), clues exist for how to extend the static theory to the dynamic realm.

To promote the broad testing of METE across varied systems we have developed an open source (GNU General Public License v2) R package, meteR (version 1.0; cran.r-project.org/ web/packages/meteR) that calculates all of METE’s predictions in an object-oriented framework. We envision that meteR will be valuable for testing the scale dependence and equilibrium assumptions of METE’s predictions. That is, METE is not constrained to work at any specific spatial or temporal resolution, or for any particular taxonomic unit. While METE is typically applied to species coexisting in a single snap shot in time in a single plot, these conventions are imposed by data availability and not any fundamental attributes of the theory. One could readily study genus or family patterns across regional or subplot extents. meteR allows rapid generation of many models built at different scales or across different clades to better examines METE’s successes and failures. We encourage in particular exploration of how well METE predictions work across gradients of disturbance, ecosystem age, latitude, elevation, phylogenetic diversity, functional diversity and level of invasion. Tests of theory along these gradients can illuminate how the mechanisms associated with those gradients drive communities away from the statistical idealizations of METE. Such an understanding will help ecologists better predict when METE represents a sufficient model for their system, what ecological processes might be needed to improve predictions, or theoretical scaling relationships among state variables.

Only further comparison to data can address these open questions surrounding METE. To detect generalities, tests must be performed in a wide range of systems, which means that these tests must be performed by a large collection of researchers with different expertise. Achieve this goal motivated out development of an R package (R Development Core 2013b), meteR, which makes the theory accessible and provides an efficient workflow for evaluating METE with empirical data.

## meteR

Our R package, meteR, helps to address two key challenges with using METE: it reduces the need for practitioners to understand many of the mathematical details of the theory, and allows users to readily explore predictions without re-deriving maximum entropy solutions by hand. The prerequisite for using meteR is an understanding of its predicted macroecological distributions and relationships (Table [tab:metrics]), while details of exact solutions, numerical methods and mathematical approximations used to derive them are relegated to (user-accessible) functions “under the hood.” Importantly, meteR automatically checks that the conditions for all approximations used in METE predictions are met, and if not, implements more computationally costly exact solutions. Thus meteR provides not only the quickest approach to testing METE because of its mathematical abstraction, but also because it is operationally optimized. We contrast meteR with other software resources for METE, which either include fewer predictive features (Kitzes and Wilber 2016) or are exclusively low-level (e.g. code available at github.com/weecology/METE) and can only be used by those already well-versed in scientific computation. Additionally meteR is the only R resource available for METE, notable due to the accessibility and popularity of R, as opposed to other programming languages, among ecologists. Furthermore we have unit tested meteR using alredy published and verified solutions John Harte (2011; Newman et al. 2014) (results from these tests can be found at github.com/cmerow/meteR/tree/master/tests/testthat), meaning that users can confidently proceed with the results produced by our package. Additionally, bugs can be reported via GitHub’s issue tracking feature (github.com/cmerow/meteR/issues).

meteR also simplifies model comparison and visualization. Fitted macroecological distributions can be readily used to predict, simulate, or compare to data for use in further analyses. These operations are performed by familiar R functions; for example each fitted distribution contains a d element which gives the function of the probability distribution for use in visualization and an r element for random number generation (in analogy to dnorm and rnorm for a Gaussian distribution). Likelihoods can be computed using the standard logLik function and plotting carried out using plot. By reducing the effort spent on these tasks, meteR makes it practical to more fully explore METE models, e.g., by rapidly testing models built from data aggregated at different spatial, temporal or biological scales. Furthermore, owing to its object-oriented construction meteR allows users to take full advantage of other packages in R, for example model comparison procedures achieved via likelihood and AIC.

meteR provides a workflow to facilitate empirical tests of METE (Fig. [fig:workflow]). meteR calculates state variables directly from a variety of data formats. Next, it calculates the core probability distributions—the *Ecosystem Structure Function* (ESF) and *Spatial Structure Function* (SSF)—from which METE’s macroecological predictions arise. These core functions are next used to construct two types of macroecological predictions: probabilistic distributions (e.g. species abundance distribution) and deterministic relationships (e.g. the species-area relationship). The predictions are readily assessed via both plotting functions and statistical tests.

## Package Features

### Inputs

meteR accepts data in several formats, each of which is represented in the *Input* column in Figure [fig:workflow]:

1. One row per individual: This is useful when individual metabolic rate measurements are available
2. Multiple individuals of the same species per row: This is useful when multiple individuals are observed with the same values; e.g., if no metabolic rates are observed or if an average species metabolic rate is used. This format is also helpful if species are only located at the plot level, and individual-level coordinates are not needed/available.
3. Spatial information: This can be provided for individuals or counts of species either by the coordinates of those observations or the and IDs of those observations from a gridded landscape.
4. State variables: One can directly specify the values of , , and , to derive only the theoretical predictions, which is useful to compare how predictions change with different state variables from hypothetical datasets.

Note that models can be fit even when some components of the dataset are missing; e.g., if metabolic rate data are unavailable, meteR can still provide predictions that do not rely on these data (e.g., SAD or SAR). If location data are missing, all non-spatial predictions are available. The data are subsequently included in all model objects (see below) so that plots comparing predictions to observations can be be automatically produced. This flexibility in data formats, along with the fact that data are included with model objects, allows users to aggregate data at different spatial, temporal and taxonomic scales and fit models rapidly to test and organize predictions under a variety of assumptions.

### Core Functions

At the heart of METE are two probability distributions that are fit using the principle of maximum information entropy, the ESF and SSF (John Harte 2011). The ESF is fit with meteESF(...) using a nonlinear equation solver (package nleqslv; Hasselman 2015) to find the Lagrange multipliers. meteESF returns an object of class meteESF, for which a variety of methods are available. The ESF is typically not compared directly to data, but rather is used to obtain a number of more familiar macroecological patterns such as the SAD, IPD and SPD (Table [tab:metrics]; Harte et al. 2008; John Harte 2011 note that we use power—the “P” in IPD and SPD to refer to metabolic rate). In meteR, all of METE’s predicted distributions and relationships (cf. Table [tab:metrics]) can be obtained simply by passing a meteESF object to the appropriate function (Fig. [fig:workflow]).

The SSF () describes spatial structure via the probability that individuals of a species are observed in a plot of area given that individuals of that species were observed in the entire study area of size . For the case where METE predicts a uniform distribution over (p. 159 John Harte 2011). The more general form of the SSF is obtained similarly to the ESF using nleqslv in meteSSF(...), which returns an object of class meteSSF inheriting from meteESF. meteSSF objects can be used directly to predict patterns of spatial aggregation such as the SAR or the endemic-area relationship (Fig. [fig:workflow]).

### Macroecological Predictions

From the ESF and SSF, all of METE’s macroecological predictions can be obtained (Table [tab:metrics]). These predictions are divided into probability distributions (class meteDist) and deterministic relationships (class meteRelat), as denoted in Table [tab:metrics].

We use methods for object classes meteESF, meteDist and meteRelat to allow users to query models using familiar tools in R. For example, we include print and plot methods to explore model output. Similarly, for likelihood-based inference, logLik and AIC can be used to compare METE predictions to other distributions (e.g. comparing log-normal and log-series distributions for the SAD). Measures of model fit comparing METE predictions to data are readily performed using residuals or mse (for mean squared error). In a hypothesis testing framework, meteR also provides tools to estimate the z-score of model fit using either the logLikZ or mseZ functions. If the z-score is smaller than 1.96 this typically corresponds to failing to reject the hypothesis that the observed data came from a METE distribution.

