Linking evolutionary and ecological theory illuminates non-stationary biodiversity

A. J. Rominger[1, 2], I. Overcast[3], H. Krehenwinkel[1], R. G. Gillespie[1], J. Harte[1, 4], and M. J. Hickerson[3]

[1]Department of Environmental Science, Policy and Management, University of California, Berkeley
[2]Santa Fe Institute
[3]Biology Department, City College of New York
[4]Energy and Resource Group, University of California, Berkeley

Corresponding author: Rominger, A.J. (ajrominger@gmail.com).

Abstract

Whether or not biodiversity dynamics are governed by stable equilibria remains an unsolved question in ecology and evolution. The existence (or non-existence) of such equilibria has wide ranging implications, for example, whether conservation should focus on conventional preservationist paradigms or adaptive management. Whether biodiversity rapidly and consistently tends toward an equilibrium also determines how species and communities will respond to global environmental change. Inference methods from phylogenetics and population genetics, and macroecological theory represent two primary lenses through which we view biodiversity patterns. Population genetics and phylogenetics provide an integrated view of changes in population demography and lineage origination over scales of generations to geological epochs but cannot be used to directly infer the ecological composition of any given time period. Conversely, macroecology, and its constituent theories, provides an ahistorical description of the commonness and rarity of species co-occurring in the same landscape. With advances in molecular methods allowing genetic data to be gathered across entire communities, the field is ready to not only combine the theoretical insights of phylo/population genetics and ecology, but also test predictions with real data.

1 Equilibrium, inference, and theory in ecology and evolution

We propose that combining insights from ecological theory 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 time is ripe to fully harness the vast amount of genetic and genomic data being generated at unprecedented scales [1–13] to address fundamental questions in ecology and evolution.

The idea of an ecological and evolutionary equilibrium has pervaded studies of biodiversity both on geologic and ecological time scales, and from global to local scales [14–22]. Biodiversity theories based on assumptions of equilibrium, both mechanistic [18, 20, 22] and statistical [see the Glossary; 19, 23, 24] have found success in predicting ahistorical patterns of diversity such as the species abundance distribution [18, 19, 25] and the species area relationship [18, 19]. However, investigation of the underlying dynamics producing these patterns has revealed that the equilibrium assumed by the theories is not realistic [26], and that many processes, equilibrial and otherwise, can generate the same macroscopic, ahistorical predictions [27, 28].

The consequences of non-equilibrium dynamics for biodiversity, from diversification to macroecology to conservation, are not well understood. The need to understand non-equilibrial biodiversity processes comes at a critical time when anthropogenic pressures are forcing biodiversity systems into states of rapid transition [29]. The extent to which ecosystems are governed by non-equilibrial processes has profound implications for conservation, which are only just beginning to be explored [30]. For example whether conservation should focus on conventional preservationist paradigms or adaptive management [30]. Whether biodiversity rapidly and consistently tends toward a steady state also determines how species

and communities will respond to global environmental change [29].

Tests of equilibrial ecological theory alone will not allow us to identify systems out of equilibrium, nor permit us to pinpoint the mechanistic causes of any observed patterns indicating non-equilibrial processes. The dynamic natures of evolutionary innovation and landscape change suggest that ecological theory could be greatly enriched by building a joint modeling framework with population genetic theory that explicitly accounts for historical changes in populations and does not rely on stationarity for generative model predictions. This would remedy two shortfalls of equilibrial theory: 1) if the theory fits observed ahistorical patterns but the implicit dynamical assumptions were wrong, we would make the wrong conclusion about the equilibrium of the system; 2) if the equilibrial theories do not fit the data we cannot know why unless we have a perspective on the temporal dynamics underlying the generation of those data.

Existing efforts to directly infer the evolutionary and demographic dynamics underlying community assembly in the context of ecological theory testing are limited by a lack of data and analytical framework (see section 4). The advent of next generation sequencing approaches to biodiversity [1, 4, 6, 12] have lifted the data barrier, but we need a tool set of bioinformatic methods (see Box 2) and meaningful predictions (see section 6 grounded in theory to make use of those data; we call for and sketch that tool set here.

2 Ecological theories and equilibrium

The development of the equilibrium theory of island biogeography (ETIB; [31]) ushered in the advent of mechanistically elegant, predictive theories of general patterns in biodiversity. The theory of MacArthur and Wilson also set the precedent of focusing on equilibrial predictions for biodiversity, instead of transient states. From this starting point, three classes of ecological theory have emerged, mechanistically niche-based theory, mechanistically neutral theory, and mechanistically agnostic, statistical mechanical theory. We will focus on neutral

67 and statistical equilibria here. In so doing, we treat niches as in effect being drivers of non-

68 equilibrium: powerful niche dynamics prevent a system from attaining a neutral or statistical

69 equilibrium. Modeling niche dynamics is difficult due to the inherent high dimensionality

70 of the parameter space implied by verbal niche models [e.g., 32], thus showing a lack of

71 neutrality or statistical equilibrium is easier than directly demonstrating niche factors. We

72 further explore the consequences of this approach the **Outstanding Questions** box.

2.1 Mechanistically neutral theory

74 Neutral theory [18] assumes that one mechanism—demographic drift—drives community as-

75 sembly. By presuming that populations do not differ in fitness nor in resource use, neutrality

76 avoids the intractability of over parameterization and arrives as an equilibrial prediction when

77 homogeneous stochastic processes of birth, death, speciation and immigration have reached

78 stationarity. Thus neutrality in ecology is analogous to neutral drift in population genetics.

2.2 Statistical theory

80 Rather than assume that any one mechanism, be it niche-based or neutral, dominates the

81 assembly of populations into a community, theories based on statistical mechanics assume

82 that all mechanisms could be valid, but their unique influence has been lost to the enormity

83 of the system and thus outcome of assembly is a community in statistical equilibrium. In

84 one class of such theories, it is assumed that whatever mechanisms are at play, they are

85 only relevant in determining the values of ecological state variables, and then if the system

86 is allowed to come to equilibrium its properties will be predicted by maximizing information

87 entropy relative to the constraints of the state variables. One example is the maximum

88 entropy theory of ecology (METE), one model realization of which assumes that the area

89 (A_0) of an ecosystem, the total number of species (S_0) in some taxonomic group, the total

90 number of individuals in those species (N_0), and the total metabolic rate of those individuals

4

91 (E_0), capture all necessary information to characterize a community because that community
92 has reached a statistical equilibrium in which the imprint of specific mechanistic forces has
93 been lost. While this theory finds widespread success in predicting ahistorical patterns of
94 species abundance, size, and spatial distribution [19, 25, 33, 34] at single snapshots in time,
95 it fails to match observed patterns in disturbed and rapidly evolving communities [19, 35].

3 Inferring non-equilibrium dynamics

97 Unlocking insight into the dynamics underlying community assembly will help us overcome
98 the limitations of analyzing ahistorical patterns with equilibrial theory. We need explicit
99 information about the rates of change of populations and species by processes of demographic
100 fluctuations, immigration, selection, speciation, and extinction.

101 While the fossil record can elucidate deep time patterns for select, well-fossilized groups
102 [36], and in limited geographic areas and temporal extents yielding good preservation [37],
103 we require an approach that is applicable across taxa and scales of space and time. Bridging
104 ecological theory with models from phylogenetics has long given us potential general-use
105 tools to gain insight into the dynamics underlying contemporary biodiversity patterns [38–
106 40], while links from population genetics have been more recently explored [38–47].

