Linking evolutionary and ecological theory illuminates non-equilibrium biodiversity

1 **Abstract**

2 Whether or not biodiversity dynamics tend toward stable equilibria remains an unsolved question
3 in ecology and evolution with important implications for our understanding of diversity and
4 its conservation. Phylo/population genetic models and macroecological theory represent two
5 primary lenses through which we view biodiversity. While phylo/population genetics provide
6 an averaged view of changes in demography and diversity over timescales of generations to
7 geological epochs, macroecology provides an ahistorical description of commonness and rarity
8 across contemporary co-occurring species. Our goal is leverage advances in community-wide high
9 throughput sequencing technology, specifically metabarcoding, to combine evolutionary and
10 macroecological approaches to gain novel insights into the non-equilibrium nature of biodiversity.
11 We help guide near future research with a call for bioinformatic advances and an outline of
12 quantitative predictions made possible by our approach.

13

14

15 Non-equilibrium, inference, and theory in ecology and evolution

16 The idea of an ecological and evolutionary equilibrium has pervaded studies of biodiversity from geological to
17 ecological, and from global to local (Sepkoski 1984; Chesson 2000; Hubbell 2001; Tilman 2004; Rabosky 2009;
18 Harte 2011). The consequences of non-equilibrium dynamics for biodiversity are not well understood and the
19 need to understand them is critical with anthropogenic pressures forcing biodiversity into states of rapid
20 transition (Blonder *et al.* 2015). Non-equilibrial processes could profoundly inform conservation in ways only
21 just beginning to be explored (Wallington *et al.* 2005).

22 Biodiversity theories based on assumptions of equilibrium, both mechanistic (Chesson 2000; Hubbell 2001;
23 Tilman 2004) and statistical **(see the Glossary)** (Pueyo *et al.* 2007; Harte 2011) have found success in
24 predicting ahistorical patterns of diversity such as the species abundance distribution (Hubbell 2001; Harte
25 2011; White *et al.* 2012) and the species area relationship (Hubbell 2001; Harte 2011). These theories assume
26 a macroscopic equilibrium in terms of these coarse-grained metrics, as opposed to focusing on details of
27 species identity (such as in Blonder *et al.* (2015)), although macroscopic and microscopic approaches are not
28 mutually exclusive. Nonetheless, the equilibrium assumed by these theories is not realistic (Ricklefs 2006),
29 and many processes, equilibrial or otherwise, can generate the same macroscopic, ahistorical predictions
30 (McGill *et al.* 2007).

31 Tests of equilibrial ecological theory alone will not allow us to identify systems out of equilibrium, nor permit
32 us to pinpoint the mechanistic causes of any non-equilibrial processes. The dynamic natures of evolutionary
33 innovation and landscape change suggest that ecological theory could be greatly enriched by synthesizing its
34 insights with inference from population genetic theory that explicitly accounts for history. This would remedy
35 two shortfalls of equilibrial theory: 1) if theory fits observed ahistorical patterns but the implicit dynamical
36 assumptions were wrong, we would make the wrong conclusion about the equilibrium of the system; 2) if
37 theories do not fit the data we cannot know why unless we have a perspective on the temporal dynamics
38 underlying those data.

39 No efforts to date have tackled these challenges. We propose that combining insights from ecological theory
40 and inference of evolutionary and demographic change from genetic data will allow us to understand and

predict the consequences of non-equilibrial processes in governing the current and future states of ecological assemblages. The advent of high throughput sequencing-enabled metabarcoding has made unprecedented data available about the biodiversity of lineages from microbes to arthropods (Taberlet *et al.* 2012; Ji *et al.* 2013; Zhou *et al.* 2013; Bohmann *et al.* 2014; Gibson *et al.* 2014; Dodsworth 2015; Leray & Knowlton 2015; Linard *et al.* 2015; Shokralla *et al.* 2015; Venkataraman *et al.* 2015; Liu *et al.* 2016). These metabarcoding data have great potential to yield synthetic insight across ecology and evolution. However, to draw such insights we need a new tool set of bioinformatic methods (**Box box:dry**) and meaningful predictions (**section ref sec:pred**) grounded in theory to make use of those data. We present the foundation of this methedological tool set here.