### Directly Working with Probability Distributions

A useful feature of meteR is its conceptual treatment of distributions using the distr package (Ruckdeschel et al. 2006). For each of the macroecological probability distributions, meteR provides functions for the density function (d), cumulative distribution (p), quantile function (q) and simulation (r). This functionality for fitted METE distributions is analogous, e.g., to dnorm, pnorm, qnorm, rnorm for the normal distribution in the stats package (R Development Core 2013b). These functions are used extensively internally in meteR functions; e.g. r functions are used simulate data from fitted distributions to evaluate model fit via z-scores in mseZ and loglikZ (see below), while p is used for plotting cumulative distributions in plot. These distribution functions enable users to extend analyses to include more nuanced hypothesis testing, simulation for theoretical studies and custom plotting.

## Sample Code

We use two datasets, both made freely available and distributed with meteR, to illustrate its functionality. The first dataset comes from a 16x16m survey of herbaceous plants in the Anza Borrego Reserve in southern California John Harte (2011) which we will use to study patterns of abundance and spatial distribution (available as anbo in meteR). Briefly, the data set consists of 16 1 contiguous subplots. In each subplot, the abundance of each species was recorded, however individual metabolic rates were not recorded. Consequently, we also study a complementary dataset which contains individual body size measurements for one plot. This dataset comes from an extensive sample of arthropods collected by canopy fogging in native Hawaiian montane forest Gruner (2007), which we will use to study patterns of abundance and power (metabolic rate) distribution (available as arth in meteR).

With the anbo data set, we illustrate how to predict the species abundance distribution (), the Species-Area relationship (), and the *Spatial Structure Function* (). The code below generates the predicted distributions and produces Figure [fig:anbo]a.

data(anbo)  
 esf1 <- meteESF(spp=anbo$spp, abund=anbo$count)   
 sad1 <- sad(esf1)  
 sad1 # print function returns useful summary  
  
Species abundance distribution predicted using raw data   
with parameters:   
 S0 N0 E0   
 24 2445 NA   
 la1 la2   
0.00037 0.00109

From the SAD object we can extract familiar probability distribution functionality and plot it, e.g.,

sad1$r(20) # simulate 20 values  
 sad1$q(seq(0,1,length=10)) # 10% quantiles  
 plot(sad1, ptype='rad', log='y') # plotting can be customized by passing   
 # arguments to generic `plot' via `...'

The SAR and EAR follow similarly; we can either compute these relationships from the ESF or produce an object bundling the theoretical SAR with the observed data (plotting this bundled object produces Fig. [fig:anbo]b:

sar1 <- downscaleSAR(esf1, A=2^(seq(-2, 4, length=7)), A0=16)  
 ear1 <- downscaleSAR(esf1, A=2^(seq(-2, 4, length=7)), A0=16, EAR=TRUE)  
 ## sar2 bundles both the predicted SAR and the observed SAR similarly to   
 ## the output of `sad(...)'  
 sar2 <- meteSAR(anbo$spp, anbo$count, anbo$row, anbo$col, Amin=1, A0=16)  
 plot(sar2, xlim=c(1, 2^8), ylim=c(1, 45), log='xy')

The upscaled SAR uses a more involved, recursive optimization routine (see eqns (7.70) and (7.71) in John Harte 2011) but is equally accessible in meteR:

sarUP <- upscaleSAR(esf1, A0=16, Aup=2^8)  
 plot(sarUP, add=TRUE, col='blue') # add to previous SAR plot

With the arth data set, we illustrate the individual power distribution () as well as assessment of model fit using likelihood z-scores. Generating power distributions follows the same routine as species abundance distributions, starting with the ESF:

data(arth)  
 esf2 <- meteESF(spp=arth$spp, abund=arth$count, power=arth$mass^(3/4))  
 ipd2 <- ipd(esf2)

plot(ipd2, ptype='rad') # showing two plotting options: rank and cumulative  
 plot(ipd2, ptype='cdf')

Note that we assume the relationship based on metabolic theory (Brown et al. 2004). The two different plotting options display the rank plot and cumulative density plots, respectively, which are shown in panels (a) and (b) of Figure [fig:arth].

To evaluate model fit, we calculated the z-score of the likelihood of the data under the “null” model that the data truly did come from the METE distribution. Thus a z-score of less than fails to reject the hypothesis that the data came from a METE distribution.

ipd2.z <- logLikZ(ipd2, nrep=999, return.sim=TRUE) # return simulated values   
 # for plotting  
 ipd2.z$z # the z-value itself  
'log Lik.' -2.713953 (df=2)

The z-value less than indicates we cannot reject METE. Figure [fig:arth] (panel c) visually confirms this:

plot(density(ipd2.z$sim)) # density of simulated likelihoods  
 abline(v=ipd2.z$obs) # observed likelihood

More detailed examples are availible as a vignette included in meteR and accessible with the command

## Conclusions

By relegating mathematical details and conveniently organizing analyses, meteR allows users to readily address new macroecological questions. Until now, tests of METE have primarily been conducted by only a handful of experts, even though many more datasets surely exist that could help us to better understand METE’s strengths and weaknesses and macroecological patterns more generally. meteR lowers the activation energy for new users to begin to study both macroecology and METE, with the hope that more researchers will become engaged and invest the necessary time to learn the theory. As such, meteR is valuable as both a teaching and research tool.

## Acknowledgements

We thank J. Harte, Y. Zhang, E. Newman, and C. Lewis for constructive input throughout the development process. We are grateful to D. Gruner and E. Newman for providing data, and the Rocky Mountain Biological Lab where data were collected. AJR acknowledges funding from NSF grant DEB-1241253, the NSF Graduate Research Fellowship and the Gordon and Betty Moore Foundation. CM acknowledges funding from USDA-NRI grant 2008-35615-19014 and NSF grants CHN-1414108 and DEB-1137366.

## Figures

![](data:application/pdf;base64,)

meteR’s workflow. meteR accepts multiple data types, which are used to calculate the core probability distributions from which all predictions arise: the *Ecosystem Structure Function* (ESF) and the *Spatial Structure Function* (SSF). From the ESF and SSF, a variety of macroecological distributions can be calculated (see Table [tab:metrics] for definitions). Each of these can be plotted or summarized in various ways and model fit is readily assessed. Furthermore, density, distribution function, quantile function and random generation is available for each function, allowing for custom plotting, simulation and model evaluation.

Species abundance distribution (a) and species area relationship (b) for the anbo dataset. The blue line in (b) depicts the upscaled SAR predicted by METE. METE fits the Anza Borrego data poorly at increasingly small scales. Understanding such systematic deviations from theory could be an opportune area of exploration, and initial work suggests that such deviations could be an indicator of anthropogenic disturbance (Newman *et al.*, in prep.) such as invasive species, which are present in the Anza Borrego dataset. Note that we have stylized the log axes using custom code not included in meteR.

Species abundance distribution (a) and species area relationship (b) for the anbo dataset. The blue line in (b) depicts the upscaled SAR predicted by METE. METE fits the Anza Borrego data poorly at increasingly small scales. Understanding such systematic deviations from theory could be an opportune area of exploration, and initial work suggests that such deviations could be an indicator of anthropogenic disturbance (Newman *et al.*, in prep.) such as invasive species, which are present in the Anza Borrego dataset. Note that we have stylized the log axes using custom code not included in meteR.

Individual power distributions shown as a rank plot (a) and cumulative density plot (b). To evaluate model fit we compare the observed likelihood of the data given the METE model (red line in panel c) to simulated likelihoods produced by simulated nrep communities from the fitted METE object and calculating their likelihoods. Thus the simulated distribution represents hypothetical likelihoods for datasets obeying METE.

Individual power distributions shown as a rank plot (a) and cumulative density plot (b). To evaluate model fit we compare the observed likelihood of the data given the METE model (red line in panel c) to simulated likelihoods produced by simulated nrep communities from the fitted METE object and calculating their likelihoods. Thus the simulated distribution represents hypothetical likelihoods for datasets obeying METE.