107 The potential for such modeling approaches was appreciated early on [48, 49], as it was
108 well recognized that population genetic data from multiple codistributed taxa could aug-
109 ment investigation of traditionally ecologically-centered questions. Yet Despite over 30 years
110 of comparative historical demography studies, there has been almost a wholesale neglect of
111 the growing body of theory from community ecology. Conversely, ecological models of com-
112 munity assembly tend to ignore the history. The opportunity to unify processes underlying
113 patterns of species diversities and abundances with distributions of historical population size
114 trajectories, colonisation times, speciation times and regional patterns of genetic connectivity
115 begs investigation.

One of the fundamental tools allowing for complex historical inference with population genetic data is coalescent theory [50–54]. Now broadly applied, coalescent theory can generate the statistical properties of any sample of alleles across the genome by modeling gene genealogies backwards in time under virtually any complex demographic history thereby allowing model-based estimation of historical parameters such as historical population size fluctuations, divergence and/or colonization times, and migration rates [55]. This approach can also be put in a multi-species, community context thanks to advances in historical demographic models [56–60].

Estimating isolation, divergence and/or speciation times has been a particularly important application of population genetic data, and use of coalescent theory is of notable importance in this endeavor because it statistically captures the stochastic discord between population divergence times and gene divergence times [61, 62]. In the context of island biogeography and invasion ecology, coalescent-based estimates of isolation times is of particular importance for understanding the dynamics and timing of island colonization, intra-island speciation, as well as invasion times [63–66].

The history of population size change is also of fundamental importance for understanding the dynamics of community assembly across a variety of ecological settings, and coalescent theory has likewise become the standard tool for estimating size change histories with population genetic and phylogeographic data on hand [67, 68]. This application of coalescent modeling has been deployed for large numbers of species from which only small numbers of genetic loci are sampled from populations [69]. Pivotal to the understanding of demographic and evolutionary histories, coalescent theory has also allowed modeling complex patterns of historical population structure [70, 71] and gene flow [72, 73].

Taking all of these elements of demographic history together (i.e. structure, divergence, expansion, size change and migration), researcher, simulation-based coalescent approaches such as approximate Bayesian computation [74, 75] have become of notable importance for

making statistical inference under complex histories when solving the likelihood function becomes intractable [76].

As important as it is for the inference of complex demographic history, coalescent theory has also become an important modeling tool for understanding how natural selection shapes patterns of genetic polymorphism [77–79]. Indeed, one of the most commonly used techniques for detecting positive selection relies on a summary statistic, Tajima's D, that can be easily simulated under the coalescent given alternative models with neutrality or selection [80]. However, population genetic models of positive and/or purifying selection also have very similar predicted Tajima's D values to those derived from neutral histories with non-stationary population growth [81–85], as well as other more complex models of selection such as polygenic adaptation and interference selection [86, 87]. Thus Tajima's D can best be seen as a metric that quantifies deviation from demographic equilibrium and used to jointly describe selection and demographic history [78, 79, 88, 89].

Ultimately, it is at the community level of inference that coalescent-based population genetic methods could be most useful for investigating ecological models that deviate from stationarity. Indeed, it is the inherent historical approach enabled by coalescent methods that can potentially enrich the ecological theoretical approaches to community assembly and stationarity.

4 Current efforts to integrate evolution into ecological theory

Current efforts to synthesize theoretical perspectives from evolution and ecology have made substantial contributions toward understanding what drives biodiversity patterns. However, a more concerted integration is needed, and indeed was not even feasible until recent and ongoing genetic, bioinformatic and theoretical advances. Approaches to date have been

hindered by one or more of several general issues: 1) lack of a solid theoretical foundation, 4) inability to distinguish multiple competing alternative hypotheses, 3) lack of comprehensive genetic data, 4) lack of bioinformatic approaches to resolve species and their abundances. Here we quickly survey the ways these shortcomings have prevented further advances and then move on to the cutting edge of the field.

Community phylogenetics [38] attempted to understand the roles of competition and environmental filtering on community assembly by assuming key ecologically-relevant traits are conserved along phylogenies; without a solid theory of trait-mediated competition and recruitment, nor a solid theory of trait evolution, this program broke down [90]. Largely lost is the opportunity to use phylogenetic information to understand the historical contingencies at play in community assembly [91, 92], a task which phylogenies might be able to perform, while they are often poor proxies for traits [90].

Joint studies of genetic and species diversity [44, 46, 93, 94] are largely correlative, lacking a strong theoretical core that could be used to make testable predictions. These studies also miss the opportunity to explore more than just diversity metrics, but full models of community assembly, population demography and molecular evolution. These studies are also held back by limited access to genetic data, a hurdle which is currently overcome by the development of next generation sequencing based community analysis approaches (see Box 1).

Phylogeographic studies of past climate change have provided insights of how specific groups have responded in non-equilibrium ways to perturbations [60, 95–97], but such studies cannot make inference about entire community-level processes, nor were they designed to tie into ecological theory and have not been leveraged by theoreticians to gain more realistic insights on demographic and geographic change.

Ecological neutral theory applied to microbial communities [98] has demonstrated that the same ecological processes that operate at macro-scales may also scale down to commu-

8

nities of microbes. However, such studies have not made use of the immense phylogenetic and functional genomic resources available for microbes. Nor has the problem of inferring abundance from metagenomic and metabarcoding data been fully resolved (see Box 2).

4.1   Emerging approaches

Several approaches have been taken that better ground synthesis of ecological and evolutionary dynamics in rigorous theory. Because the NTB is implicitly an evo-ecological theory [18, 99], despite typically being treated as ahistorical, it is natural to include evolutionary information into inference about the theory's parameters. Etienne cast the solution of the NTB's species abundance distribution as a coalescent problem [100] while Jabot and Chave [101] used approximate Bayesian computation to improve estimates of the NTB's fundamental biodiversity number using phylogenetic information. Efforts have also been made to validate the underlying assumption of ecological equivalence, a key assumption of the NTB, from a phylogenetic perspective [102]. While these efforts improved inference of the parameters involved in making ahistorical predictions of species abundance, they did not aim to improve the underlying realism of the evolutionary dynamic presumed by the NTB. For example, while the NTB accurately predicts phylogenetic tree shape (sensu [101]) it does not accurately reflect tree tempo [103]. The time to equilibration in the NTB is also unrealistically long [26]. While protracted speciation has been proposed to correct some of these tempo problems in the NTB [104], it remains to be tested, by a framework such as the one we propose, whether these theoretical advances can accurately predict joint patterns of population genetics, phylogenies, and communities.

Another approach has tested the ahistorical predictions of equilibrial ecological theory through evolutionary snapshots of community assembly and change. Several applications of the NTB in the fossil record have been used to show changes over geologic time in community assembly mechanisms [105, 106]. In a similar theme, Rominger et al. [35] used the

geologic chronosequence of the Hawaiian Islands in combination with METE to investigate how evolutionary changes in community assembly drove non-equilibrial patterns in networks of plants and herbivorous insects. While Rominger et al. used genetic information to understand how evolutionary rates vary between different arthropod clades in response to the geologic chronosequence, these evolutionary snapshot studies lack a quantitative reconciliation of mechanisms inferred by analyses of ahistorical theory with independently inferred dynamics, either from genetic data or stratigraphic time series.

5   What is needed now

A key limitation of using ahistorical theory to infer dynamic mechanisms is that multiple mechanisms, from simple and equilibrial to complex, can map onto the same ahistorical pattern, such as the species abundance distribution [107–111]. This means that even when a theory describes the data well, we do not really know the dynamics that led to that good fit— an interpretational pitfall common in many studies that claim mechanistic insight even in novel evolutionary study systems [18, 105, 106]. Studies that do not have a strong theoretical foundation, and instead rely on qualitative predictions such as higher or lower phylogenetic dispersion [38], further exacerbate the problem of many mechanisms mapping onto single phenomenological predictions.

