

¹ Detection of eco-evolutionary dynamics in metacommunities using
² Joint Species Distribution Models

³ Jelena H. Pantel, Ruben Hermann

⁴ Ecological Modelling, Faculty of Biology, University of Duisburg-Essen, Universitätsstraße
⁵ 5, 45141 Essen, Germany

⁶ **Abstract**

⁷ Biodiversity at the metacommunity scale is typically influenced by a number of environmental, spatial, biotic,
⁸ and stochastic factors. At the same time, populations of individual species are evolving in ways that can be
⁹ impacted by the same factors, as sites present different selection pressures and site connectivity can impact
¹⁰ gene flow and genetic drift. Identifying the relative impacts of environmental, spatial, biotic, and other drivers
¹¹ of community composition across spatial and temporal scales has been greatly facilitated by joint species
¹² distribution models, but these models have not yet considered the role that evolution can play in driving
¹³ changes in composition. I used Hierarchical Models of Species Communities (HMSC) to analyze data from a
¹⁴ simulation model of an evolving metacommunity, in order to establish whether these models can sufficiently
¹⁵ quantify the contribution of phenotypic evolution for metacommunity composition. The models successfully
¹⁶ partitioned variance contributed by environmental, spatial, and phenotypic drivers, and also estimated site-
¹⁷ and year-specific covariance. I also included estimates for eco-evolutionary interactions, and the model
¹⁸ successfully estimated their role as well. The study of eco-evolutionary dynamics may require data that
¹⁹ reflects numerous complex, interacting processes and its necessary to have flexible, generalized statistical
²⁰ models to analyze this data. HMSC models present one promising path for analysis of eco-evolutionary
²¹ dynamics in multi-species communities.

22 Introduction

23 Analysis of biodiversity at large spatial scales can be complex, because numerous abiotic factors, spatial
24 structure and connectivity, interactions between species, and also stochastic processes such as drift and
25 priority effects operate and determine observed patterns. Statistical models in ecology have made tremendous
26 progress in reflecting these numerous ecological processes at multiple scales, and there are now a diversity
27 of joint species distribution models (Pollock et al. 2014 ([Pollock et al. 2014](#)); Pichler & Hartig 2021 ([Pichler
and Hartig 2021](#))) or multi-species occupancy models (REF; Devarajan et al. 2020). These models can
28 incorporate critical components that structure variance in biological data such as observer, measurement,
29 and process error, error propagation, the existence of present but undetected species or life stages (REF)
30 and other forms of imperfect detection (REF). Many of these models increasingly have implementations in R
31 packages (e.g. spOccupancy, Doser et al. 2022), and there are intriguing extensions of these occupancy models
32 with consideration of the error and data types particularly associated with eDNA (e.g. EDNAOCCUPANCY,
33 Dorazio & Erickson 2018) and citizen science data (Altwegg & Nichols 2019).

34 In addition to ecological processes, biodiversity at different levels of ecological organization – populations,
35 communities, and ecosystems – are partially determined by the genotypic and phenotypic properties of
36 organisms as well. These ecological processes can impact selection on genes and phenotypes, potentially
37 resulting in dynamics feedback loops between ecological and evolutionary processes (Lion 2018; Barbour et
38 al. 2022). Although an increasing number of theoretical (REF) and empirical studies have evaluated the
39 impacts of evolutionary dynamics for community assembly and metacommunity dynamics (REF), analytical
40 methods to consider this additional complexity lag behind the scope of data collected in eco-evolutionary
41 studies. A series of methods papers (Hairston et al. 2005; Ellner et al. 2011; Govaert et al. 2016) has
42 made progress in utilizing an ANOVA framework to partition variance (in phenotypes or in other ecological
43 response variables) across contributing ecological and evolutionary components. These approaches have
44 been applied to categorical ecological and evolutionary treatments with two levels. For example, the study
45 of terHorst et al. (2014) evaluated the impacts of a categorical evolutionary treatment, where host plants were
46 experimentally adapted to a dry or wet soil environment, and a categorical environmental treatment, where
47