## Tables

[1]

>

m#1

L4cm p1.2cm L5.4cm p2cm p2cm Predictions & Acronym & Functional Form & Type & meteR function  
Ecosystem Structure Function & ESF & & Core Distribution & meteESF  
Spatial Structure Function & SSF & & Core Distribution & meteSSF  
Distribution of Species Abundance & SAD & & Distribution & sad  
Distribution of individual Power across community & IPD & & Distribution & ipd  
Distribution of individual Power for species with individuals & SIPD & & Distribution & sipd  
Distribution of average (over individuals) metabolic rates over species & SPD & & Distribution & spd  
Average individual metabolic rate for species with individuals & ebar & & Relationship & ebar  
Species Area Relationship & SAR & & Relationship & meteSAR, downscaleSAR, upscaleSAR  
Endemics Area Relationship & EAR & & Relationship & meteSAR, downscaleSAR

# Conclusion

Using two seminal paleontological datasets (Sepkoski 1992; Alroy et al. 2008) I have shown that fluctuations in marine biodiversity over the past 550 million years results from the superposition of many independently fluctuating subsystems whose fluctuations are Gaussian but give rise to non-Gaussian patterns when combined. These independent subsystems correspond to lineages of closely related animal taxa, implying that diversification within lineages is driven by random additive interactions with the environment. These findings thus challenge the idea that changes in origination and extinction through deep geologic time are the result of complicated evolutionary interactions among organisms and between organisms and their environment (Bak and Sneppen 1993; Solé et al. 1997; Newman and Roberts 1995). However, I demonstrated that the evolutionary process responsible for generating new lineages varies slowly through time, possibly driven by non-random evolutionary innovations in the physiology and demography of new lineages. This slow change between lineages produces patterns of apparent complexity earlier ascribed to unnecessarily complicated mechanisms. I have further shown, using permutational null models, that these findings are not an artifact of how fossils are taxonomically classified but rather capture true underlying biological processes.

To further explore the importance of biological evolution in driving unique non-equilibrial patterns I synthesized population genetic and trophic network data for Hawaiian arthropods to show that as assembly by immigration (in communities assembled on young substrates) gives way to evolutionary processes (in communities assembled on old substrates), arthropod herbivore networks momentarily reach a steady state as predicted by equilibrial statistical mechanics (Rominger et al. 2015). On the young and old end of this spectrum different eco-evolutionary mechanisms lead to deviations from statistical mechanical theory: incomplete assembly and non-equilibrium adaptive evolution, respectively. Using population genetic data from other arthropod lineages I show that assembly and differentiation rates differ according to the trophic level of the organisms, implying that different trophic levels will reach an equilibrium at different periods and for different durations along the chronosequence. This study provides a framework for using island systems combined with simple equilibrial theory building to understand how complex communities emerge from ecological (population dynamics, dispersal, trophic interactions) and evolutionary (genetic structuring, adaptation, speciation, extinction) processes.

Finally, I argue that to fully realize the utility of statistical mechanics in the study of biodiversity, we must test these theories across many different systems. I advocate for further exploring how the evolutionary process and rapid ecological transitions could drive deviations from theory by proposing tests of the maximum entropy theory of ecology (J. Harte 2011) across gradients of disturbance, diversity and evolutionary age. To facilitate these novel tests I provided ecologists with the open source R package meteR.

“Top-down” approaches to biodiversity theory will be critical for building universal predictions about the the diversity of life (J. Harte 2011; Krakauer et al. 2011). These theories require that many biological details be course-grained but in turn promise unprecedented predictive power. Testing these theories in real ecosystems can illuminate where more detailed, biologically-grounded mechanisms are further needed. In this thesis I have used principles form statistical mechanics to develop and test new top-down theory that explains fluctuations in diversity through deep geologic time and identifies key evolutionary processes as important mechanisms to include in a synthetic eco-evolutionary theory of biodiversity. This work motivates exciting research directions for the near future which I sketch below.

## Future Work

### Adaptive landscapes and non-Markovian memory

Ecological theory has of late been dominated by neutral models (e.g. Rominger, Miller, and Collins 2009; Rominger et al. 2015). This approach needs robust alternative hypotheses because a lack of alternative theories rooted in classical ecology could be one limitation preventing a more rigorous competition between deterministic and statistical theories of biodiversity. My work with super-statistics in the fossil record (Chapter 1) motivates a new approach to ecological theory using super-statistics to parsimoniously capture the non-neutrality of species and relate that non-neutrality to the non-equilibrium process of diversification on an adaptive landscape.

### Metabarcoding and Ecological Theory

Combining test of ecological theory with massive phylogenetic data, as motivated by my synthesis of population genetics with trophic networks (Chapter 2), could determine what aspects of evolutionary history specifically shape ecological communities, but massive phylogenetic data on the scale of large ecological studies is limiting. A new approach dubbed “metabarcoding” (Taberlet et al. 2012) harnesses next generation sequencing technology to produce massive amounts of genetic data for ecological samples. This approach, however, has known pitfalls (e.g. bias in primer affinities between taxa Clarke et al. 2014) but these could be overcome bioinformatically using hierarchical models, already commonplace in ecology (Royle and Dorazio 2008).

# Supplement to “Punctuated non-equilibrium and niche conservatism explain biodiversity fluctuations through the Phanerozoic”

## Limit distribution of a time-averaged homogeneous origination-extinction process

Fossil taxa gain and lose taxa according to an origination-extinction process. We assume that most fossil occurrences of a taxon come from the period of its history when it is dominant and in steady state. In a time slice of duration during such a period of steady state the latent per capita rates of origination and extinction would be equal (i.e. ) and the number of origination or extinctions events (call such events ) each follow an inhomogeneous Poisson process with rate where is the number of species or genera in the taxon of interest at time . Allowing to vary smoothly with time, and invoking the comunicative property of the Poisson distribution, we arrive at the number of extinction *or* origination events in being distributed

Under the steady state assumption we can approximate by , the steady state diversity, leading to

Assuming the of each time period in the Paleobiology Database or Sepkoski’s compendium to be approximately equal (i.e. equal durations of major stratographic units) then the distribution of fluctuations within taxa will be asymptotically Gaussian.

## Additional super-statistical analyses

To evaluate the sensitivity of our super-statistical analysis on the particular data used and we tested our predictions on different data sets (see below). The fact that it works in all different applications indicates that it is robust to vagaries of different recording strategies and bias corrections in paleobiology. This could mean that much of the raw signal in massive fossil datasets, at least signals regarding fluctuations, are not artifacts of sampling, as has been proposed before Hannisdal and Peters (2011).

### Raw PBDB data

We calculated the super-statistical prediction at the order level from raw genus diversity recorded in the PBDB without correcting for taphonomic or sampling bias (Fig. [fig:supprawPBDBPx]). The super-statistical calculation also closely fits the raw data as in the case of sampling and publication bias-corrected data.

### Different taxonomic ranks in PBDB data

As noted in the main text, the super-statistical prediction predictably breaks down at higher taxonomic scales. In Figure [fig:suppPBDBPxcls] we present this worsening fit graphically using class level data with three-timer and publication corrected PBDB data

### Sepkoski’s compendium

Sepkoski’s compendium Sepkoski (1992) provided the first hypothesis of Phanerozoic diversification. As such, it has served as a benchmark for further investigation into large-scale paleobiological patternsAlroy et al. (2008). We conducted the same super-statistical analysis as in the main text and find comparable results. Specifically, the super-statistical prediction far out preforms the null Gaussian model (Fig. [fig:suppsepkPx]) and worsens with increasing taxonomic scale (Fig. [fig:suppsepkPx]), again implying the uniqueness of orders.

## Supplemental Figures

![](data:application/pdf;base64,)

Relationship between number of publications and genus diversity as recorded by the PBDB.

![](data:application/pdf;base64,)

Comparison of SQS method Alroy (2010) (solid black line) with the raw data (dashed black) and our three-timer and publication bias correction method (red). The time-series of all marine invertebrate genera shows general agreement with the only major deviations toward the modern (A). Despite these differences the distribution of fluctuations in genus diversity across all marine invertebrates show good agreement (B).