Quantitative theoretical foundations and direct information about dynamics can break this many-to-one mapping of mechanism onto theoretical prediction. This nicely parallels calls to incorporate additional information into community ecology and macroecological studies [112]. We propose here a needed framework for integrating the dynamics inferred from population and phylogenetic approaches with ahistorical, equilibrial ecological theory. There are two complementary options for incorporating the insights of both ahistorical ecological theory and genetic inference methods:

- Option 1: using dynamics from genetic inference to predict and understand deviations from ahistorical theories. This amounts to separately fitting ahistorical theory to typical macroecological data (often species abundance, but potentially including body size and trophic network links) while also fitting population genetic and/or phylogenetic models to genetic data simultaneously captured for the entire community. Doing so requires serious bioinformatic advances that would allow the joint capture of genetic or genomic data from entire community samples, while also estimating accurate abundances for each species in those same samples. This is discussed in Box 2.

- Option 2: building off existing theories, develop new joint models that simultaneously predict macroecological and population genetic patterns. This amounts to building hierarchical models that take genetic data as input and integrate over all possible community states given explicit models of community assembly and population coalescence. Such a model approach also represents a major bioinformatic challenge, which is also discussed in Box 2.

5.1   What we could gain from this framework

Given the insights that could be gained from either option 1 or 2 above, we could finally understand why ahistorical theories fail when they do—is it because of rapid population change, or evolution/long-distance dispersal of novel ecological strategies? We could predict whether a system that obeys the ahistorical predictions of equilibrial ecological theory is in fact undergoing major non-equilibrial evolution. We could better understand and forecast how/if systems out of equilibrium are likely to relax back to equilibrial patterns. With such a framework we could even flip the direction of causal inference and understand ecological drivers of diversification dynamics. This last point bears directly on long-standing and open debates about the importance of competitive limits on diversification. Competition and limiting similarity have a long history of study as drivers of diversification. This has culminated

in ideas of diversity-dependent diversification[113–115]. What has not been done is link this back to ecological assembly mechanisms, but the opportunity seems ripe considering the abundance of work on niche differences and fitness differences[20, 116–118]. There has even been work on this from a phylogenetic viewpoint[119, 120]. Conclusions about phylogenetic patterns (e.g. diversification slowdowns) would be more believable and robust if combined with population genetic inference (e.g. declining populations) and community patterns (e.g. deviation from equilibrium).

6 Evo-ecological predictions for systems out of equilibrium

The data needed to fully test a non-equilibrial theory of ecology and evolution, synthesizing historical and contemporary biodiversity patterns, are unprecedented in scale and depth. Put simply, we require knowing the species identities of each individual in a sample as well as information on some portion of their genomes such that we can estimate historical demography and diversification. In Box 2 we discuss progress toward generating such data. We highlight two promising routes: 1) estimating abundance from targeted capture high throughput sequencing data (i.e. metabarcoding) to be used in ahistorical ecological theory testing, and then separately fitting models of demography and diversification; and 2) jointly estimating the parameters of coupled models of community assembly and population demographics. Assuming these two approaches are within reach (as we demonstrate in **Bioinformatic advances**), we now discuss hypotheses to be tested in our non-equilibrium framework.

6.1 Cycles of non-equilibrium

Ecosystems are likely to experience consistence disturbances which can occur on ecological time-scales, such as primary success, or evolutionary time scales, such as evolution of novel innovations that lead to new ecosystem processes [121, 122]. We hypothesize that these consistence disturbances will lead to cycles of non-equilibrium in observed biodiversity patterns.

Figure 1 shows how we hypothesize biodiversity to transition between different phases of equilibrium and non-equilibrium. The figure presents deviation between data and ahistorical ecological theory on the y-axis, versus deviations from demographic or evolutionary equilibrium on the x-axis. A clockwise cycle through this space would indicate:

- Panel I → Panel II: following rapid ecological disturbance, ecological metrics diverge from equilibrium

- Panel II → III: ecological non-equilibrium spurs evolutionary non-equilibrium leading to both ecological and evolutionary metrics diverging from equilibrium values

- Panel III → IV: ecological relaxation to equilibrium after evolutionary innovations provide the means for populations to re-equilibrate to their environments

- Panel IV → I: finally a potential return to equilibrium of both ecological and evolutionary metrics once evolutionary processes have also relaxed to their equilibrium

Cycles could also be much shorter, with a system only transitioning back and forth between Panel I and Panel II. This scenario corresponds to the system being driven only by rapid ecological disturbance, and this disturbance itself following a stationary dynamic leading to no net evolutionary response.

Cycles through this space could also occur in a counterclockwise direction, being initiated by an evolutionary innovation. Under such a scenario we hypothesize the cycle to proceed:

- Panel I → IV: non-equilibrium evolution (including sweepstakes dispersal) leading to departure from evolutionary equilibrium before departure from ecological equilibrium

- Panel IV → III: non-equilibrial ecological response to non-equilibrium evolutionary innovation

- Panel III → I: ecological and evolutionary relaxation

We hypothesize that the final transition will be directly to a joint equilibrium in ecological and evolutionary metrics (Panel I) because a transition from panel III to panel II is unlikely, given that ecological rate are faster than evolutionary rates.

A complete cycle cannot be observed without a time machine, but by combining ahistorical ecological theory and population/phylogenetic inference methods with community-level genetic data we can identify where on the cycle our focal systems are located. Such an approach assumes that abundance data have been estimated from sequence data, ahistorical ecological theories have been fit to those abundance data, and models of population demography and/or diversification have been separately fit to the underlying sequence data. To better under how our focal systems have transitioned between different equilibrium and non-equilibrium phases, we must more deeply explore the joint inference of community assembly and evolutionary processes. In the following sections we do just that for each transition shown in Figure 1. We bring to bear other aspects of joint eco-evolutionary inference, in particular the 1) relationship between lineage age (inferred from molecular data) and lineage abundance; 2) the nature of deviation from ecological metrics, specifically the shape of the species abundance distribution; and 3) the nature of deviation from evolutionary metrics, specifically inference of past population change and selection.

6.2 Systems undergoing rapid ecological change

For systems whose metrics conform to demographic predictions of equilibrium, but deviate from equilibrial ecological theory, we predict that rapid ecological change underlies their dynamics. However, more information is needed to confirm that the system is being driven primarily by rapid ecological change. The first line of evidence could come from a lack of correlation between lineage age and lineage abundance—this would indicate that slow eco-evolutionary drift is interrupted by frequent perturbations to populations, making their size independent of age (Fig. 2). Actual abundance should similarly be uncorrelated with inference of effective population size from genetic data. Further support for the ecology-only hypothesis could come from a lack of directional selection detected in community-wide surveys of large genomic regions (see Box 1); however, this is not yet feasible at the scale of entire community samples. Taken as a whole, systems in which ecological metrics deviate from equilibrial theory while demographic and macroevolutionary metrics conform to equilibrial theory presents an opportunity to understand and test hypotheses relating to disturbance, assembly, and the shape of the species abundance distribution [e.g., 19].