48 experiments were conducted in a dry or wet soil environment, for bacterial and fungal community richness
49 and diversity. Hiltunen and Becks (2014) compared the relative importance of the (ecological) change in prey
50 density and the (evolutionary) change in prey defense traits for the change in predator density from one time
51 point to the next. However, these methods are ultimately limited by the same requirements that can limit
52 application of ANOVA - data should follow a normal distribution, the processes that structure the target
53 response variable must be linear and additive, and complicated structuring mechanisms must be reduced into
54 one or a few target categories labeled as ‘ecology’ or ‘evolution’. Flexible models, which can vary depending
55 on the mechanisms and structure of the diverse kinds of data that most studies of eco-evolutionary dynamics
56 will collect, are needed for rigorous inference about eco-evolutionary processes. These models should consider
57 the mechanistic basis of data structure, with realistic models of variance and uncertainty, and ideally can be
58 used to make predictions about future states of the system.

59 Two existing statistical models are well suited for analysis of eco-evolutionary data at the microevolutionary
60 and metacommunity scale. Lasky et al. (2020) developed an integrated reaction norm model, linking genetic,
61 phenotypic, and demographic processes, and Benito Garzón et al. (2019) presented a species distribution
62 model with local adaptation and phenotypic plasticity (Δ SDMs). The model of Lasky et al. is an excellent
63 candidate for spatially complex population dynamics, and awaits an extension to the multispecies level. The
64 Δ SDM models described by Benito Garzón et al. truly disentangle the role that plastic and evolutionary
65 trait divergence can play in species distributions, but they also require common garden experimental data
66 across sites to estimate these effects. Data typically collected in a metacommunity survey (of eukaryotes,
67 which thus focuses on the species level) may only be time series of population size and trait values, with
68 surveys of environmental properties and spatial connectivity as potential predictors.

69 One particular metacommunity statistical model, the Hierarchical Modelling of Species Communities
70 (HMSC) framework (Ovaskainen et al. 2017; Tikhonov et al. 2020), was built explicitly to consider the
71 multitude of processes that can structure species occupancy and abundance data across time and space and
72 is thus well suited to consider analysis of metacommunity eco-evolutionary data (Leibold et al. 2022). This
73 model of environmental filtering via variation and covariation in how species respond to their environment
74 considers the impacts of traits and phylogenetic relationships, and uses latent variables to account for

75 the numerous unobserved environmental and spatial features that are difficult to exhaustively measure
 76 in community data. The associated R package has immense flexibility and accompanying data analysis
 77 examples and code, and has been applied to study environmental responses of diverse assemblages of
 78 organisms (REF). In this study, I use a simulation of an evolving metacommunity model to generate realistic
 79 time series of population sizes and mean trait values for multiple species in a spatially and environmentally
 80 complex landscape. I then use HMSC to analyze the resulting data, to determine whether HMSC can
 81 successfully estimate the impacts of trait evolution for community composition, and to generate relative
 82 contributions of fixed environmental, spatial, and trait evolution drivers as well as random effects of
 83 spatial and temporal covariance among species. Finally, in order to determine whether HMSC is useful
 84 for hypothesis testing about the drivers of eco-evolutionary dynamics in communities in a landscape, I
 85 also conducted a few numerical experiments with the evolving metacommunity simulation - I varied the
 86 degree of environmental fluctuations in the landscape, and I varied the strength of species interactions, to
 87 determine whether these potential drivers of increased selection pressure alter the importance of evolution
 88 for metacommunity composition. For these experiments, I hypothesized that XX and YY. I found that
 89 XXXX.