![](data:application/pdf;base64,)

Super-statistical prediction of raw (i.e. not bias corrected) order-level fluctuations in genus diversity recorded in the PBDB. Grey dots are the full data of orders, while black ones are orders with more than 15 points. The red line is our theoretical prediction and the blue line the best Gaussian fit to the data.

![](data:application/pdf;base64,)

Super-statistical prediction of bias corrected class-level fluctuations in genus diversity recorded in the PBDB. Grey dots are the full data of orders, while black ones are orders with more than 15 points. The red line is our theoretical prediction and the blue line the best Gaussian fit to the data. Note at the class level the fit is predictably worse, see main text for discussion.

![](data:application/pdf;base64,)

Super-statistical prediction (red line) of fluctuations in genus diversity recorded in Sepkoski’s compendium of marine invertebrates compared to maximum likelihood normal distribution (blue line). Super-statistical theory explains order level fluctuations well (A) with increasingly poorer fits at the class (B) and phylum (C) levels.

# Supplement to “Community assembly on isolated islands: Macroecology meets evolution”

## Compilation of networks and metric validation

Researchers have put forward a set of “network metrics,” including nestedness (Bascompte et al. 2003; Ulrich, Almeida-Neto, and Gotelli 2009) and modularity (Newman and Girvan 2004; Olesen et al. 2007), to understand the complex structure of ecological networks. Null models are used to evaluate the statistical significance of these metrics and to compare between networks of different size (Ulrich, Almeida-Neto, and Gotelli 2009). We compare the results derived from two common null models: the “probabilistic null” of Bascompte et al. (2003) using the relative degree distributions of plants and herbivores as weights while randomizing links and suffers from high Type II error (Ulrich, Almeida-Neto, and Gotelli 2009); the “fixed-fixed null” (Ulrich, Almeida-Neto, and Gotelli 2009) maintains the exact number of links assigned to each species while randomizing which interactors fill the requisite set of links and suffers from high Type I error (Ulrich, Almeida-Neto, and Gotelli 2009). We find that using these different null models does not change any trends in our network statistics across the Hawaiian chronosequence but different null models do influence the sign and significance of the network metrics (Fig. [figSupp:netMetComp]). We therefore do not interpret the sing or significance of the metrics but only their relative trends across site age.

Because these networks are based on opportunistic data associated with species descriptions, and not based on standardized ecological surveys, we cannot interpret patterns in network metrics without evaluating possible sampling biases (Nielsen and Bascompte 2007; Gibson et al. 2011; Rivera-Hutinel et al. 2012). To do so we rarify networks by the number of Hemiptera species included and, for each subsampled network, re-calculate nestedness and modularity z-scores. This rarefaction procedure shows that nestedness is very sensitive to network size (Fig. [figSupp:rfy]), a known property of nestedness (Nielsen and Bascompte 2007; Gibson et al. 2011; Rivera-Hutinel et al. 2012). However the relative nestedness z-scores across networks remain qualitatively similar to those observed for the complete networks (Fig. [figSupp:rfy]). Modularity depends on network size in a more variable way (Fig. [figSupp:rfy]). Modularity is expected to decrease with network size (Rivera-Hutinel et al. 2012) and so the marked increase in modularity with network size on Haleakala is unexpected. However in light of the large number of highly specialized taxa this pattern is more reasonable—if most species only have within module links then removing these species through subsampling will only reduce overall modularity. Thus this pattern speaks to the high level of specialization on Haleakala, and to a lesser extent at Kokee which also shows a slight increase in modularity with network size (Fig. [figSupp:rfy]).

## R Scripts for Maximum Entorpy Analysis and Monte Carlo Methods

### Needed Functions

require(distr)  
  
## d, p and r functions for the maxent link distribution  
## also a funciton to calculate the MLE and log likelihood  
  
dmelink <- function(x, la) {  
 exp(-la\*(x-1)) - exp(-la\*x)  
}  
  
pmelink <- function(x, la, lower.tail=TRUE) {  
 if(length(x) > 1) {  
 cp <- sapply(x, function(q) sum(dmelink(1:q, la)))  
 } else {  
 cp <- sum(dmelink(1:x, la))  
 }  
   
 if(lower.tail) {  
 return(cp)  
 } else {  
 return(1 - cp)  
 }  
}  
  
rmelink <- function(n, la, X0) {  
 sample(X0, n, rep=TRUE, prob=dmelink(1:X0, la))  
}  
  
  
qmelink <- function(x, la) {  
 fun <- DiscreteDistribution(1:3000, dmelink(1:3000, la))  
 fun@q(x)  
}  
  
## likelihood functions (for simple cases maxent solution is equivilant   
## to maximum likelihood solution so we use MLE for computational ease)  
melink.mle <- function(x) {  
 log(1 + 1 / mean(x - 1))  
}  
  
melink.logLik <- function(x) {  
 la <- melink.mle(x)  
   
 length(x) \* log(1 - exp(-la)) - la \* sum(x - 1)  
}  
  
  
## function to make rank function (for plotting) of maxent link distrib  
melink.rankFun <- function(x) {  
 qmelink(seq(1-1/(length(x)+1), 1/(length(x)+1),   
 by=-1/(length(x)+1)), melink.mle(x))  
}  
  
## mean squared error function for maxent link distrib  
melink.mse <- function(x) {  
 mean((sort(x, TRUE) - melink.rankFun(x))^2)  
}  
  
## monte carlo method for calculating distribution of mse values   
## under model where maxent truely generated the link distribution  
sim.melink.z <- function(x, X0, nsim=999) {  
 la <- melink.mle(x)  
   
 res <- replicate(nsim, {  
 # browser()  
 sim <- rmelink(length(x), la, X0)  
 return(melink.mse(sim))  
 })  
   
 obs.mse <- melink.mse(x)  
   
 return(list(obs=obs.mse, z=(obs.mse - mean(res)) / sd(res), sim=res))  
}

### Example Use

## set a Lagrange multiplier value and simulate data  
la <- 0.01  
x <- rmelink(2000, la, 2000)  
  
## fit the maxEnt model to simulated data  
x <- rmelink(2000, la, 2000)  
  
## evalute if fitted maxEnt model matches the data  
melink.mle(x) # the MLE should be near 0.01  
plot(sort(x, TRUE)) # the plotted data should look like the theory  
lines(pmelink(1:200, melink.mle(x)), col='red')  
  
## test that the logLike function is returning the correct value  
melink.logLik(x) - sum(log(dmelink(x, melink.mle(x)))) < 1e-12  
  
## test that the log likelihood simulation is working  
sim.melink.z(x, 200) # should be a small z-value  
sim.melink.z(rbinom(100, 200, 0.1), 200) # should be a large z-value

## Supplemental Figures

![](data:application/pdf;base64,)

Comparison of different null models (“Prob” and “Quasiswap”) used to standardize network metrics and comparison of different algorithms for assessing modularity (“edge between” and “walk drap”). Choice of null model has a strong influence on the sign and magnitude of metrics but not on their relative trends. The different modularity algorithms lead to largely similar qualitative patterns.

![](data:application/pdf;base64,)

Metrics NODF and modularity calculated for networks based on more biogeographically conservative assignment of Hemiptera to localities. Colors and metric specifics as in Figure [figSupp:netMetComp].

![](data:application/pdf;base64,)

Result from rarification analysis showing sensitivity of network metrics to number of Hemiptera sampled.

Albert, Réka, and Albert-László Barabási. 2002. “Statistical Mechanics of Complex Networks.” *Reviews of Modern Physics* 74 (1). APS: 47.

Albrecht, Matthias, Matthias Riesen, and Bernhard Schmid. 2010. “Plant–pollinator Network Assembly Along the Chronosequence of a Glacier Foreland.” *Oikos* 119 (10). Wiley Online Library: 1610–24.