6.3 Non-equilibrium ecological communities fostering non-equilibrium evolution

A lack of equilibrium in an ecological assemblage means that the system will experience change on its trajectory toward a future possibility of equilibrium. If ecological relaxation does not occur—by chance, or because no population present is equipped with the adaptations to accommodate the new environmental that drove the system from equilibrium to begin with—then the system is open to evolutionary innovation. Such innovation could take the form of elevated speciation or long-distance immigration of populations with different ecological strategies than those present, relating to the idea that community assembly is a race between processes with potentially different, but stochastic rates [44], thus any process

might by chance produce the observed outcome. In this way, speciation and sweepstakes immigration (and its human-added counterpart, invasion) represent similar processes. Speciation and sweepstakes/invasion will yield very different phylogenetic signals, however their population genetic signals in a non-equilibrium community may be very similar (e.g. rapid population expansion). Thus where non-equilibrium communities foster non-equilibrium diversification (either through speciation or invasion) we expect to see a negative relationship between lineage age and abundance (Fig. 2) and signs of rapid population expansion in the newest arrivals to the system.

Local extinctions results from non-equilibrium community dynamics are unlikely to scale up to non-equilibrial macroevolution, so long as the extinctions are not global. For such ecosystem collapse to occur, continental ecosystems would have to be self-organized to the point of criticality, a hypothesis which has largely been refuted [123].

6.4 Non-equilibrium evolution fostering non-equilibrium ecological dynamics

If evolutionary processes, or their counterpart in the form of sweepstakes immigration/invasion, generate new ecological strategies in a community, this itself constitutes a form of disturbance pushing the system to reorganize, thus our predicted cycle in Figure 1 from Panel I to IV to III. Evolutionary change would have to be extremely rapid to force ecological metrics out of equilibrium, because evolutionary change could be accommodated by ecological communities if it occurs slowly enough, thus we would expect to see phylogenetic signals of adaptive radiation, and corresponding signals of strong selection in genomic-scale sequence data.

6.5   Ecological relaxation

Ecological metrics can return to equilibrium either by ecological means (local immigration from the source pool, local extinction, rapid population changes) or by evolutionary means (novel innovations arising through selection and adaptation or sweepstakes immigration). In either case, communities are likely to return to equilibrium given enough time without disturbance. Because ecological rates are typically faster than evolutionary rates, this ecological relaxation is likely to happen more quickly than evolutionary relaxation, and thus genetic inference may reveal a time-averaged demographic signature of non-equilibrium.

6.6   Evolutionary relaxation

Given sufficient time in ecological equilibrium, the time averaged demographic record revealed by genetic inference will likely also re-equilibrate.

7   Harnessing evo-ecological measures of non-equilibrium for a changing world

Inference of community dynamics that accounts for non-equilibrium historical complexities needs to expand empirical dimensions beyond species abundances and diversities to include axes of information that are historically dynamic with respect to generative models that link spatial-temporal processes and regional genetics patterns across and between species. Although comparative historical demographic models are advancing [56–60], testing community-scale hypotheses with milti-taxa data would be profoundly improved and enriched if population genetic model were grounded in macroecological and biogeographic theory. Conversely, it has been long recognized that models in community ecology have been overly reliant on ahistorical patterns, such as the species abundance distribution, which are

17

by themselves often insufficient for distinguishing competing models of assembly [27, 28]. The field is ready to fully merge these two approaches using the wetlab, bioinformatic, and theoretical-conceptual approaches we have promoted here. The time is right to do so as biodiversity scientists must face an increasingly non-equilibrium world and its consequences for our fundamental understanding of what forces govern the diversity of life and how we can best harmonize human activities with it.

References

[1] Yu, D.W. *et al.* (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* 3, 613–623

[2] Pompanon, F. *et al.* (2012) Who is eating what: diet assessment using next generation sequencing. *Molecular ecology* 21, 1931–1950

[3] Taberlet, P. *et al.* (2012) Towards next-generation biodiversity assessment using dna metabarcoding. *Molecular Ecology* 21, 2045–2050

[4] Ji, Y. *et al.* (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology letters* 16, 1245–1257

[5] Zhou, X. *et al.* (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without pcr amplification. *Gigascience* 2, 4

[6] Tang, M. *et al.* (2014) Multiplex sequencing of pooled mitochondrial genomes—a crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic acids research* 42, e166–e166

[7] Bohmann, K. *et al.* (2014) Environmental dna for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution* 29, 358–367

[8] Gibson, J. *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

[9] Shokralla, S. *et al.* (2015) Massively parallel multiplex dna sequencing for specimen identification using an illumina miseq platform. *Scientific reports* 5, 9687

[10] Linard, B. *et al.* (2015) Metagenome skimming of insect specimen pools: potential for comparative genomics. *Genome biology and evolution* 7, 1474–1489

[11] Leray, M. and 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

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

[13] Liu, S. *et al.* (2016) Mitochondrial capture enriches mito-dna 100 fold, enabling pcr-free mitogenomics biodiversity analysis. *Molecular ecology resources* 16, 470–479

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

[15] Alroy, J. (2010) The shifting balance of diversity among major marine animal groups. *Science* 329, 1191–1194

[16] Rabosky, D.L. and Lovette, I.J. (2008) Density-dependent diversification in north american wood warblers. *Proceedings of the Royal Society B: Biological Sciences* 275, 2363–2371

[17] Rabosky, D.L. (2009) Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. *Ecol. Lett.* 12, 735–743

[18] Hubbell, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*, vol. 32. Princeton University Press

[19] Harte, J. (2011) *The Maximum Entropy Theory of Ecology.* Oxford University Press

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

[21] Adler, P.B. *et al.* (2010) Coexistence of perennial plants: an embarrassment of niches. *Ecol. Lett.* , no–no

[22] 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

[23] Pueyo, S. *et al.* (2007) The maximum entropy formalism and the idiosyncratic theory of biodiversity. *Ecol. Lett.* 10, 1017–1028

[24] Shipley, B. *et al.* (2006) From plant traits to plant communities: A statistical mechanistic approach to biodiversity. *Science*

[25] White, E.P. *et al.* (2012) Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology* 93, 1772–1778

[26] Ricklefs, R.E. (2006) The unified neutral theory of biodiversity: do the numbers add up? *Ecology* 87, 1424–1431

[27] McGill, B.J. *et al.* (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* 10, 995–1015

[28] McGill, B.J. (2010) Towards a unification of unified theories of biodiversity. *Ecology Letters* 13, 627–642

[29] Barnosky, A.D. *et al.* (2012) Approaching a state shift in earth/'s biosphere. *Nature* 486, 52–58

[30] Wallington, T.J. *et al.* (2005) Implications of current ecological thinking for biodiversity conservation: A review of the salient issues. *Ecol. Soc.* 10

[31] MacArthur, R.H. and Wilson, E.O. (1967) *The theory of island biogeography.* Princeton University Press

[32] Hutchinson, G.E. (1959) Homage to santa rosalia or why are there so many kinds of animals? *The American Naturalist* 93, 145–159

[33] Xiao, X. *et al.* (2015) A strong test of the maximum entropy theory of ecology. *Am. Nat.* 185, E70–E80

[34] Harte, J. *et al.* (2009) Biodiversity scales from plots to biomes with a universal species-area curve. *Ecol. Lett.* 12, 789–797

[35] Rominger, A.J. *et al.* (2015) Community assembly on isolated islands: macroecology meets evolution. *Glob. Ecol. Biogeogr.*

[36] Alroy, J. *et al.* (2008) Phanerozoic trends in the global diversity of marine invertebrates. *Science* 321, 97–100

[37] Harnik, P.G. (2011) Direct and indirect effects of biological factors on extinction risk in fossil bivalves. *Proceedings of the National Academy of Sciences* 108, 13594–13599

[38] Webb, C.O. *et al.* (2002) Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* 33, 475–505

[39] Emerson, B.C. (2002) Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. *Mol. Ecol.* 11, 951–966

[40] Lavergne, S. *et al.* (2010) Biodiversity and climate change: Integrating evolutionary and ecological responses of species and communities. *Annu. Rev. Ecol. Evol. Syst.* 41, 321–350