90 Methods

91 Simulation model

92 Population growth for species in the metacommunity simulation follows a Leslie-Gower model (a discrete-
 93 time version of a Lotka-Volterra model; Beverton & Holt 1957; Leslie & Gower 1958). I consider the
 94 impact of trait evolution for growth using a discrete time quantitative genetic model of evolutionary rescue
 95 (Gomulkiewicz & Holt 1995). The model for population size is as follows:

$$N_{i,t+1} = \frac{\hat{W} e^{\frac{-[(w + (1-h^2)P)(E-x_t)]^2}{2(P+w)}} N_{i,t}}{1 + \alpha_{ii} N_{i,t} + \sum_{j \neq i}^S \alpha_{ij} N_{j,t}}$$

96 where $N_{i,t}$ is the population size of species i at time t , \hat{W} is calculated as $\hat{W} = W_{max} \sqrt{\left(\frac{w}{P+w}\right)}$, W_{max} is the

97 species' maximum per-capita growth rate, w is the width of the Gaussian fitness function (which determines
 98 the strength of selection), P is the width of the distribution of the phenotype x , h^2 is the heritability of
 99 the trait x , E is the local environmental optimum trait value, $x_{i,t}$ is the trait value of species i at time t ,
 100 α_{ii} is the intraspecific competition coefficient (the per capita impact of species i on itself) and α_{ij} is the
 101 interspecific competition coefficient. Populations have a critical density N_c , below which the population is
 102 subject to extinction due to demographic stochasticity at a probability of p (Gomulkiewicz & Holt 1995).

103 This model was used in Pantel & Becks (2023) to evaluate the consequences of adaptive evolution for
 104 coexistence in a three-species system. To expand this model to a metacommunity, I consider the evolution of
 105 multiple species ($S = 15$) in a landscape of patches ($k = 50$). The patches have values of an environmental
 106 property E (drawn from a uniform distribution ranging from 0 to 1) that determines the local optimum
 107 phenotype E , i.e. where species experience their absolute fitness W_{max} , and patches also have spatial locations
 108 X and Y (both drawn from $U(0, 1)$). The patches thus have a connectivity matrix \mathbf{D} (here given by their
 109 Euclidean distance), as well as a connectivity matrix \mathbf{C} that is a Gaussian function of \mathbf{D} and a species
 110 dispersal rate d_i .

111 To simulate phenotypic evolution as well as species growth and competition in the landscape, I used parameter
 112 values of $W_{max} = 2$, $P = 1$, $w = 10$, $N_c = 100$, and $p = 0.001$. The simulated metacommunity was
 113 initialized in a way that mimics community assembly in small vernal pools that periodically flood (and
 114 thus are reset seasonally). In the simulated pool system, local communities are drawn at random from the
 115 regional species pool, and thus their initial degree of maladaptation in the pool is initially random as well.

116 The initial species richness of each site s_{0k} was drawn from a random Poisson distribution, $s_{0k} \sim Pois(0.75)$,
 117 the initial population size for all species $N_{0i} \sim Pois(10)$, and the initial degree of maladaptation $\beta_{i0} \sim$
 118 $Gamma(0.75, 1)$. From this β_{i0} , the initial distance to the local patch's optimum phenotype was calculated
 119 as $d_{i0} = \sqrt{B_{i0}(w + P)}$ and the initial trait value for each species x_{i0} was calculated by subtracting d_{i0} from
 120 the lowest site environmental value E_k .

121 The species interaction matrix was fixed throughout the simulation, with $\alpha_{ii} = 0.00125$ and
 122 $\alpha_{ij} \sim U(0.0015, 1)$. After draws for all random parameter values, the same resulting initial commu-
 123 nity was used for all simulations, to allow direct comparison across dispersal and heritability values (Figure

¹²⁴ @ref(fig:figS1)). Simulations were run across a range of dispersal values (identical for all species; $d = 0$,
¹²⁵ $10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1) and heritability
¹²⁶ values (identical for all species: $h^2 = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$). The simulations were run
¹²⁷ in MATLAB (version R2022a) for 1000 time steps and produced records of N_{ikt} and x_{ikt} .