Allesina, Stefano, and Si Tang. 2012. “Stability Criteria for Complex Ecosystems.” *Nature* 483 (7388). Nature Publishing Group: 205–8.

Almeida-Neto, M., P. Gumarães, P.R. Gumarães, R.D. Loyola, and W. Ulrich. 2008. “A Consistent Metric for Nestedness Analysis in Ecological Systems: Reconciling Concept and Measurement.” *Oikos* 117: 1227–39.

Alroy, J. 2010. “The Shifting Balance of Diversity Among Major Marine Animal Groups.” *Science* 329 (5996): 1191–94. doi:[10.1126/science.1189910](http://dx.doi.org/10.1126/science.1189910).

Alroy, John, Martin Aberhan, David J. Bottjer, Michael Foote, Franz T. Fürsich, Peter J. Harries, Austin J. W. Hendy, et al. 2008. “Phanerozoic Trends in the Global Diversity of Marine Invertebrates.” *Science* 321: 97–100. doi:[10.1126/science.1156963](http://dx.doi.org/10.1126/science.1156963).

Bak, Per, and Kim Sneppen. 1993. “Punctuated Equilibrium and Criticality in a Simple Model of Evolution.” *Phys. Rev. Lett.* 71: 4083–86. <http://link.aps.org/doi/10.1103/PhysRevLett.71.4083>.

Banavar, JR, and A Maritan. 2007. “The Maximum Relative Entropy Principle. Á Bhttp.” *Arxiv. Org/abs/cond-Mat/0703622*.

Banavar, JR, A Maritan, and I Volkov. 2010. “Applications of the principle of maximum entropy: from physics to ecology.” *Journal of Physics: Condensed Matter* 22: 063101.

Bascompte, J., P. Jordano, C. J. Melian, and J. M. Olesen. 2003. “The Nested Assembly of Plant-Animal Mutualistic Networks.” *Proc. Natl Acad. Sci. USA* 100: 9383–87.

Bascompte, Jordi, and Pedro Jordano. 2007. “Plant-Animal Mutualistic Networks: The Architecture of Biodiversity.” *Annu. Rev. Ecol. Evol. Syst.* 38: 567–93.

Beck, C., and E.G.D. Cohen. 2003. “Superstatistics.” *Physica A: Statistical Mechanics and Its Applications* 322: 267–75. doi:[10.1016/S0378-4371(03)00019-0](http://dx.doi.org/10.1016/S0378-4371(03)00019-0).

Beck, Christian. 2004. “Superstatistics in Hydrodynamic Turbulence.” *Physica D: Nonlinear Phenomena* 193: 195–207.

Bennett, Gordon M, and Patrick M O’Grady. 2013. “Historical Biogeography and Ecological Opportunity in the Adaptive Radiation of Native Hawaiian Leafhoppers (Cicadellidae: Nesophrosyne).” *Journal of Biogeography* 40 (8). Wiley Online Library: 1512–23.

Benton, M.J. 1995. “Diversification and Extinction in the History of Life.” *Science* 268: 52–58. doi:[10.1126/science.7701342](http://dx.doi.org/10.1126/science.7701342).

Berlow, Eric L, Jennifer A Dunne, Neo D Martinez, Philip B Stark, Richard J Williams, and Ulrich Brose. 2009. “Simple Prediction of Interaction Strengths in Complex Food Webs.” *Proceedings of the National Academy of Sciences* 106 (1). National Acad Sciences: 187–91.

Borer, Elizabeth T, Eric W Seabloom, Daniel S Gruner, W Stanley Harpole, Helmut Hillebrand, Eric M Lind, Peter B Adler, et al. 2014. “Herbivores and Nutrients Control Grassland Plant Diversity via Light Limitation.” *Nature*. Nature Publishing Group.

Brewer, Michael S, Rebecca A Carter, Peter JP Croucher, and Rosemary G Gillespie. 2015. “Shifting Habitats, Morphology, and Selective Pressures: Developmental Polyphenism in an Adaptive Radiation of Hawaiian Spiders.” *Evolution* 69 (1). Wiley Online Library: 162–78.

Brose, Ulrich, Eric L Berlow, and Neo D Martinez. 2005. “Scaling up Keystone Effects from Simple to Complex Ecological Networks.” *Ecology Letters* 8 (12). Wiley Online Library: 1317–25.

Brose, Ulrich, Richard J Williams, and Neo D Martinez. 2006. “Allometric Scaling Enhances Stability in Complex Food Webs.” *Ecology Letters* 9 (11). Wiley Online Library: 1228–36.

Brown, J., J. Gillooly, A. Allen, V. Savage, and G. West. 2004. “Toward a Metabolic Theory of Ecology.” *Ecology* 85 (7): 1771–89.

Brown, J.H. 1995. “Macroecology.” Chicago, Illinois, USA: The University of Chicago Press.

Brown, James H. 1995. *Macroecology*. University of Chicago Press.

Butler, Marguerite A, and Aaron A King. 2004. “Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution.” *The American Naturalist* 164 (6). JSTOR: 683–95.

Cavender-Bares, Jeannine, Kenneth H Kozak, Paul VA Fine, and Steven W Kembel. 2009. “The Merging of Community Ecology and Phylogenetic Biology.” *Ecology Letters* 12 (7). Wiley Online Library: 693–715.

Chave, Jerôme. 2004. “Neutral Theory and Community Ecology.” *Ecology Letters* 7 (3). Wiley Online Library: 241–53.

Clarke, Laurence J, Julien Soubrier, Laura S Weyrich, and Alan Cooper. 2014. “Environmental Metabarcodes for Insects: In Silico PCR Reveals Potential for Taxonomic Bias.” *Molecular Ecology Resources* 14 (6). Wiley Online Library: 1160–70.

Cox, John C, Jonathan E Ingersoll Jr, and Stephen A Ross. 1985. “A Theory of the Term Structure of Interest Rates.” *Econometrica: Journal of the Econometric Society*. JSTOR, 385–407.

Csardi, Gabor, and Tamas Nepusz. 2006. “The Igraph Software Package for Complex Network Research.” *InterJournal* Complex Systems: 1695. <http://igraph.sf.net>.

Daily, Gretchen. 1997. *Nature’s Services: Societal Dependence on Natural Ecosystems*. Island Press.

Damuth, J. 1981. “Population density and body size in mammals.” *Nature*.

Dewar, Roderick C, and Annabel Porté. 2008. “Statistical Mechanics Unifies Different Ecological Patterns.” *Journal of Theoretical Biology* 251 (3). Elsevier: 389–403.

Donatti, Camila I, Paulo R Guimarães, Mauro Galetti, Marco Aurélio Pizo, Flavia Marquitti, and Rodolfo Dirzo. 2011. “Analysis of a Hyper-Diverse Seed Dispersal Network: Modularity and Underlying Mechanisms.” *Ecol. Lett.* 14 (8). Wiley Online Library: 773–81.

Eldredge, Niles, and Stephen Jay Gould. 1972. “Punctuated Equilibria: An Alternative to Phyletic Gradualism.” *Models in Paleobiology* 82. San Francisco: 115.

Erwin, Douglas H. 2007. “Disparity: Morphological Pattern and Developmental Context.” *Palaeontology* 50 (1). Wiley Online Library: 57–73.

———. 2012. “Novelties That Change Carrying Capacity.” *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 318 (6). Wiley Online Library: 460–65.

Erwin, Douglas H. 1998. “The End and the Beginning: Recoveries from Mass Extinctions.” *Trends in Ecology and Evolution* 13: 344–49.

Excoffier, Laurent, and Heidi EL Lischer. 2010. “Arlequin Suite Ver 3.5: A New Series of Programs to Perform Population Genetics Analyses Under Linux and Windows.” *Molecular Ecology Resources* 10 (3): 564–67.

Fisher, Ronald A, A Steven Corbet, and Carrington B Williams. 1943. “The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population.” *The Journal of Animal Ecology*. JSTOR, 42–58.