[41] Li, L. and Ma, Z.S. (2016) Testing the neutral theory of biodiversity with human microbiome datasets. *Sci. Rep.* 6, 31448

[42] McGaughran, A. (2015) Integrating a population genomics focus into biogeographic and macroecological research. *Front. Ecol. Evol.* 3

[43] Laroche, F. *et al.* (2015) A neutral theory for interpreting correlations between species and genetic diversity in communities. *Am. Nat.* 185, 59–69

22

[44] Vanoverbeke, J. *et al.* (2015) Community assembly is a race between immigration and adaptation: eco-evolutionary interactions across spatial scales. *Ecography*

[45] Vellend, M. and Geber, M.A. (2005) Connections between species diversity and genetic diversity. *Ecol. Lett.* 8, 767–781

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

[47] Dexter, K.G. *et al.* (2012) Historical effects on beta diversity and community assembly in amazonian trees. *Proc. Natl. Acad. Sci. U. S. A.* 109, 7787–7792

[48] Avise, J.C. *et al.* (1987) Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* 18, 489–522

[49] Avise, J.C. (1998) The history and purview of phylogeography: a personal reflection. *Mol. Ecol.* 7, 371–379

[50] Hudson, R.R. (1983) Properties of a neutral model with intragenic recombination. *Theor. Popul. Biol.* 23, 183–201

[51] Tajima, F. (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105, 437–460

[52] Kingman, J.F.C. (1982) The coalescent. *Stochastic Process. Appl.* 13, 235–248

[53] Kingman, J.F.C. (1982) On the genealogy of large populations. *J. Appl. Probab.* 19a, 27–43

[54] Rosenberg, N.A. and Nordborg, M. (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet.* 3, 380–390

[55] Wakeley, J. (2008) *Coalescent Theory.* Roberts and Company Publishers

[56] Xue, A.T. and Hickerson, M.J. (2015) The aggregate site frequency spectrum (aSFS) for comparative population genomic inference. *Mol. Ecol.*

[57] Hickerson, M.J. *et al.* (2006) Test for simultaneous divergence using approximate bayesian computation. *Evolution* 60, 2435–2453

[58] Carstens, B.C. *et al.* (2016) Community trees: Identifying codiversification in the páramo dipteran community. *Evolution* 70, 1080–1093

[59] Chan, Y.L. *et al.* (2014) Detecting concerted demographic response across community assemblages using hierarchical approximate bayesian computation. *Mol. Biol. Evol.* , msu187

[60] Satler, J.D. and Carstens, B.C. (2016) Phylogeographic concordance factors quantify phylogeographic congruence among co-distributed species in the sarracenia alata pitcher plant system. *Evolution*

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

[62] Edwards, S.V. and Beerli, P. (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* 54, 1839–1854

[63] Estoup, A. and Clegg, S.M. (2003) Bayesian inferences on the recent island colonization history by the bird zosterops lateralis lateralis. *Mol. Ecol.* 12, 657–674

[64] Estoup, A. *et al.* (2004) Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, bufo marinus. *Evolution* 58, 2021–2036

[65] Hickerson, M.J. and Meyer, C. (2008) Testing comparative phylogeographic models of marine vicariance and dispersal using a hierarchical bayesian approach. *BMC Evol. Biol.* 8, 322

[66] Gray, M.M. *et al.* (2014) Demographic history of a recent invasion of house mice on the isolated island of gough. *Mol. Ecol.* 23, 1923–1939

[67] Kuhner, M.K. *et al.* (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* 149, 429–434

[68] Slatkin, M. and Hudson, R.R. (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129, 555–562

[69] Drummond, A.J. *et al.* (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22, 1185–1192

[70] Prado-Martinez, J. *et al.* (2013) Great ape genetic diversity and population history. *Nature* 499, 471–475

[71] Bahlo, M. and Griffiths, R.C. (2000) Inference from gene trees in a subdivided population. *Theor. Popul. Biol.* 57, 79–95

[72] Beerli, P. and Felsenstein, J. (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4563–4568

[73] Hey, J. and Nielsen, R. (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of drosophila pseudoobscura and d. persimilis. *Genetics* 167, 747–760

[74] Beaumont, M. (2010) Approximate bayesian computation in evolution and ecology. *Annu. Rev. Ecol. Evol. Syst.* 41, 379–406

[75] Pritchard, J.K. *et al.* (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* 16, 1791–1798

[76] Sunnåker, M. *et al.* (2013) Approximate bayesian computation. *PLoS Comput. Biol.* 9, e1002803. doi:10.1371/journal.pcbi.1002803

[77] Kim, Y. and Stephan, W. (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160, 765–777

[78] Kern, A.D. and Schrider, D.R. (2016) Discoal: flexible coalescent simulations with selection. *Bioinformatics* 32, 3839–3841

[79] Ewing, G.B. and Jensen, J.D. (2016) The consequences of not accounting for background selection in demographic inference. *Mol. Ecol.*

[80] Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595

[81] Freedman, A.H. *et al.* (2016) Demographically-based evaluation of genomic regions under selection in domestic dogs. *PLoS Genet.* 12, e1005851

[82] Barton, N.H. (1998) The effect of hitch-hiking on neutral genealogies. *Genet. Res.* 72, 123–133

[83] Barton, N.H. (2000) Genetic hitchhiking. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 1553–1562

[84] Jensen, J.D. *et al.* (2005) Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics* 170, 1401–1410

[85] Schrider, D.R. *et al.* (2016) Effects of linked selective sweeps on demographic inference and model selection. *Genetics* 204, 1207–1223

[86] Stephan, W. (2016) Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Mol. Ecol.* 25, 79–88

[87] Good, B.H. *et al.* (2014) Genetic diversity in the interference selection limit. *PLoS Genet.* 10, e1004222

[88] Phung, T.N. *et al.* (2016) Determining the effect of natural selection on linked neutral divergence across species. *PLoS Genet.* 12, e1006199

[89] Roux, C. *et al.* (2016) Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biol.* 14, e2000234

[90] Losos, J.B. (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol. Lett.* 11, 995–1007

[91] Ricklefs, R.E. (2007) History and diversity: Explorations at the intersection of ecology and evolution. *Am. Nat.* 170, S56–S70

[92] Emerson, B.C. and Gillespie, R.G. (2008) Phylogenetic analysis of community assembly and structure over space and time. *Trends Ecol. Evol.*

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

[94] Vellend, M. *et al.* (2014) Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. *Mol. Ecol.* 23, 2890–2901

[95] Arbogast, B.S. and Kenagy, G.J. (2001) Comparative phylogeography as an integrative approach to historical biogeography. *J. Biogeogr.* 28, 819–825

[96] Smith, B.T. *et al.* (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

[97] Hickerson, M.J. and Cunningham, C.W. (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

[98] Venkataraman, A. *et al.* (2015) Application of a neutral community model to assess structuring of the human lung microbiome. *MBio* 6

[99] Hubbell, S.P. (2005) The neutral theory of biodiversity and biogeography and stephen jay gould. *Paleobiology* 31, 122–132

[100] Etienne, R.S. and Olff, H. (2004) A novel genealogical approach to neutral biodiversity theory. *Ecol. Lett.* 7, 170–175

[101] Jabot, F. and 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

[102] Burbrink, F.T. *et al.* (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

[103] Davies, T.J. *et al.* (2011) NEUTRAL BIODIVERSITY THEORY CAN EXPLAIN THE IMBALANCE OF PHYLOGENETIC TREES BUT NOT THE TEMPO OF THEIR DIVERSIFICATION. *Evolution* 65, 1841–1850