¹²⁸ Statistical model

¹²⁹ To determine the relative influence of environmental and spatial properties, population dynamics, and evolu-
¹³⁰ tionary dynamics for community structure, I applied hierarchical modelling of species communities (HMSC,
¹³¹ Ovaskainen et al. 2017; Tikhonov et al. 2020) to model the abundance at each site and year for each species
¹³² N_{ikt} . The abundance values were modeled as $N_{ikt} \sim N(L_{ikt}, \sigma_2)$, where the linear predictors \mathbf{L} are the sum
¹³³ of fixed and random effects $\mathbf{L}^F + \mathbf{L}^R$ (Tikhonov et al. 2020). The fixed effects were (1) E_k , the value of the
¹³⁴ site's environmental parameter, (2) the value of each species' population size the time step before $N_{ik,t-1}$,
¹³⁵ (3) the absolute value of the change in each species' trait value from the previous time step to the next
¹³⁶ $|\Delta x_{ik,t-1 \rightarrow t}|$, (4) and an interaction term for population size by change in trait value $N_{ik,t-1} \times |\Delta x_{ik,t-1 \rightarrow t}|$,
¹³⁷ which led to a total of 46 predictors. The random effects were modeled at the site-level ($\epsilon_{ik} \sim N(0, \Omega^S)$,
¹³⁸ where Ω^S is the site species-to-species covariance matrix) and at the day level ($\epsilon_{it} \sim N(0, \Omega^T)$, where Ω^T
¹³⁹ is the time species-to-species covariance matrix) using a latent variable approach. In this approach, species-
¹⁴⁰ to-species covariance matrices are computed as $\Omega = \lambda \top \lambda$, where λ are the loadings of each species with
¹⁴¹ latent variables, and latent variables are modeled as $\epsilon_{ik} = \sum_B \eta_{kB} \lambda_{Bi}$ and $\epsilon_{it} = \sum_B \eta_{tB} \lambda_{Bi}$ for B latent
¹⁴² variables considered (Ovaskainen et al. 2016). The spatial latent variables represent a spatial model, where
¹⁴³ species respond to some latent spatial predictor associated with the xy coordinates of the site (Tikhonov et
¹⁴⁴ al. 2020 ([Tikhonov et al. 2020](#))). Analysis and was conducted in R (version 4.1.1) using packages 'Hmsc',
¹⁴⁵ 'R.matlab', and XX (Bengtsson 2022; Tikhonov et al. 2022; XX).

¹⁴⁶ Results

¹⁴⁷ The simulation model followed population size and trait value dynamics for species in the landscape across
¹⁴⁸ 1000 time steps, and results depended on the dispersal level and heritability values (Figure @ref(fig:fig1)).

149 For dispersal, results matched predictions of existing metacommunity models demonstrating that diversity
150 is higher for intermediate rates of dispersal (as species can increasingly reach more suitable patches at
151 intermediate dispersal, but then experience mass effects at higher dispersal levels; Mouquet & Loreau 2003
152 ([Mouquet and Loreau 2003](#)); Figure @ref(fig:fig2)), and that among-site similarity (measured as the average
153 among-site Hellinger distance; Legendre & Gallagher 2001 ([Legendre and Gallagher 2001](#))) decreases with
154 increasing dispersal level (Figure @ref(fig:fig3)). The inclusion of trait heritability did have a strong effect
155 on community diversity, although there was no indication that increasing heritability beyond $h^2 = 0.1$ had
156 any impact (Figure @ref(fig:figS2)).

157 Discussion

158 I have analysed data collected by Herman Bumpus ([Bumpus 1898](#)) on the relationship between sparrow
159 (*Passer domesticus*) total length and survival following an unusually severe storm. I found that sparrows that
160 died in the storm were longer than sparrows that survived, which suggests that higher sparrow body length
161 decreased survival. Of course, it is not possible to definitively conclude a causal relationship between any
162 aspect of body size and sparrow survival, and even the available data collected by Bumpus would permit a
163 more thoughtful analysis than that conducted in this study (see [Appendix Table 1](#)).

