

# Supporting Information

## **Tropical forests can maintain hyperdiversity because of enemies**

Taal Levi, Michael Barfield, Shane Barrantes, Christopher Sullivan, Robert D. Holt, John Terborgh

S1. Derivation of general exclusion zone equation

S2. Average number of species in a 9 cell exclusion zone assuming equal abundances and random placement

S3. Absorbing Markov chain with equations for a general exclusion zone

S4. Software to implement simulations

S5. Results using other diversity metrics

Figures S1-S3

Table S1

## S1. Derivation of general exclusion zone equation

Assume that the site of the newly dead tree is unavailable to conspecifics of the dead tree and trees in the  $E - 1$  surrounding sites (for example,  $E=9$  for a 3 by 3 square exclusion zone) due to Janzen-Connell effects. To increase in abundance, species  $i$  must not include a tree in the  $E$ -cell exclusion zone. If the absence of species  $i$  in a cell is independent of its absence from adjacent cells, this has probability  $(1 - N_i/J)^E = (1 - f_i)^E$ . Since J-C effects prevent species  $i$  from being in cells within an exclusion distance of each other, species  $i$  should be more uniformly distributed than this, so this is an upper bound on the actual probability. As species  $i$  approaches a spatially uniform arrangement, the probability of its absence in the exclusion zone approaches  $1 - Ef_i$ . (This can be shown as follows. Let  $p_k$  be the probability that there are  $k$  species- $i$  trees in a random  $E$ -cell exclusion zone. Then the expected value of the number of these trees is  $\sum_{k=0}^E k p_k$ . However, if species  $i$  is uniform in space and  $f_i < 1/E$ , then there can only be 0 or 1 tree in an exclusion zone. So the expected value of the number of species- $i$  trees is  $0(p_0) + 1(p_1) = p_1 = 1 - p_0$ . But since  $f_i$  is the average density, the expected value of the number of species- $i$  trees in  $E$  cells must be  $Ef_i$ . Therefore assuming spatial uniformity  $p_0 = 1 - Ef_i$ . Since uniformity in space is the most regular spatial distribution possible,  $1 - Ef_i$  is a lower bound. Assuming  $f_i$  is small, the upper  $[(1 - f_i)^E]$  and lower  $[1 - Ef_i]$  bounds are close to each other (the linear term is the result of dropping higher order terms from a binomial or Taylor expansion), and we use the latter, simpler, expression that is also consistent with a relatively uniform spatial distribution.

If species  $i$  is not in the exclusion zone, the probability that the dead tree's replacement on the first attempt is species  $i$  is  $P_1 = N_i/(J - E) \approx f_i$  (since  $J \gg E$ ). If species  $i$  is not selected on the first try (which has probability  $1 - f_i$ ), a prohibited species is selected with probability  $E/(S - 1)$ , since there are  $E$  cells in the exclusion zone and we are assuming that they are occupied by  $S - 1$  different equally abundant species. In this case, species  $i$  has another chance to colonize with probability  $P_2 \approx (1 - f_i)E/(S - 1)$ . The probably that species  $i$  is selected is again  $P_1$ , and if instead a prohibited species is selected randomly, species  $i$  gets a third chance with probability  $P_2$ , and so on. The probability of species  $i$  replacing the dead tree (given that it is not an excluded species) is the sum of the probability that  $i$  is chosen on the first or subsequent tries, which is a geometric series. The probability that species  $i$  increases in abundance is the product of the probability defined by the geometric series,  $\sum_{n=0}^{\infty} P_1 P_2^n$ , which equals  $\frac{P_1}{1 - P_2}$ , and the probability that species  $i$  is not in the exclusion zone,  $(1 - Ef_i)$ . Therefore,

$$\Pr(N_i + 1|N_i) \approx (1 - Ef_i) \sum_{n=0}^{\infty} P_1 P_2^n = \frac{(1 - Ef_i)P_1}{1 - P_2} \approx (1 - Ef_i) \left( \frac{f_i}{1 - \frac{[1 - f_i]E}{S - 1}} \right) = \frac{(1 - Ef_i)f_i(S - 1)}{S - 1 - E + Ef_i}.$$

Species  $i$  decreases in abundance if it dies, which has probability  $\Pr(N_i - 1|N_i) = N_i/J = f_i$ , since a dead tree cannot be replaced by another of the same species. The ratio between the probability of increase and of decrease is therefore

$$\Pr(N_i + 1|N_i) / \Pr(N_i - 1|N_i) \approx (1 - Ef_i)(S - 1) / (S - 1 - E + Ef_i).$$

We assumed that the non-focal species (all except species  $i$ ) are equally abundant to make the derivation simpler, and because we do not know the actual distribution of abundances. However, we can show the effect of heterogeneity in abundances. Assume that the focal species is assigned species number  $S$ , and the non-focal species have abundances  $N_j$  (and the total abundance of non-focal species  $i$  is the sum of  $N_j$  from  $j = 1$  to  $S - 1$ ). Now the probability that a tree of species  $j$  is selected to replace the dead tree (given that a tree that is species  $i$  is not) is