[104] Rosindell, J. *et al.* (2010) Protracted speciation revitalizes the neutral theory of biodiversity. *Ecol. Lett.* 13, 716–727

[105] Olszewski, T.D. and Erwin, D.H. (2004) Dynamic response of permian brachiopod communities to long-term environmental change. *Nature* 428, 738–741

[106] Wagner, P.J. *et al.* (2006) Abundance distributions imply elevated complexity of post-paleozoic marine ecosystems. *Science* 314, 1289–1292

[107] Kendall, D.G. (1948) On the generalized "Birth-and-Death" process. *Ann. Math. Stat.* 19, 1–15

[108] Kendall, D.G. (1948) On some modes of population growth leading to r. a. fisher's logarithmic series distribution. *Biometrika* 35, 6–15

[109] Engen, S. and Lande, R. (1996) Population dynamic models generating the lognormal species abundance distribution. *Math. Biosci.* 132, 169–183

[110] Engen, S. and Lande, R. (1996) Population dynamic models generating species abundance distributions of the gamma type. *J. Theor. Biol.* 178, 325–331

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

[112] McGill, B.J. *et al.* (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* 10, 995–1015

[113] Etienne, R.S. and Haegeman, B. (2012) A conceptual and statistical framework for adaptive radiations with a key role for diversity dependence. *Am. Nat.* 180, E75–E89

[114] Rabosky, D.L. (2013) Diversity-dependence, ecological speciation, and the role of competition in macroevolution. *Annu. Rev. Ecol. Evol. Syst.* 44

[115] Rabosky, D.L. and Lovette, I.J. (2008) Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? *Evolution* 62, 1866–1875

[116] Adler, P.B. *et al.* (2007) A niche for neutrality. *Ecol. Lett.* 10, 95–104

[117] HilleRisLambers, J. *et al.* (2012) Rethinking community assembly through the lens of coexistence theory. *Annual Review of*

[118] Levine, J.M. and HilleRisLambers, J. (2009) The importance of niches for the maintenance of species diversity. *Nature* 461, 254–257

[119] Mayfield, M.M. and Levine, J.M. (2010) Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecol. Lett.* 13, 1085–1093

[120] Godoy, O. *et al.* (2014) Phylogenetic relatedness and the determinants of competitive outcomes. *Ecol. Lett.* 17, 836–844

[121] Van Valen, L. (1973) A new evolutionary law. *Evolutionary theory* 1, 1–30

[122] Benton, M.J. (2009) The red queen and the court jester: species diversity and the role of biotic and abiotic factors through time. *science* 323, 728–732

[123] Kirchner, J.W. and Weil, A. (1998) No fractals in fossil extinction statistics. *Nature* 395, 337–338

[124] Krehenwinkel, H. *et al.* (2016) A cost-efficient and simple protocol to enrich prey dna from extractions of predatory arthropods for large-scale gut content analysis by illumina sequencing. *Methods in Ecology and Evolution*

[125] Elbrecht, V. and Leese, F. (2015) Can dna-based ecosystem assessments quantify species abundance? testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PloS one* 10, e0130324

[126] Coissac, E. *et al.* (2016) From barcodes to genomes: extending the concept of dna barcoding. *Molecular ecology*

[127] Nielsen, H.B. *et al.* (2014) Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nature biotechnology* 32, 822–828

[128] Angly, F.E. *et al.* (2014) Copyrighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome* 2, 11

[129] Saitoh, S. *et al.* (2016) A quantitative protocol for dna metabarcoding of springtails (collembola) 1. *Genome* 59, 705–723

[130] Overcast, I.A. and Hickerson, M.J. (in prep.) Integrating community assembly models and comparative population genetics.

[131] Rosindell, J. *et al.* (2015) Unifying ecology and macroevolution with individual-based theory. *Ecol. Lett.* 18, 472–482

[132] Kelleher, J. *et al.* (2016) Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput Biol* 12, e1004842

Boxes

Box 1: Wetlab techniques

Next generation sequencing technology has ushered in a revolution in evolutionary biology and ecology. The large scale recovery from bulk samples (e.g. passive arthropod traps) of species richness, food web structure, cryptic species promise unprecedented new insights into ecosystem function and assembly [3, 8, 9, 124]. Two approaches, differing in cost and effectiveness, have emerged.

Metabarcoding describes the targeted PCR amplification and next generation sequencing of short DNA barcode markers (typically 300-500 bp) from community samples [1, 4]. The resulting amplicon sequences can be clustered into OTUs or grafted onto more well supported phylogenies. Even minute traces of taxa in environmental samples can be detected using metabarcoding [7]. Amplicon sequencing is cheap, requires a small workload and thus allows rapid inventories of species composition and species interactions in whole ecosystems [2, 8, 11]. However, the preferential amplification of some taxa during PCR can lead to highly skewed abundance estimates [1, 125] from metabarcoding libraries.

Metagenomic approaches , in contrast, avoid marker specific amplification bias by sequencing libraries constructed either from untreated genomic DNA [6, 10, 12], or after targeted enrichment of genomic regions [13]. While being more laborious, expensive and computationally demanding than metabarcoding, metagenomics thus offers improved accuracy in detecting species composition and abundance [5]. Moreover, the assembly of high coverage metagenomic datasets recovers large contiguous sequence stretches, even from rare members in a community, offering high phylogenetic resolution at the whole community level [126]. Due to large genome sizes and high genomic complexity, metazoan metagenomics is currently mostly limited to the assembly of fairly short high copy regions. Particularly mitochondrial

32

and chloroplast genomes as well as nuclear ribosomal clusters are popular targets [12, 126]. In contrast, microbial metagenomic studies now routinely assemble complete genomes and characterize gene content and metabolic pathways even from complex communities [127]. This allows unprecedented insights into functional genetic process underlying community assembly and evolutionary change of communities to environmental stress. Such whole genome based community analysis is not yet feasible for macroorganisms. However, considering the ever increasing throughput and read length of next generation sequencing technology, as well as growing number of whole genomes, it might well become a possibility in the near future, opening up unprecedented new research avenues for community ecology and evolution.

Box 2: Bioinformatic advances

While species richness can be routinely identified by sequencing bulk samples using high throughput methods, estimating species abundance remains challenging [125] and severely limits the application of high throughput sequencing methods to many community-level studies. We propose two complementary approaches to estimate species abundance from high throughput data. The first approach estimates abundance free from any models of community assembly, the second jointly estimates the parameters of a specific assembly model of interest along with the parameters of a coalescent-based population genetic model.

Model-free abundance estimation. We propose a pipeline (Fig. I) where raw reads are generated and assembled into a phylogeny using standard approaches, and potentially aided by additionally available sequence data in a super tree or super matrix approach. The numbers of sequences assigned to each terminal tip are then used in a Bayesian hierarchical model which seeks to estimate the true number of organisms representing each terminal tip, accounting for sequencing biases originating from, e.g. primer affinity and copy number differences between taxa. Information on phylogenetic relatedness can inform modeled correlations in biases between taxa [128, e.g. copy number is known to be phylogenetically

33

conserved at least in microbes]. This approach is particularly tailored to metabarcoding data. In a potentially powerful extension, and thanks to the proposed Bayesian framework, information from sequencing experiments that seek to calibrate metabarcoding studies [e.g., 124, 129] can be used to build meaningfully informative priors and improve model accuracy. Through a simulation study (described in the supplement) we show that true underlying abundances can be accurately estimated (Fig. II).