Foote, Michael, James S Crampton, Alan G Beu, and Roger A Cooper. 2008. “On the Bidirectional Relationship Between Geographic Range and Taxonomic Duration.” *Paleobiology* 34 (4): 421–33.

Fuentes, Austin AND Vicente, Miguel A. AND Gerig. 2009. “Universal Behavior of Extreme Price Movements in Stock Markets.” *PLoS ONE* 4. Public Library of Science: e8243. doi:[10.1371/journal.pone.0008243](http://dx.doi.org/10.1371/journal.pone.0008243).

Gavrilets, Sergey, and Aaron Vose. 2005. “Dynamic Patterns of Adaptive Radiation.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (50). National Acad Sciences: 18040–45.

Gibson, Rachel H., Ben Knott, Tim Eberlein, and Jane Memmott. 2011. “Sampling Method Influences the Structure of Plant-Pollinator Networks.” *Oikos* 120 (6): 822–31.

Gilinsky, Norman L. 1994. “Volatility and the Phanerozoic Decline of Background Extinction Intensity.” *Paleobiology*. JSTOR, 445–58.

Gillespie, R. 2004. “Community Assembly Through Adaptive Radiation in Hawaiian Spiders.” *Science* 303 (5656). American Association for the Advancement of Science: 356–59.

Gillespie, R.G. 2009. “Adaptive Radiation.” In *Guide to Ecology*, edited by S. Levin, S. Carpenter, C. Godfray, A. Kinzig, M. Loreau, J.B. Losos, B. Walker, and D. Wilcove, 143–52. Princeton, NJ: Princeton University Press.

Gillespie, RG, and GK Roderick. 2014. “Geology and Climate Drive Diversification.” *Nature* 509: 207–98.

Golan, Amos, George Judge, and Douglas Miller. 1996. “Maximum Entropy Econometrics: Robust Estimation with Limited Data.” Chichester (United Kingdom) John Wiley; Sons.

Goodman, K.R., S.C. Welter, and G.K. Roderick. 2012. “Genetic Divergence Is Decoupled from Ecological Diversification in the Hawaiian Nesosydne Planthoppers.” *Evolution* 66: 2798–2814.

Goodman, KR, JP Kelley, SC Welter, GK Roderick, and DO Elias. 2015. “Rapid Diversification of Sexual Signals in Hawaiian Nesosydne Planthoppers (Hemiptera: Delphacidae): The Relative Role of Neutral and Selective Forces.” *Journal of Evolutionary Biology* 28 (2). Wiley Online Library: 415–27.

Gould, Stephen Jay. 1985. “The Paradox of the First Tier: An Agenda for Paleobiology.” *Paleobiology* 11 (01). Cambridge Univ Press: 2–12.

Gruner, Daniel S. 2007. “Geological Age, Ecosystem Development, and Local Resource Constraints on Arthropod Community Structure in the Hawaiian Islands.” *Biological Journal of the Linnean Society* 90 (3). Wiley Online Library: 551–70.

Gull, SF, and TJ Newton. 1986. “Maximum Entropy Tomography.” *Applied Optics* 25 (1). Optical Society of America: 156–60.

Haegeman, Bart, and Michel Loreau. 2008. “Limitations of Entropy Maximization in Ecology.” *Oikos* 117 (11). Wiley Online Library: 1700–1710.

Hannisdal, Bjarte, and Shanan E. Peters. 2011. “Phanerozoic Earth System Evolution and Marine Biodiversity.” *Science* 334 (6059): 1121–24. doi:[10.1126/science.1210695](http://dx.doi.org/10.1126/science.1210695).

Harnik, Paul G. 2011. “Direct and Indirect Effects of Biological Factors on Extinction Risk in Fossil Bivalves.” *Proceedings of the National Academy of Sciences* 108 (33). National Acad Sciences: 13594–99.

Harte, J, T Zillio, E Conlisk, and AB Smith. 2008. “Maximum entropy and the state-variable approach to macroecology.” *Ecology* 89 (10): 2700–2711.

Harte, J. 2011. *The Maximum Entropy Theory of Ecology*. Oxford University Press.

Harte, John. 2011. *Maximum entropy and ecology: a theory of abundance, distribution, and energetics*. Oxford, UK: Oxford University Press.

Harte, John, Adam B Smith, and David Storch. 2009. “Biodiversity scales from plots to biomes with a universal species-area curve.” *Ecology Letters* 12 (8): 789–97.

Hasselman, B. 2015. “nleqslv: Solve systems of non linear equations. R package version 2.1. 1.”

Holman, Eric W. 1989. “Some Evolutionary Correlates of Higher Taxa.” *Paleobiology*. JSTOR, 357–63.

Hopkins, Melanie J, Carl Simpson, and Wolfgang Kiessling. 2014. “Differential Niche Dynamics Among Major Marine Invertebrate Clades.” *Ecology Letters* 17 (3). Wiley Online Library: 314–23.

Hubbell, S. P. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*. Vol. 32. Princeton University Press.

Jablonski, David. 2008. “Species Selection: Theory and Data.” *Annual Review of Ecology, Evolution, and Systematics* 39: 501–24.

Jaynes, E. 2003. “Probability Theory: The Logic of Science.” *Cambridge University Press*.

Jaynes, E T. 1982. “On the rationale of maximum-entropy methods.” *Proceedings of the IEEE* 70 (9): 939–52.

Jaynes, ET. 1957. “Information theory and statistical mechanics. II.” *Physical Review* 108 (2): 171–90.

Jetz, W, GH Thomas, JB Joy, K Hartmann, and AO Mooers. 2012. “The Global Diversity of Birds in Space and Time.” *Nature* 491 (7424). Nature Publishing Group: 444–48.

Jordano, P., J. Bascompte, and J. M. Olesen. 2003. “Invariant Properties in Coevolutionary Networks of Plant-Animal Interactions.” *Ecol. Lett.* 6: 69–81.

Kirchner, James W., and Anne Weil. 1998. “No Fractals in Fossil Extinction Statistics.” *Nature* 395: 337–38. <http://dx.doi.org/10.1038/26384>.

Kitayama, Kanehiro, and Dieter Mueller-Dombois. 1995. “Vegetation Changes Along Gradients of Long-Term Soil Development in the Hawaiian Montane Rainforest Zone.” *Vegetatio* 120 (1). Springer: 1–20.

Kitzes, Justin, and Mark Wilber. 2016. “Macroeco: Reproducible Ecological Pattern Analysis in Python.” *Ecography*. Blackwell Publishing Ltd.

Krakauer, David C, James P Collins, Douglas Erwin, Jessica C Flack, Walter Fontana, Manfred D Laubichler, Sonja J Prohaska, Geoffrey B West, and Peter F Stadler. 2011. “The Challenges and Scope of Theoretical Biology.” *Journal of Theoretical Biology* 276 (1). Elsevier: 269–76.

Lewinsohn, Thomas M, Vojtech Novotny, and Yves Basset. 2005. “Insects on Plants: Diversity of Herbivore Assemblages Revisited.” *Annual Review of Ecology, Evolution, and Systematics*. JSTOR, 597–620.

Liow, Lee Hsiang, and Nils Chr Stenseth. 2007. “The Rise and Fall of Species: Implications for Macroevolutionary and Macroecological Studies.” *Proceedings of the Royal Society B: Biological Sciences* 274 (1626). The Royal Society: 2745–52.

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

Madin, Joshua S., John Alroy, Martin Aberhan, Franz T. Fürsich, Wolfgang Kiessling, Matthew A. Kosnik, and Peter J. Wagner. 2006. “Statistical Independence of Escalatory Ecological Trends in Phanerozoic Marine Invertebrates.” *Science* 312: 897–900. doi:[10.1126/science.1123591](http://dx.doi.org/10.1126/science.1123591).