164 Overall, this document demonstrates how high quality, professional looking documents can be written using
165 Rmarkdown. The [underlying code](#) for this manuscript is publicly available, along with [accompanying notes](#)
166 to understand how it was written. By using Rmarkdown to write manuscripts, authors can more easily use
167 version control (e.g., git) throughout the writing process. The ability to easily integrate citations through
168 BibTeX, LaTeX tools, and dynamic R code can also make writing much more efficient and more enjoyable.
169 Further, obtaining the benefits of using Rmarkdown does not need to come with the cost of isolating
170 colleagues who prefer to work with Word or LaTeX because Rmarkdown can easily be converted to these
171 formats (in the case of Word, with the push of a button). By learning all of the tools used in this manuscript,
172 readers should have all of the necessary knowledge to get started writing and collaborating in Rmarkdown.

173 References

- 174 Bumpus, H. C. 1898. Eleventh lecture. The elimination of the unfit as illustrated by the introduced sparrow,
175 *Passer domesticus*. (A fourth contribution to the study of variation.). Biological Lectures: Woods Hole
176 Marine Biological Laboratory 209–225.
- 177 Legendre, P., and E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species
178 data. *Oecologia* 129:271–280. Springer.
- 179 Mouquet, N., and M. Loreau. 2003. [Community patterns in source-sink metacommunities](#). *The American*
180 *Naturalist* 162:544–557.
- 181 Pichler, M., and F. Hartig. 2021. [A new joint species distribution model for faster and more accurate](#)
182 [inference of species associations from big community data](#). *Methods in Ecology and Evolution* 12:2159–
183 2173.
- 184 Pollock, L. J., R. Tingley, W. K. Morris, N. Golding, R. B. O'Hara, K. M. Parris, P. A. Vesk, and M. A.
185 McCarthy. 2014. [Understanding co-occurrence by modelling species simultaneously with a joint species](#)
186 [distribution model \(JSDM\)](#). *Methods in Ecology and Evolution* 5:397–406.
- 187 Tikhonov, G., Ø. H. Opedal, N. Abrego, A. Lehikoinen, M. M. J. de Jonge, J. Oksanen, and O. Ovaskainen.
188 2020. [Joint species distribution modelling with the r-package hmsc](#). *Methods in Ecology and Evolution*
189 11:442–447.