Ecological theories and non-equilibrium

Neutral and statistical theories in ecology focus on macroscopic patterns, and equilibrium is presumed to be relevant to those patterns. Our goal throughout is not to validate neutral or statistical theories—quite the opposite, we propose new data dimensions, namely genetics, to help better test alternative hypotheses against these null theories, thereby gaining insight into what non-neutral and non-statistical mechanisms are at play in systems of interest.

Non-neutral and non-statistical models (e.g., Tilman (2004); Chesson (2000)) also invoke ideas of equilibrium in their derivation. However, these equilibria focus on the micro-scale details of species interactions, which could in fact lead to instability and non-equilibrium at larger scales (cite: romGEB, critical transition in food web stuff from Beth, random matrix theory). Thus, detailed species interactions could in fact be drivers of non-equilibrium and thus interesting hypotheses to test as alternatives to neutral/statistical models.

To use neutral/statistical theories as null models, we need a robust measure of goodness of fit. The emerging consensus is that likelihood-based test statistics should be preferred (Baldridge *et al.* 2016). The "exact test" of Etienne (2007) has been extended by Rominger & Merow (2017) into a simple z-score which can parsimoniously describe the goodness of fit between theory and pattern. We advocate its use in our proposed framework.

The unified neutral theory of biodiversity (UNTB) (Hubbell 2001) is a useful null because it assumes that one mechanism—demographic drift—drives community assembly. Equilibrium occurs when homogeneous stochastic processes of birth, death, speciation and immigration have reached stationarity. Thus neutrality in ecology is analogous to neutral drift in population genetics (Hubbell 2001) (also cite molecular neutral theory).

Rather than assuming any one mechanism dominates the assembly of populations into a community, statistical theories assume all mechanisms could be valid, but their unique influences have been lost to the enormity of the system and thus the outcome of assembly is a community in statistical equilibrium [Harte (2011); pueyo2007]. The mechanistic agnosticism is what makes statistical theories useful nulls. These statistical theories are also consistent with niche-based equilibria (Pueyo *et al.* 2007; Neill *et al.* 2009) if the complicated, individual or population level models with many mechanistic drivers were to be upscaled to entire communities.

The maximum entropy theory of ecology (METE) (Harte 2011) derives its predictions by condensing the many bits of mechanistic information down into ecological state variables and then mathematically maximizing information entropy conditional on those state variables. METE can predict multiple ahistorical patterns, including distributions of species abundance, body size, spatial aggregation, and trophic links (Harte 2011; Rominger *et al.* 2015; Rominger & Merow 2017), making for a stronger null theory (McGill 2003). However, multiple dynamics can still map to this handful of metrics (McGill *et al.* 2007) and while extensive testing often supports METE's predictions [Harte (2011); white2012; xiao2015] at single snapshots in time, METE fails to match observed patterns in disturbed and rapidly evolving communities [Rominger *et al.* (2015); harte2011]. Just like deviations from UNTB, we cannot know the cause of these departures from theoretical predictions without adding tests of theories and metrics that capture temporal dynamics.

Inferring non-equilibrium dynamics

Unlocking insight into the dynamics underlying community assembly is key to overcoming the limitations of analyzing ahistorical patterns with equilibrial theory. While the fossil record could be used for this task, it has limited temporal, spatial, and taxonomic resolution. Here we instead focus on population/phylogenetic insights into rates of change of populations and species because while there are real limitations in the accuracy and resolution of temporal dynamics with population genetic (cite) and phylogenetic (cite) methods, they can, in principle, be applied to any extant group. Additionally, despite limitations in resolving detailed temporal dynamics, robust metrics of deviation from simple, stationary birth-death and/or speciation-extinction processes have been well-established and widely used for population genetic and phylogenetic data (cite about Tajima D, gamma stats, etc).