Magnacca, Karl N, David Foote, and PATRICK M O’Grady. 2008. “A Review of the Endemic Hawaiian Drosophilidae and Their Host Plants.” *Zootaxa* 1728: 1–58.

Mayr, Ernst. 1965. “Numerical Phenetics and Taxonomic Theory.” *Systematic Zoology* 14 (2). JSTOR: 73–97.

McGill, Brian J, and Jeffrey C Nekola. 2010. “Mechanisms in Macroecology: AWOL or Purloined Letter? Towards a Pragmatic View of Mechanism.” *Oikos* 119 (4). Wiley Online Library: 591–603.

Mendelson, Tamra C, and Kerry L Shaw. 2005. “Sexual Behaviour: Rapid Speciation in an Arthropod.” *Nature* 433 (7024). Nature Publishing Group: 375–76.

Miller, M.P. 1997. “Tools for Population Genetic Analysis (TFPGA), 1.3: A Windows Program for the Analysis of Allozyme and Molecular Population Genetic Data. Distributed by the Author: Http://www.marksgeneticsoftware.net/tfpga.htm.”

Nekola, Jeffrey C, and James H Brown. 2007. “The Wealth of Species: Ecological Communities, Complex Systems and the Legacy of Frank Preston.” *Ecology Letters* 10 (3). Wiley Online Library: 188–96.

Newman, CM, JE Cohen, and C Kipnis. 1985. “Neo-Darwinian Evolution Implies Punctuated Equilibria.” *Nature* 315 (6018). Nature Publishing Group: 400–401.

Newman, Erica A, Mary Ellen Harte, Natalie Lowell, Mark Wilber, and John Harte. 2014. “Empirical Tests of Within-and Across-Species Energetics in a Diverse Plant Community.” *Ecology* 95 (10). Wiley Online Library: 2815–25.

Newman, M, and M Girvan. 2004. “Finding and Evaluating Community Structure in Networks.” *Physical Review E* 69: 026113.

Newman, M. E. J., and B. W. Roberts. 1995. “Mass Extinction: Evolution and the Effects of External Influences on Unfit Species.” *Proceedings of the Royal Society of London B* 260: 31–37. doi:[10.1098/rspb.1995.0055](http://dx.doi.org/10.1098/rspb.1995.0055).

Nielsen, Anders, and Jordi Bascompte. 2007. “Ecological Networks, Nestedness and Sampling Effort.” *Journal of Ecology* 95 (5). Wiley Online Library: 1134–41.

Nuismer, Scott L., Pedro Jordano, and Jordi Bascompte. 2013. “Coevolution and the Architecture of Mutualistic Networks.” *Evolution* 67 (2): 338–54.

Odling-Smee, F John, Kevin N Laland, and Marcus W Feldman. 2003. *Niche Construction: The Neglected Process in Evolution*. 37. Princeton University Press.

Oksanen, Jari, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O’Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, and Helene Wagner. 2013. *Vegan: Community Ecology Package*. <http://CRAN.R-project.org/package=vegan>.

Olesen, J., J. Bascompte, Y. Dupont, and P. Jordano. 2007. “The Modularity of Pollination Networks.” *Proc. Natl Acad. Sci. USA* 104: 19891–96.

O’Grady, Patrick M, Richard T Lapoint, James Bonacum, Jackline Lasola, Elaine Owen, Yifei Wu, and Rob DeSalle. 2011. “Phylogenetic and Ecological Relationships of the Hawaiian Drosophila Inferred by Mitochondrial DNA Analysis.” *Molecular Phylogenetics and Evolution* 58 (2). Elsevier: 244–56.

Percy, D.M. 2003. “Radiation, Diversity and Host Plant Interactions Among Island and Continental Legume-Feeding Psyllids.” *Evolution* 57: 2540–56.

Percy, D.M., and M. Kennedy. 2006. “Psyllid Communication: Acoustic Diversity, Mate Recognition and Phylogenetic Signal.” *Invertebrate Systematics* 20: 431–45.

Phillips, Steven J, Robert P Anderson, and Robert E Schapire. 2006a. “Maximum Entropy Modeling of Species Geographic Distributions.” *Ecological Modelling* 190 (3). Elsevier: 231–59.

———. 2006b. “Maximum entropy modeling of species geographic distributions.” *Ecological Modelling* 190: 231–59.

Preston, Frank W. 1950. “Gas Laws and Wealth Laws.” *The Scientific Monthly* 71: 309–11.

———. 1962a. “The Canonical Distribution of Commonness and Rarity: Part I.” *Ecology* 43 (2). Eco Soc America: 185–215.

———. 1962b. “The Canonical Distribution of Commonness and Rarity: Part II.” *Ecology* 43 (3). JSTOR: 410–32.

Price, J.P., and D.A. Clague. 2002. “How Old Is the Hawaiian Biota? Geology and Phylogeny Suggest Recent Divergence.” *Proc R Soc Lond B Biol Sci* 269: 2429–35.

Price, Jonathan P. 2012. *Mapping Plant Species Ranges in the Hawaiian Islands: Developing a Methodology and Associated GIS Layers*. US Department of the Interior, US Geological Survey.

Price, Trevor D, Daniel M Hooper, Caitlyn D Buchanan, Ulf S Johansson, D Thomas Tietze, Per Alström, Urban Olsson, et al. 2014. “Niche Filling Slows the Diversification of Himalayan Songbirds.” *Nature* 509 (7499). Nature Publishing Group: 222–25.

Pueyo, Salvador, Fangliang He, and Tommaso Zillio. 2007. “The Maximum Entropy Formalism and the Idiosyncratic Theory of Biodiversity.” *Ecology Letters* 10 (11). Wiley Online Library: 1017–28.

Quental, Tiago B, and Charles R Marshall. 2013. “How the Red Queen Drives Terrestrial Mammals to Extinction.” *Science*. American Association for the Advancement of Science.

R Development Core, Team. 2013a. “R: A language and environment for statistical computing.” *Vienna, Austria*.

———. 2013b. “R: A language and environment for statistical computing.” *Vienna, Austria*.

Rabosky, Daniel L. 2006. “LASER: A Maximum Likelihood Toolkit for Detecting Temporal Shifts in Diversification Rates from Molecular Phylogenies.” *Evolutionary Bioinformatics Online* 2. Libertas Academica: 247.

———. 2009. “Ecological Limits and Diversification Rate: Alternative Paradigms to Explain the Variation in Species Richness Among Clades and Regions.” *Ecology Letters* 12 (8). Wiley Online Library: 735–43.

Raup, David M, J John Sepkoski Jr, and others. 1982. “Mass Extinctions in the Marine Fossil Record.” *Science* 215 (4539): 1501–3.

Ricklefs, Robert E. 2004. “A Comprehensive Framework for Global Patterns in Biodiversity.” *Ecology Letters* 7 (1). Wiley Online Library: 1–15.

———. 2006. “The Unified Neutral Theory of Biodiversity: Do the Numbers Add up?” *Ecology* 87 (6). Eco Soc America: 1424–31.

Rivera-Hutinel, A, RO Bustamante, VH Marín, and R Medel. 2012. “Effects of Sampling Completeness on the Structure of Plant-Pollinator Networks.” *Ecology* 93 (7). Eco Soc America: 1593–1603.

Roderick, G. K., and D.M. Percy. 2008. “Host Plant Use, Diversification, and Coevolution: Insights from Remote Oceanic Islands.” In *Specialization, Speciation, and Radiation. Evolutionary Biology of Herbivorous Insects*, edited by K.J. Tilmon, 151–61. Berkeley: University of California Press.

Roderick, G.K., P.J.P. Croucher, A.G. Vandergast, and R.G. Gillespie. 2012. “Species Differentiation on a Dynamic Landscape: Shifts in Metapopulation and Genetic Structure Using the Chronology of the Hawaiian Archipelago.” *Evolutionary Biology* 32: 192–206.