190 Figures & Tables

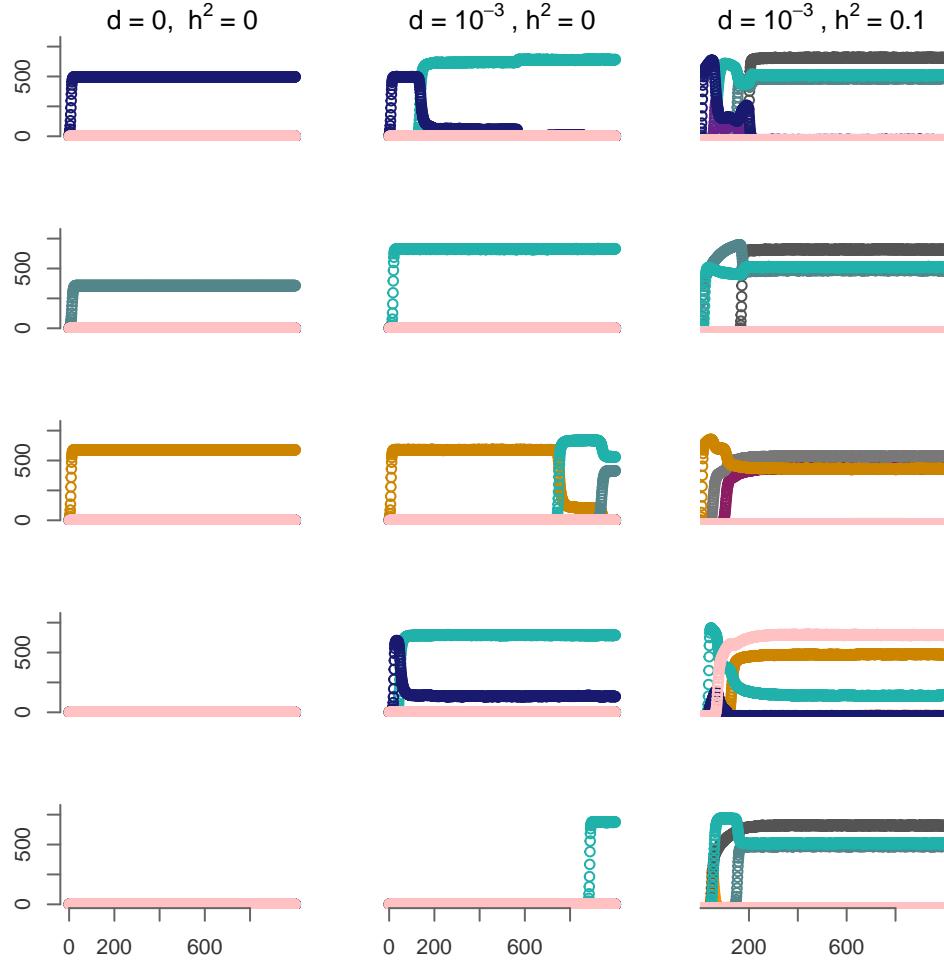


Figure 1: Time series of population size (y -axis) over time (x -axis) for all species in 5 of 50 total patches, for varying levels of dispersal d and heritability h^2 . All patches have a distinct value for E , the environmental parameter (that determines the local population mean fitness; here $E = 0.60198194, 0.45054160, 0.08382138, 0.07596669$, and 0.23991615 for patches from top to bottom row). Each of the $s = 15$ species is plotted with a unique color. The left column has no dispersal ($d=0$) and no evolution ($h^2=0$). The middle column includes intermediate dispersal ($d = 10^{-3}$) and no evolution, while the third column has intermediate dispersal and evolution ($h^2=0.1$).