Current efforts to integrate evolution into ecological theory

While quantitatively integrating theory from ecology, population genetics, and phylogenetics has not yet been achieved, existing efforts to synthesize perspectives from evolution and ecology point toward promising directions despite being hindered by one or more general issues: 1) lack of a solid theoretical foundation, 2) inability to distinguish multiple competing alternative hypotheses, 3) lack of comprehensive genetic/genomic data, and 4) lack of bioinformatic approaches to resolve species and their abundances with high throughput sequencing data.

Phylogenetic information has been incorporated into studies of the UNTB to better understand its ultimate equilibrium (Jabot & Chave 2009; Burbrink *et al.* 2015). However, phylogenetic reasoning also points out the flaws in the UNTB's presumed equilibrium (Ricklefs 2006). Attempts to correct the assumed dynamics of UNTB through "protracted speciation" (Rosindell *et al.* 2010) are promising, and while their implications for diversification have been considered (Etienne & Rosindell 2011), these predictions have not been integrated with demographic and phylogeographic approaches (e.g., (Edwards & Beerli 2000; Charlesworth 2010; Prado-Martinez *et al.* 2013)) that have the potential to validate or falsify presumed mechanisms of lineage divergence. Such demographic studies, particularly phylogeographic investigations of past climate change, have highlighted the non-equilibrium responses of specific groups to perturbations (Hickerson & Cunningham 2005; Smith *et al.* 2012), but no attempt has been made to scale up these observations to implications at the level of entire communities. The recent growth in joint studies of genetic and species diversity (Vellend 2005; Papadopoulou *et al.* 2011) have been useful in linking population genetics with ecological and biogeographic concepts.

Studies have also used chronosequences or the fossil record in combination with neutral and/or statistical theory to investigate changes over geologic time in community assembly mechanisms (Wagner *et al.* 2006; Rominger *et al.* 2015). While these studies have documented interesting shifts in assembly mechanisms, including departures from equilibrium likely resulting from evolutionary innovations, understanding exactly how the evolution of innovation is responsible for these departures cannot be achieved without more concerted integration with insights from evolutionary theory.

Data and Code Availability

All data and code needed to reproduce the results of this manuscript are available at https://github.com/ecoevomatics/EcoEvoNonStat and a detailed description of the analytical approach is available in the supplement.

References

Baldridge E, Harris DJ, Xiao X, White EP (2016) An extensive comparison of species-abundance distribution models. *PeerJ*, **4**, e2823.

Blonder B, Nogués-Bravo D, Borregaard MK, *et al.* (2015) Linking environmental filtering and disequilibrium to biogeography with a community climate framework. *Ecology*, **96**, 972–985.

Bohmann K, Evans A, Gilbert MTP, *et al.* (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, **29**, 358–367.

Burbrink FT, McKelvy AD, Pyron RA, Myers EA (2015) Predicting community structure in snakes on eastern nearctic islands using ecological neutral theory and phylogenetic methods. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20151700.

Charlesworth D (2010) Don't forget the ancestral polymorphisms. *Heredity*, **105**, 509–510.

Chesson P (2000) Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.*, **31**, 343–366.

Dodsworth S (2015) Genome skimming for next-generation biodiversity analysis. *Trends in plant science*, **20**, 525–527.

Edwards SV, Beerli P (2000) Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.

Etienne RS (2007) A neutral sampling formula for multiple samples and an 'exact' test of neutrality. *Ecology letters*, **10**, 608–618.

Etienne RS, Rosindell J (2011) Prolonging the past counteracts the pull of the present: Protracted speciation can explain observed slowdowns in diversification. *Systematic Biology*, syr091.

Gibson J, Shokralla S, Porter TM, *et al.* (2014) Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences*, **111**, 8007–8012.

Harte J (2011) *The maximum entropy theory of ecology*. Oxford University Press.

Hickerson MJ, Cunningham CW (2005) Contrasting quaternary histories in an ecologically divergent pair of low-dispersing intertidal fish (xiphister) revealed by multi-locus DNA analysis. *Evolution*, **59**, 344–360.