Joint inference of community assembly and population genetic models. Coupling individual-based, forward-time models of community assembly with backwards-time hierarchical multi-taxa coalescent models permits inference about the values of the parameters in both models. This framework is flexible enough to incorporate multiple refugia, colonization routes, ongoing migration and both neutral and deterministic processes of assembly on time scales of hundreds of thousands of years (Fig. III). A forthcoming implementation [gimmeSADπ; 130] simulates an individual-based forward time community dispersal model [131] linked with the msPrime coalescent simulator [132]. This has been accomplished by rescaling the time dependent local abundance distributions into time dependent effective population size distributions while allowing for heterogeneity in migration and colonization rates. This simulation model can be combined with random forest classifiers and hierarchical ABC to enable testing alternative assembly models, including models that have not yet reached their theoretical equilibria.

Glossary

ABC Approximate Bayesian Computation. A method of approximating the posterior sample of parameters in a complex model whose likelihood function cannot be analytically solved by simulating realizations of the model, computing summary statistics from those realizations, and probabilistically accepting or rejecting the parameter values leading to those summary statistics based on their agreement with the observed statistics computed

₇₆₄ from the real data.

₇₆₅ ahistorical   Patterns or theories which do not contain information about the historical
₇₆₆ processes that gave rise to them

₇₆₇ coalescent   A backwards in time model approach in population genetics in which alleles in
₇₆₈ the sample are traced to their ancestors under demographic models of interest.

₇₆₉ statistical equilibrium   In the context of biodiversity, a description of a steady state
₇₇₀ arrived at not by the force of one or a few deterministic mechanisms but by the stationary,
₇₇₁ statistical behavior of very large collections of mechanistic drivers acting on large assemblages
₇₇₂ of organisms.

₇₇₃ Tajima's D   A metric of non-stationary evolution computed as the difference between
₇₇₄ two distinct derivations of theoretical genetic diversity. If neutral molecular evolution in a
₇₇₅ constant population holds, both derivations should be equal, and otherwise if assumptions
₇₇₆ of constant demography and neutral selection are violated, will be unequal.

₇₇₇ Outstanding Questions

₇₇₈ 1. How can fossil data be best integrated with ahistorical ecological theory and diversity
₇₇₉   dynamics as informed by phylogenetic and population genetic inference?

₇₈₀ 2. How can functional genomics be used to better distinguish between purely demographic
₇₈₁   and niche-based drivers of non-equilibrium? Can understanding the functional content
₇₈₂   of genomes across taxa in a community help predict potential for non-equilibrium re-
₇₈₃   sponses to future perturbations? Functional genomics is still very much in development
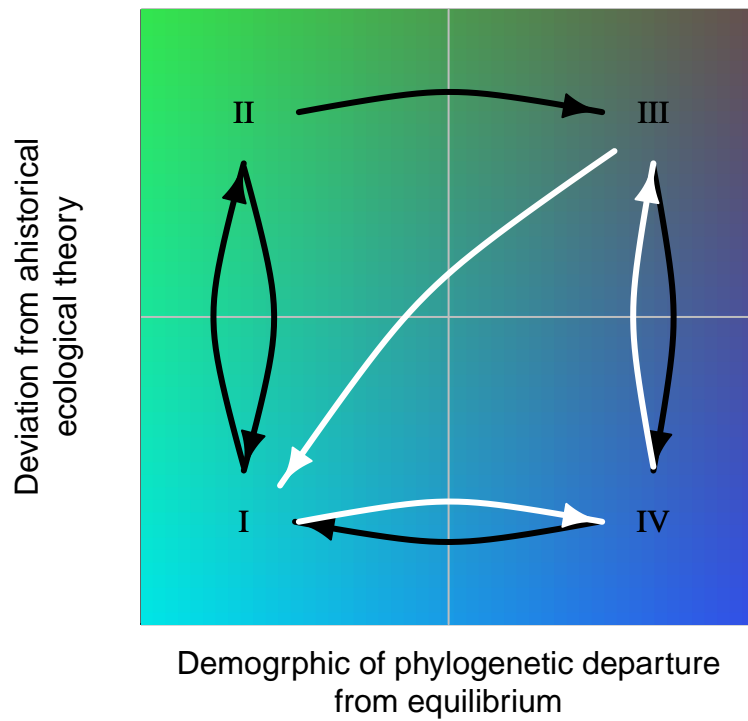₇₈₄   (see Box 1), but future prospects are exciting.

Figures

Figure 1: Hypothesized cycles between different states of equilibrium and non-equilibrium in ecological metrics (y-axis) and evolutionary metrics (x-axis). Panels I–IV are discussed in the next. Colors correspond to deviation from ahistorical ecological theory and evolutionary equilibrium. Black cycle corresponds to non-equilibrium initiated by ecological disturbance (with potential to continue to evolutionary non-equilibrium or relaxation to equilibrium). White cycle is initiated by evolutionary innovation.
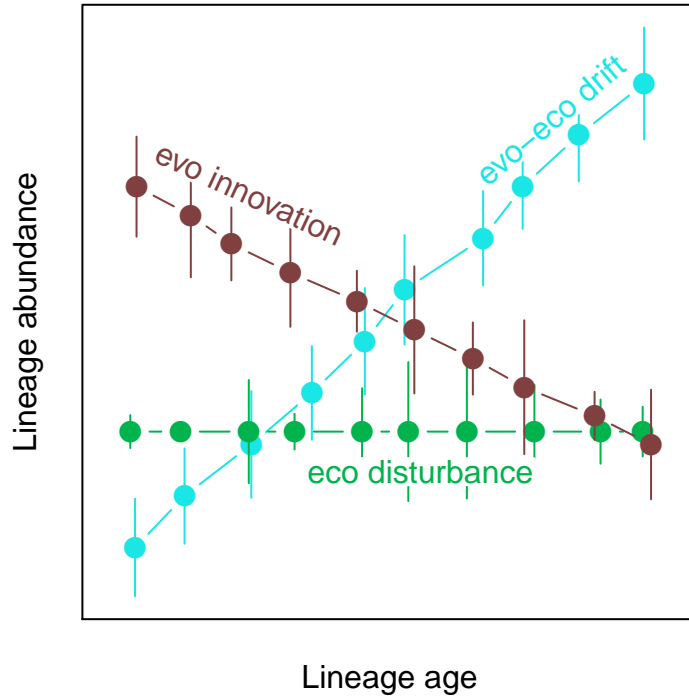
Figure 2: Hypothesized relationships between lineage age and abundance under different evo-ecological scenarios. Colors correspond to panels in Figure 1: teal is evo-ecological equilibrium; green is rapid transition to ecological non-equilibrium following short timescale disturbance; dark brown is non-equilibrium in both ecological and evolutionary metrics.
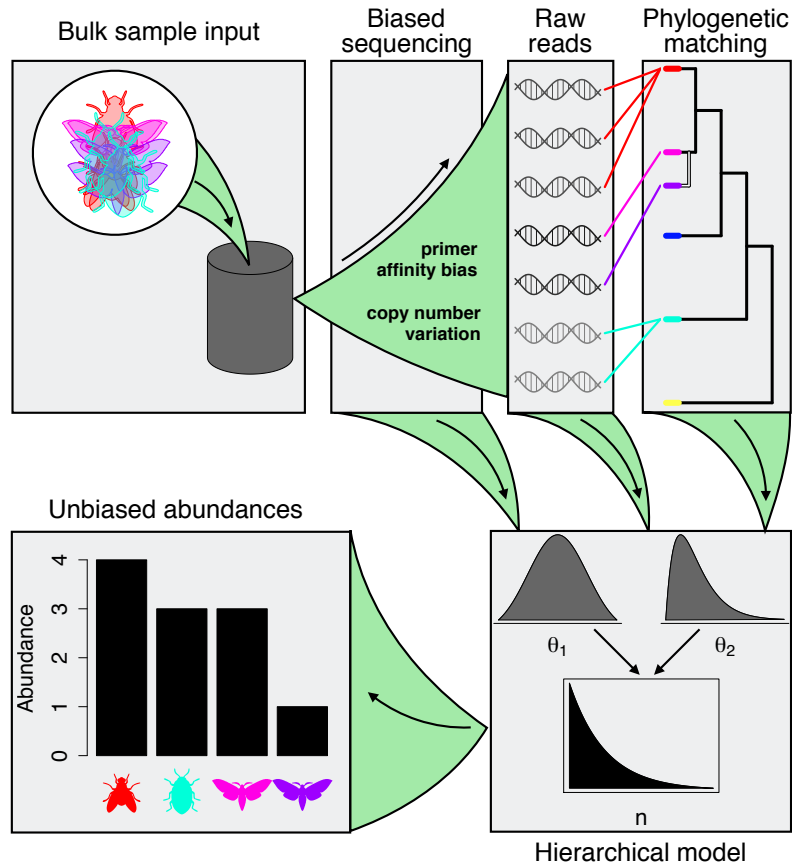
Box 2 figures