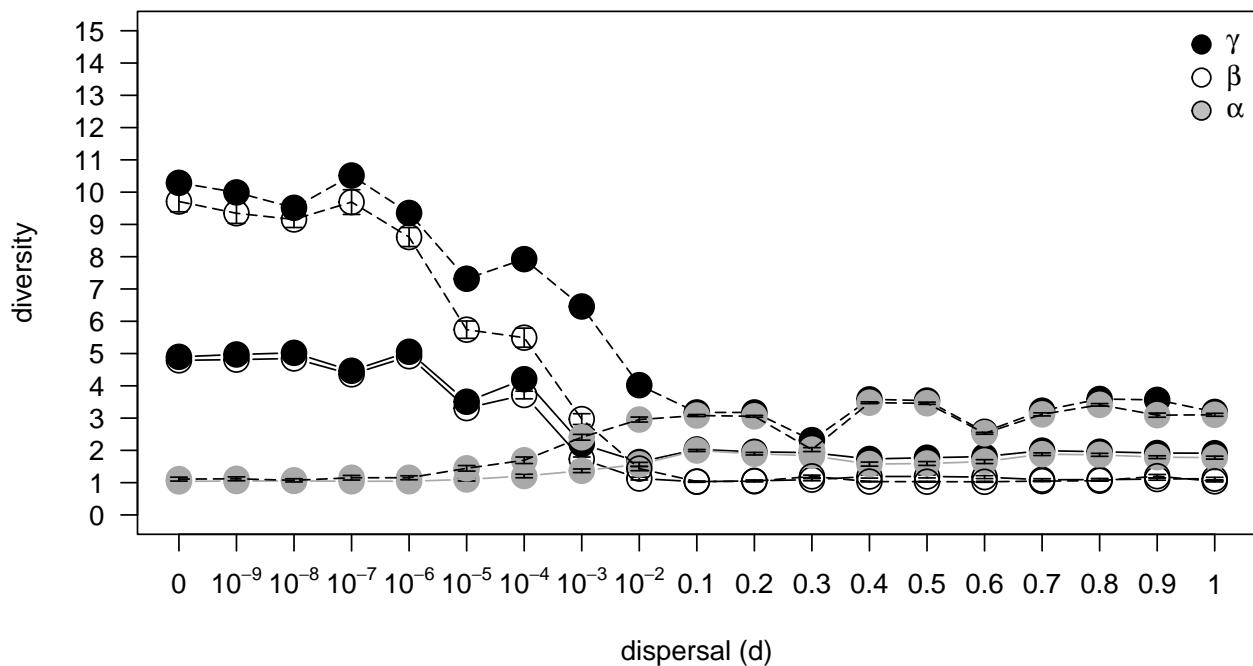


Figure 2: Inverse Simpson's diversity (2D) at the local (average α value across all 50 sites, with error bars at ± 1 standard error) and regional (γ) level, and turnover between sites (average β value across all 50 sites, with error bars at ± 1 SE) for $h^2=0$ (solid lines) and $h^2=0.1$ (dashed lines), indicating that local adaptation rescues numerous species that would otherwise become extinct.

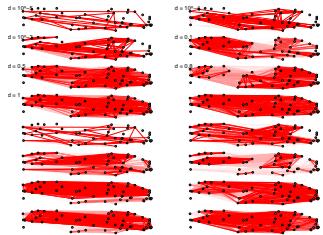


Figure 3: (A) Plot with among-site similarity plotted across dispersal levels for $h^2=0$ and $h^2=0.1$, for dispersal levels XXX. (B) Plot with average among-site similarity (error bars at ± 1 standard error are given but very small) across $h2$ levels.

¹⁹¹ **Appendix 1**

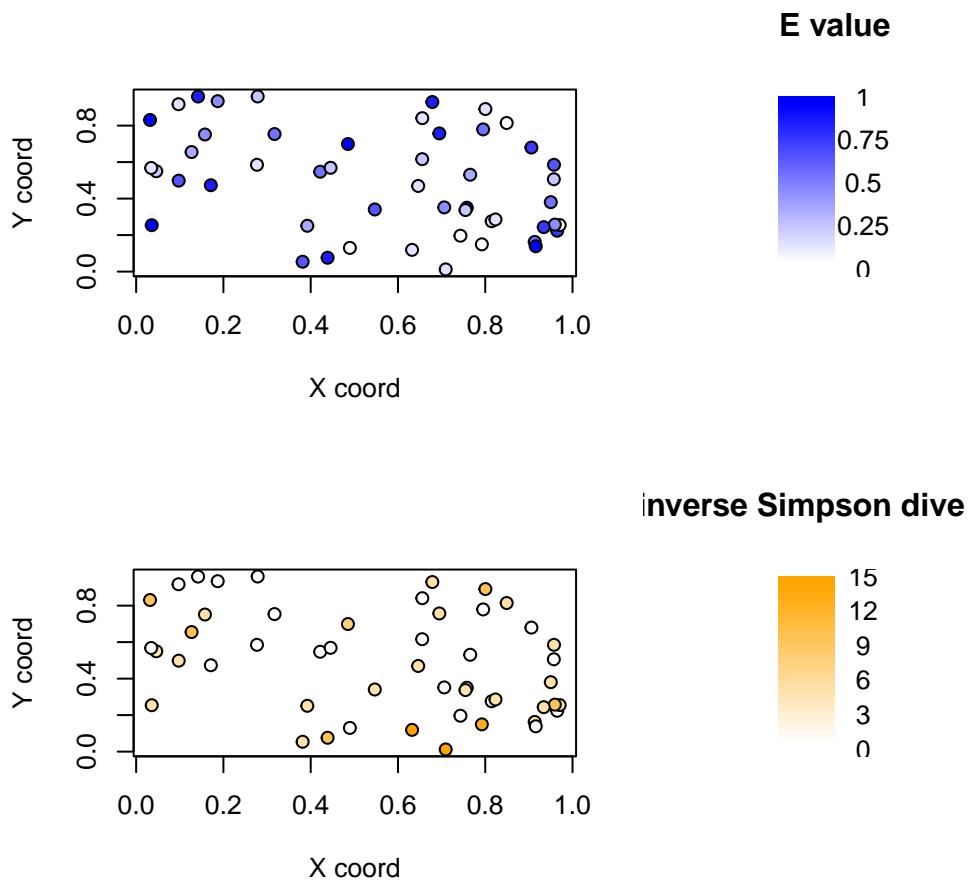


Figure 4: Figure S1. Spatial maps showing position of all 50 sites in the simulation, their value for the single environmental variable E, and their value for initial Simpson's diversity at the beginning of the simulation.

