**Appendix S2. Description of simulations run using *scape* – a phylogenetically informed community assembly simulation platform in the R package *pez***

This appendix describes the *R*-based simulation platform (Pearse *et al.*, 2015) (known hereafter as ‘*scape*’) used to assemble communities within a spatially explicit landscape, *via* phylogenetic and biogeographical assembly mechanisms. In the *scape* simulation, ‘taxa’ with known phylogenetic relationships are assembled into local ‘communities’ which, when combined, define a ‘landscape’.

*Generating phylogenetic trees*

Phylogenetic trees defining the evolutionary relationship between taxa were generated using the *sim.bdtree* function in the R package *geiger* (Harmon *et al.*, 2008). This function simulates trees under a uniform birth–death process, stopping when the desired number of tips is obtained. An existing phylogenetic tree could also be used. Trees must be made ultrametric before proceeding. We generated 100 random trees, each with 64 tips, observing indices of tree structure (δ statistic and *Ic* values) to ensure an even distribution of tree shapes was achieved.

*Simulating landscapes*

Landscape size is specified as the square of the number of communities forming a single side. This square landscape is also described spatially with paired *X* and *Y* coordinates, such that each community in the landscape has a unique pair of *X* and *Y* coordinates. Environmental conditions are defined along both the *X* and *Y* dimension, currently, the same gradient of values occurs along both dimensions such that the environmental conditions in the environment are symmetrical along the diagonal. Taxa then have optima or ‘niches’ which relate to these environmental conditions. We generated landscapes in a 16 by 16 grid, thus comprising 256 total communities.

Once the number of taxa, their evolutionary relationships (i.e. the phylogeny), and the landscape size and conditions are decided, the desired assembly parameters must be chosen. In general, choices reflect whether there is a phylogenetic signal for species’ environmental optima, how strong that signal is, and whether it has the effect of repulsion of related species (similar species are less likely to co-occur) or attraction of related species (similar species are more likely to co-occur) or neither. Where there is a phylogenetic signal for environmental optima, this is achieved by using the *corBlomberg* function in the R package *ape* (Paradis, Claude & Strimmer, 2004) to modify the phylogenetic tree’s variance–covariance matrix accordingly. In the case of repulsion, the resulting pattern of co-occurrences is similar to that from competition, while for attraction it is more similar to the expectation from environmental filtering or facilitation. Parameters also control whether there is a phylogenetic signal for range size, and its strength. Related species may, for example, have similar range sizes.

Landscape characteristics were also varied so as to alter species’ distributions. The average range width was set as a proportion of the total landscape size (here, 0.2). This represents the number of cells in the landscape a species is, on average, likely to be found in. Where range size lacks a phylogenetic signal, species’ ranges are drawn randomly from a normal distribution with a given mean size. Ranges may be spatially coherent, where species’ probabilities of presence have a hump-shaped distribution. Range sizes need not be spatially coherent, in which case sites in which species are present are more dispersed through space.

Scape works by updating continually the probability of presence of each species on the landscape based on the assembly parameters chosen. The probability of presence may be affected by species’ environmental optima, the presence of other species and their degree of relatedness, the average range size, the range size of similar species, and whether or not range sizes are identical. In addition, the amount of stochasticity in the application of these assembly rules can be adjusted by adjusting the amount of random variation in range size and species’ environmental niche values.

The resulting matrix shows the probability of presence of each species for each site (i.e. a site \* species matrix), which is used to produce a site-by-species presence/absence matrix (Fig. S1). A species abundance matrix, where probabilities are scaled by a provided maximum species abundance (*K*), was also produced. Here maximum abundance was set at 100 individuals per community. This matrix was used to calculate abundance-based metrics.

For each tree, we simulated eight types of communities meant to reflect a wide range of possible parameter combinations, from strong niche (and so, phylogenetically correlated) structure to random assembly (Table S1).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Parameter name** |  |  |  |  |  |  |
| **Landscape type** | **g.center** | **g.range** | **g.repulse** | **repulse** | **signal.center** | **signal.range** | **same.range** |
| 1 | 0.2 | 1 | 0.2 | TRUE | TRUE | FALSE | FALSE |
| 2 | 20 | 1 | 1 | FALSE | TRUE | FALSE | FALSE |
| 3 | 0.2 | 0.2 | 0.2 | TRUE | TRUE | TRUE | FALSE |
| 4 | 20 | 20 | 1 | FALSE | TRUE | TRUE | FALSE |
| 5 | 1 | 1 | 1 | FALSE | FALSE | FALSE | FALSE |
| 6 | 1 | 1 | 1 | FALSE | FALSE | FALSE | TRUE |
| 7 | 1 | 1 | 1 | FALSE | FALSE | TRUE | FALSE |
| 8 | 1 | 1 | 1 | FALSE | FALSE | TRUE | TRUE |
|  |  |  |  |  |  |  |  |

Table S1. Parameter values used for the eight types of landscapes simulated using *scape*.



**Fig. S1.** *scape* workflow showing the necessary input information and the path by which these are used to produce output from the assembly model. R code for the *scape* simulation is available in the R package *pez* (Pearse *et al*., 2015). *K* is the carrying capacity (maximum number of individuals) for each community.