Hubbell SP (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.

Jabot F, Chave J (2009) Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests. *Ecol. Lett.*, **12**, 239–248.

Ji Y, Ashton L, Pedley SM, *et al.* (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology letters*, **16**, 1245–1257.

Leray M, Knowlton N (2015) DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, **112**, 2076–2081.

Linard B, Crampton-Platt A, Gillett CP, Timmermans MJ, Vogler AP (2015) Metagenome skimming of insect specimen pools: Potential for comparative genomics. *Genome biology and evolution*, **7**, 1474–1489.

Liu S, Wang X, Xie L, *et al.* (2016) Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. *Molecular ecology resources*, **16**, 470–479.

McGill B (2003) Strong and weak tests of macroecological theory. *Oikos*, **102**, 679–685.

McGill BJ, Etienne RS, Gray JS, *et al.* (2007) Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.*, **10**, 995–1015.

Neill C, Daufresne T, Jones CG (2009) A competitive coexistence principle? *Oikos*, **118**, 1570–1578.

Papadopoulou A, Anastasiou I, Spagopoulou F, *et al.* (2011) Testing the Species–Genetic diversity correlation in the aegean archipelago: Toward a Haplotype-Based macroecology? *Am. Nat.*, **178**, 241–255.

Prado-Martinez J, Sudmant PH, Kidd JM, *et al.* (2013) Great ape genetic diversity and population history. *Nature*, **499**, 471–475.

Pueyo S, He F, Zillio T (2007) The maximum entropy formalism and the idiosyncratic theory of biodiversity. *Ecology Letters*, **10**, 1017–1028.

Rabosky DL (2009) Ecological limits and diversification rate: Alternative paradigms to explain the variation in species richness among clades and regions. *Ecol. Lett.*, **12**, 735–743.

Ricklefs RE (2006) The unified neutral theory of biodiversity: Do the numbers add up? *Ecology*, **87**, 1424–1431.

Rominger AJ, Goodman KR, Lim JY, *et al.* (2015) Community assembly on isolated islands: Macroecology meets evolution. *Glob. Ecol. Biogeogr.*

Rominger AJ, Merow C (2017) meteR: An r package for testing the maximum entropy theory of ecology.

Methods in Ecology and Evolution, **8**, 241–247.

Rosindell J, Cornell SJ, Hubbell SP, Etienne RS (2010) Protracted speciation revitalizes the neutral theory of biodiversity. *Ecol. Lett.*, **13**, 716–727.

Sepkoski JJ (1984) A kinetic model of phanerozoic taxonomic diversity. III. Post-Paleozoic families and mass extinctions. *Paleobiology*, **10**, 246–267.

Shokralla S, Porter TM, Gibson JF, *et al.* (2015) Massively parallel multiplex DNA sequencing for specimen identification using an illumina MiSeq platform. *Scientific reports*, **5**, 9687.

Smith BT, Amei A, Klicka J (2012) Evaluating the role of contracting and expanding rainforest in initiating cycles of speciation across the isthmus of panama. *Proc. Biol. Sci.*, **279**, 3520–3526.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, **21**, 2045–2050.

Tilman D (2004) Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 10854–10861.

Vellend M (2005) Species diversity and genetic diversity: Parallel processes and correlated patterns. *Am. Nat.*, **166**, 199–215.

Venkataraman A, Bassis CM, Beck JM, *et al.* (2015) Application of a neutral community model to assess structuring of the human lung microbiome. *MBio*, **6**.

Wagner PJ, Kosnik MA, Lidgard S (2006) Abundance distributions imply elevated complexity of post-paleozoic marine ecosystems. *Science*, **314**, 1289–1292.

Wallington TJ, Hobbs RJ, Moore SA (2005) Implications of current ecological thinking for biodiversity conservation: A review of the salient issues. *Ecol. Soc.*, **10**.

White EP, Thibault KM, Xiao X (2012) Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology*, **93**, 1772–1778.

Zhou X, Li Y, Liu S, *et al.* (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *Gigascience*, **2**, 4.