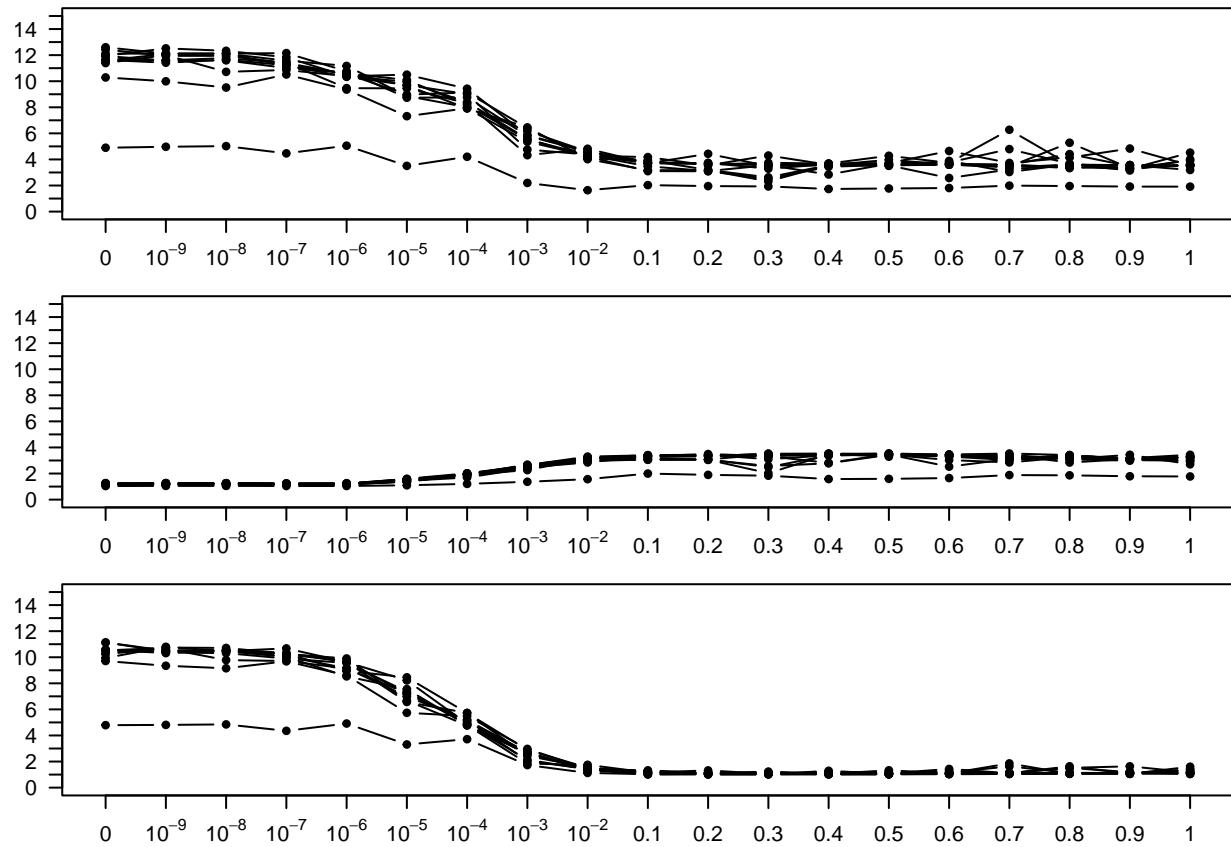


Figure 5: Plots with (A) gamma, (B) alpha, and (C) beta diversity across levels of dispersal (x-axis), for values of heritability ranging from 0-1.