Rominger, A. J., T. E. X. Miller, and S. L. Collins. 2009. “Relative Contributions of Neutral and Niche-Based Processes to the Structure of a Desert Grassland Grasshopper Community.” *Oecologia* 161 (4). Springer: 791–800.

Rominger, AJ, KR Goodman, JY Lim, EE Armstrong, LE Becking, GM Bennett, MS Brewer, et al. 2015. “Community Assembly on Isolated Islands: Macroecology Meets Evolution.” *Global Ecology and Biogeography*.

Rosindell, James, and Albert B. Phillimore. 2011. “A Unified Model of Island Biogeography Sheds Light on the Zone of Radiation.” *Ecology Letters* 14 (6): 552–60.

Rosindell, James, Stephen P Hubbell, and Rampal S Etienne. 2011. “*The Unified Neutral Theory of Biodiversity and Biogeography* at Age Ten.” *Trends in Ecology & Evolution* 26 (7). Elsevier: 340–48.

Roussev, Vassil. 2010. “Data Fingerprinting with Similarity Digests.” In *Advances in Digital Forensics Vi*, 207–26. Springer.

Roy, Kaustuv, Gene Hunt, David Jablonski, Andrew Z Krug, and James W Valentine. 2009. “A Macroevolutionary Perspective on Species Range Limits.” *Proceedings of the Royal Society B: Biological Sciences* 276 (1661). The Royal Society: 1485–93.

Royle, J. A., and R. M. Dorazio. 2008. *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*. Academic Press.

Ruckdeschel, P, M Kohl, T Stabla, and F Camphausen. 2006. “S4 classes for distributions.” *R News*.

Sepkoski, J. J. 1984. “A Kinetic Model of Phanerozoic Taxonomic Diversity. III. Post-Paleozoic Families and Mass Extinctions.” *Paleobiology* 10: 246–67. <http://www.jstor.org/stable/2400399>.

Sepkoski, J. John. 1992. *A Compendium of Fossil Marine Animal Families*. Milwaukee, WI: Milwaukee Public Museum.

Sethna, James. 2006. *Statistical Mechanics: Entropy, Order Parameters, and Complexity*. Vol. 14. Oxford University Press.

Shipley, B, D Vile, and E Garnier. 2006. “From Plant Traits to Plant Communities: A Statistical Mechanistic Approach to Biodiversity.” *Science*.

Simpson, GG. 1953. “Horotely, Bradytely, and Tachytely.” *The Major Features of Evolution*, 313–37.

Skilling, J. 1988. *Maximum Entropy and Bayesian Methods in Science and Engineering, Volume 1, ed*. GJ Erickson; CR Smith.

Slatkin, Montgomery. 1987. “Gene Flow and the Geographic Structure of Natural Populations.” *Science* 236 (4803). American Association for the Advancement of Science: 787–92.

Slatyer, Rachel A, Megan Hirst, and Jason P Sexton. 2013. “Niche Breadth Predicts Geographical Range Size: A General Ecological Pattern.” *Ecology Letters* 16 (8). Wiley Online Library: 1104–14.

Solé, R. V., S. C. Manrubia, M. Benton, and P. Bak. 1997. “Self-Similarity of Extinction Statistics in the Fossil Record.” *Nature* 388: 764–67.

Suweis, Samir, Filippo Simini, Jayanth R Banavar, and Amos Maritan. 2014. “Emergence of Structural and Dynamical Properties of Ecological Mutualistic Networks.” *Nature* 500 (7463). Nature Publishing Group: 449–52.

Taberlet, Pierre, Eric Coissac, Francois Pompanon, Christian Brochmann, and Eske Willerslev. 2012. “Towards Next-Generation Biodiversity Assessment Using DNA Metabarcoding.” *Molecular Ecology* 21 (8). Wiley Online Library: 2045–50.

Thompson, John N, Christopher Schwind, Paulo R Guimarães, and Magne Friberg. 2013. “Diversification Through Multitrait Evolution in a Coevolving Interaction.” *Proceedings of the National Academy of Sciences* 110 (28). National Acad Sciences: 11487–92.

Tilman, David. 2004. “Niche Tradeoffs, Neutrality, and Community Structure: A Stochastic Theory of Resource Competition, Invasion, and Community Assembly.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (30). National Acad Sciences: 10854–61.

Ulrich, Werner, Mário Almeida-Neto, and Nicholas J Gotelli. 2009. “A Consumer’s Guide to Nestedness Analysis.” *Oikos* 118 (1). Wiley Online Library: 3–17.

Vázquez, D. P., N. Blüthgen, L. Cagnolo, and N. P. Chacoff. 2009. “Uniting Pattern and Process in Plant-Animal Mutualistic Networks: A Review.” *Annals of Botany* 103 (9): 1445–57.

Vázquez, D.P., R. Poulin, B.R. Krasnov, and G.I. Shenbrot. 2005. “Species Abundance and the Distribution of Specialization in Host-Parasite Interaction Networks.” *Journal of Animal Ecology* 74 (5): 946–55.

Vermeij, G. J. 1987. *Evolution and Escalation*. Princeton, N.J.: Princeton University Press.

Vitousek, Peter Morrison. 2004. *Nutrient Cycling and Limitation: Hawai’i as a Model System*. Princeton University Press.

Vrba, Elisabeth S. 1983. “Macroevolutionary Trends: New Perspectives on the Roles of Adaptation and Incidental Effect.” *Science* 221 (4608): 387–89.

Wagner, W.L., and V. Funk. 1995. *Hawaiian Biogeography Evolution on a Hot Spot Archipelago*. Washington, DC: Smithsonian Institution Press.

Warren, Ben H, Daniel Simberloff, Robert E Ricklefs, Robin Aguilée, Fabien L Condamine, Dominique Gravel, Hélène Morlon, et al. 2015. “Islands as Model Systems in Ecology and Evolution: Prospects Fifty Years After MacArthur-Wilson.” *Ecology Letters* 18 (2). Wiley Online Library: 200–217.

White, Ethan P, SK Morgan Ernest, Andrew J Kerkhoff, and Brian J Enquist. 2007. “Relationships Between Body Size and Abundance in Ecology.” *Trends in Ecology & Evolution* 22 (6). Elsevier: 323–30.

White, Ethan P, Katherine M Thibault, and Xiao Xiao. 2012. “Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model.” *Ecology* 93 (8): 1772–78.

Whittaker, Robert J, Kostas A Triantis, and Richard J Ladle. 2008. “A General Dynamic Theory of Oceanic Island Biogeography.” *Journal of Biogeography* 35 (6). Wiley Online Library: 977–94.

Wiens, John J, R Alexander Pyron, and Daniel S Moen. 2011. “Phylogenetic Origins of Local-Scale Diversity Patterns and the Causes of Amazonian Megadiversity.” *Ecology Letters* 14 (7). Wiley Online Library: 643–52.

Williams, Richard J. 2010. “Simple MaxEnt Models Explain Food Web Degree Distributions.” *Theoretical Ecology* 3 (1). Springer: 45–52.

Williams, Richard J, and Neo D Martinez. 2000. “Simple Rules Yield Complex Food Webs.” *Nature* 404 (6774). Nature Publishing Group: 180–83.

Williams, RJ. 2010. “Simple MaxEnt models explain food web degree distributions.” *Theoretical Ecology* 3 (1): 45–52.

Xiao, Xiao, Daniel J McGlinn, and Ethan P White. 2015. “A Strong Test of the Maximum Entropy Theory of Ecology.” *The American Naturalist* 185 (3). JSTOR: E70–80.

Yeakel, Justin D, Paulo R Guimarães, Hervé Bocherens, and Paul L Koch. 2013. “The Impact of Climate Change on the Structure of Pleistocene Food Webs Across the Mammoth Steppe.” *Proceedings of the Royal Society of London B: Biological Sciences* 280 (1762). The Royal Society: 20130239.