Figure I: Pipeline to estimate true abundances from metabarcoding data. The pipeline follows sequence generation, matching sequences to a phylogeny (generated from the sequences themselves, or better yet from higher coverage data) and finally Bayesian hierarchical modeling leading to abundance estimates.
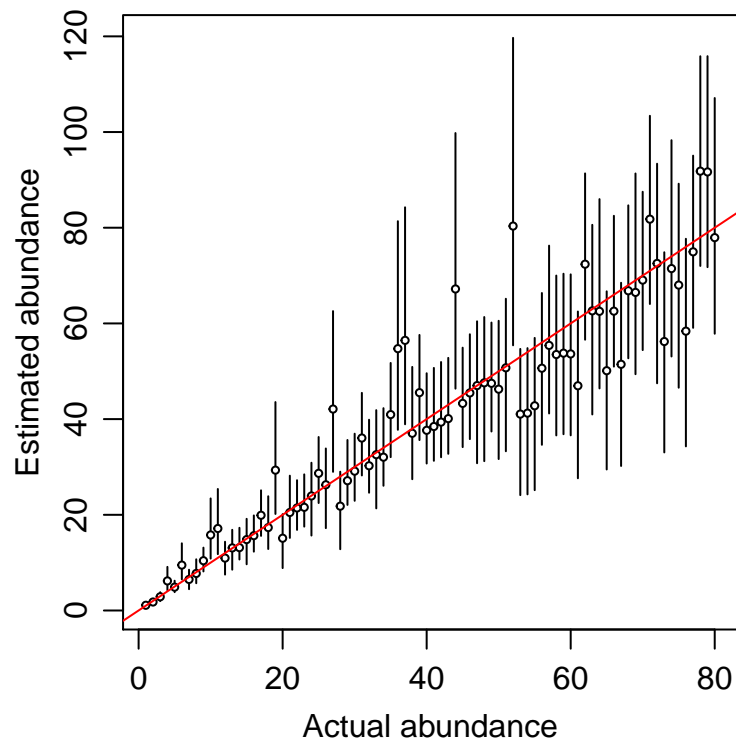
Figure II: Demonstration of agreement between actual and estimated abundances. Actual (simulated) abundances are on the x-axis, which the y-axis shows estimated abundances (error bars are 95% maximum credible intervals). The simulation study is described in the supplement.
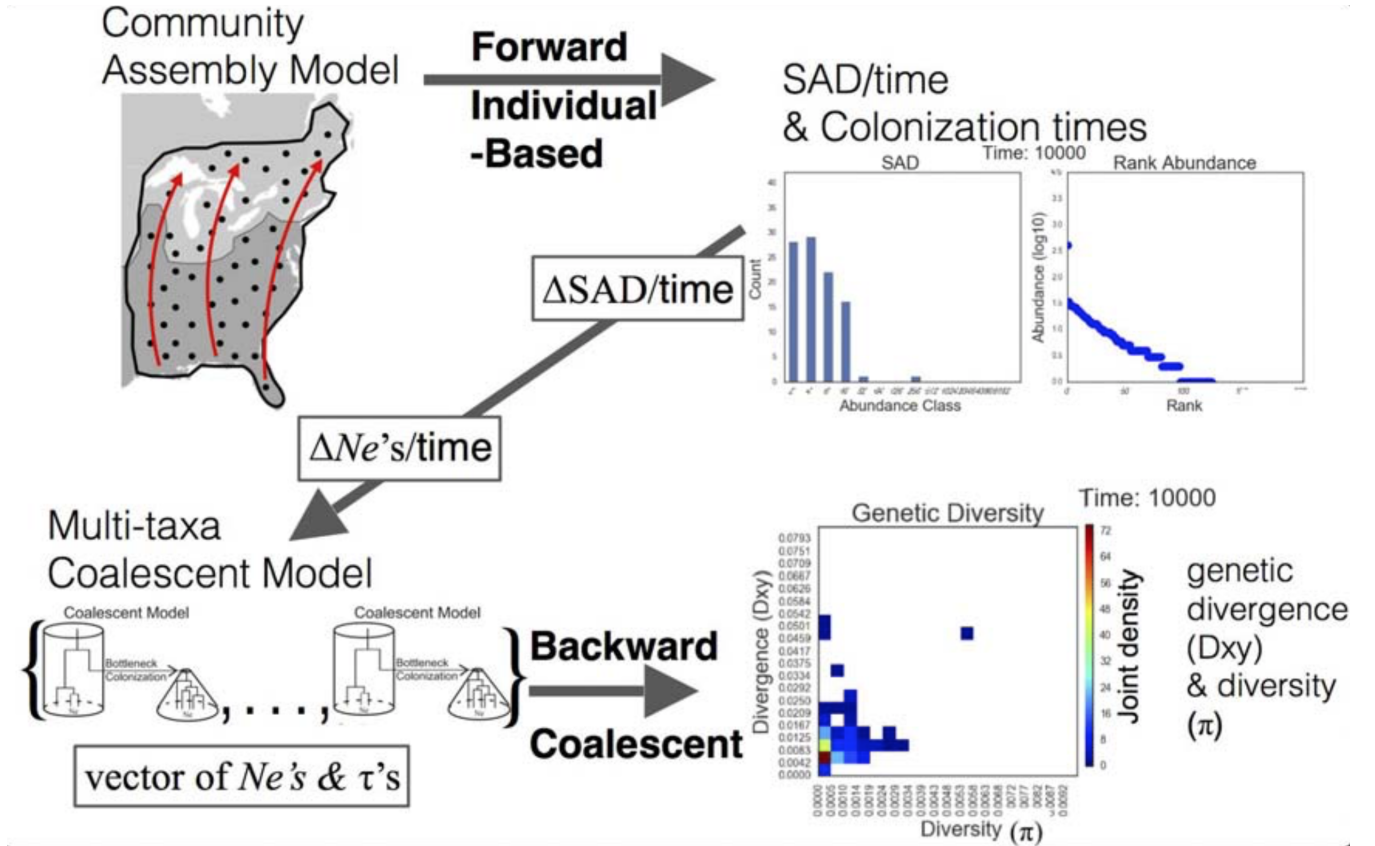
Figure III: The gimmeSADπ pipeline. The forward time models involves multi-regional expansion generating local abundance distributions over time with heterogeneity in colonization times. These temporally dynamic local abundances are re-scaled into local n_e distributions over time to generate multi-species genetic data the the coalescent, which is summarized here with a time-dependent joint spectrum of genetic diversity statistics.