$N_j/(J - N_i)$  (species  $j$ 's relative abundance among non- $i$  trees). The probability that species  $j$  is in a particular cell in the exclusion zone (give that species  $i$  is not) is the same. The probability that it is in any cell in the exclusion zone is approximately  $EN_j/(J - N_i)$  (using the same argument giving the approximate probability of species  $i$  being absent from the exclusion zone). Therefore, the expression for  $P_2$  above including this heterogeneity would be

$$P_2 \approx (1 - f_i) \sum_{j=1}^{S-1} \frac{N_j}{J - N_i} \frac{EN_j}{J - N_i} = \frac{(1 - f_i)E}{(J - N_i)^2} \sum_{j=1}^{S-1} N_j^2 = \frac{(1 - f_i)E(S-1)\bar{N}_j^2(1 + cv_j^2)}{(J - N_i)^2} = \frac{(1 - f_i)E(1 + cv_j^2)}{S - 1},$$

where  $\bar{N}_j$  and  $cv_j$  are the mean and coefficient of variation of  $N_j$ , respectively. The fact that  $\bar{N}_j$  is  $(J - N_i)/(S - 1)$  is used in the last step. If species other than species  $i$  are equally abundant, then  $cv_j = 0$  and  $P_2 \approx (1 - f_i)E/(S - 1)$  as above. If there is variation, then this increases  $P_2$  from this value, because abundant species are more likely to be in the exclusion zone and more likely to be the selected replacement, meaning there are more rejections. Having  $S - 1$  non-focal species with coefficient of variation  $cv_j$  is equivalent to having  $(S - 1)/(1 + cv_j^2)$  non-focal species in determining the ratio of probability of increase to decrease of the focal species. Since this is a lower number of non-focal species, the rare species advantage is increased by heterogeneity and diversity maintenance is stronger.

## S2. Average number of species in a 9 cell exclusion zone assuming equal abundances and random placement

The average number of species in a 9 cell exclusion zone (one-crown exclusion distance) is  $E[m] = \sum_{m=1}^9 mp_m$ , where  $p_m$  is the probability that there are  $m$  different species in an exclusion zone. These probabilities can be found under the assumption that all species are equally likely to be found in each cell (independent of the species in other cells). In this case, all arrangements of species in the cells are equally likely, so the probability of  $m$  species is the number of arrangements with  $m$  species divided by the total number of arrangements. If there are  $S$  species, there are  $S$  possible species in each cell, and the total number of species arrangements in 9 cells is  $S^9$ .

If the number of species is large, the most common situation will be a different species in each of the 9 cells, so  $m = 9$ . The number of arrangements that have 9 different species is the product of the number of 9-species combinations that can be chosen from  $S$  species, which is  $\binom{S}{9}$ , and the number of ways to arrange those in 9 cells, which is  $9!$ . Therefore,  $p_9 = \binom{S}{9} \frac{9!}{S^9}$ .

The next most common number of species is 8, which results from having one species in 2 cells and different species in all the others. The number of ways to select one species to fill two cells from the  $S$  species is  $\binom{S}{1}$ , and the number of ways to select the 7 others from the remaining  $S - 1$ , is  $\binom{S-1}{7}$ , which gives the number of species combinations as  $\binom{S}{1} \binom{S-1}{7}$ . There are again  $9!$

permutations of 9 trees, but since 2 of the trees are of the same species, permutations that differ only by having the cells of these two trees reversed are the same in terms of species arrangement.

Therefore, the number of permutations is  $9!/2!$ . This gives a probability of

$p_8 = \binom{S}{1} \binom{S-1}{7} \frac{9!}{2!S^9}$ . There are two ways to get 7 species: one species can occupy 3 cells and the other cells each have a unique species, or two species can have 2 each and the others each

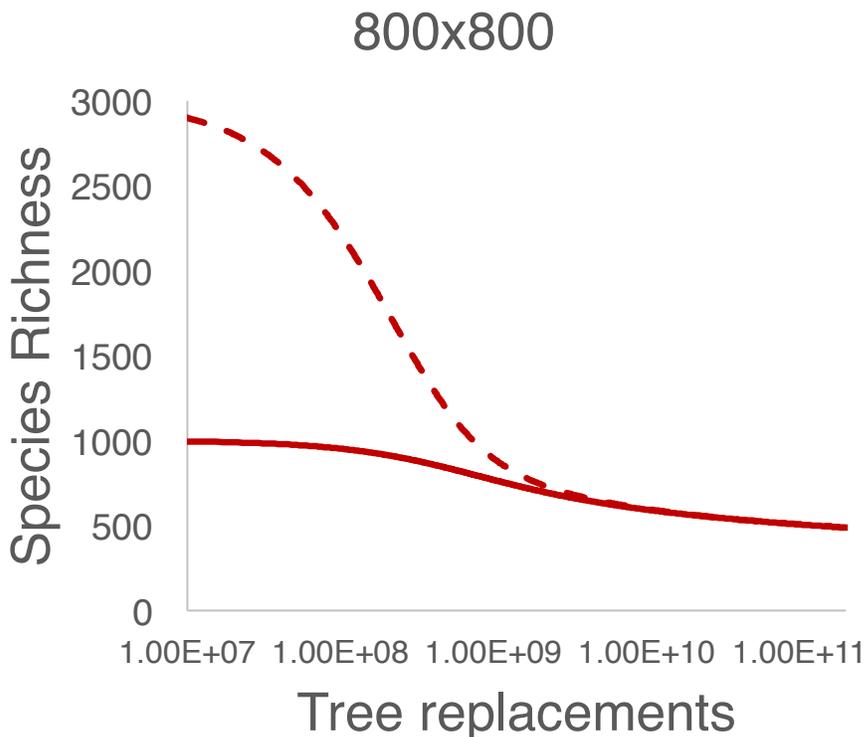
have a different species. Summing these gives  $p_7 = \binom{S}{1} \binom{S-1}{6} \frac{9!}{3!S^9} + \binom{S}{2} \binom{S-2}{5} \frac{9!}{2!2!S^9}$ . The

first term is for the case with one species in 3 cells and is directly analogous to the 8-species case above. For the second term, two species are initially chosen (those with abundance 2) and then the remaining 5 species, and the number of permutations is divided by  $2!$  for each two-cell species.  $p_m$  for other  $m$  can be found similarly.

For  $S = 100$ ,  $p_9 = 0.6903$ ,  $p_8 = 0.2701$  and  $p_7 = 0.0373$ , which sums to over 99.7% of the total probability. The contribution of these terms to the expected value of the number of species, is  $\sum_{m=7}^9 mp_m = 7p_7 + 8p_8 + 9p_9 = 8.634$ . (Including the other values of  $m$  gives an expected number of species of 8.648.) For higher  $S$ , this value would be even larger, so the assumption that an exclusion zone has 9 species if all species are equally abundant is reasonable, especially since exclusion zones might produce a more uniform spatial distribution, which should increase the expected value relative to the random arrangement case presented here.

### S3. Absorbing Markov chain with equations for a general exclusion zone

The equations for probabilities of increase and decrease of a focal species  $i$  can be used to define an absorbing Markov chain (the states are feasible values of  $N_i$ ; only state 0 is absorbing if there is an exclusion zone). The probabilities of transitions between each pair of non-absorbing states define a matrix  $\mathbf{P}$ , where  $P_{ij}$  is the probability of going from state  $i$  to  $j$ . The expected number of steps until extinction of species  $i$  starting with  $N_i = k$  is the  $k^{\text{th}}$  element of the vector  $(\mathbf{I} - \mathbf{P})^{-1}\mathbf{1}$ , where  $\mathbf{I}$  is an identity matrix and  $\mathbf{1}$  is a vector of 1s. We first calculated this for the initial number of species in a simulation,  $S_0$ , using starting abundance  $k = J/S_0$  (the mean species abundance). The calculated extinction time used the assumption that all non-focal species remain equal in abundance. Since all species vary in abundance, we divided this time to extinction of species  $i$  by  $S_0$  to approximate the time until the first extinction (when  $S$  goes from  $S_0$  to  $S_0 - 1$ ). We then repeated the procedure for  $S_0 - 1$  species, to give the time from first to second extinction, and so on for each lower value of  $S$  until the cumulative total time was at least  $2 \times 10^{11}$ .  $S$  as a function of cumulative time was then plotted for  $E=1, 9, 25$  and  $49$  for  $J = 40,000$  (Fig. 3c) and  $160,000$  (Fig. 3d). We additionally compared projected declines of species richness through time with the initial conditions  $S_0=1000$  to  $S_0=3000$  (Fig. S1). All projections based on the absorbing Markov chain using approximate equations from the text are consistent with the results from numerical simulation.



**Figure S1. Comparing projections of species richness through time with initial species richness of 3000 and 1000 for an exclusion zone of one tree crown on an 800x800 landscape. The projection from the absorbing Markov chain is based on the equations in the main text. The solutions from Markovian process are remarkably consistent with the results from simulation.**

## S4. Software to implement simulations

To implement simulations with exclusion zones on large spatial and long temporal scales, we developed an algorithm and provide a software package for high-performance computing called FEZ (Forest Exclusion Zones), with the bulk of the simulation itself written in C++ for speed. The software can be found on the Levi Lab's software page (<http://people.oregonstate.edu/~levit/software/>). FEZ packaged release was created utilizing the GNU tool Autoconf with a standard `./configure, make, make install` setup. FEZ's managing system is a python script named `controller.py` that creates and maintains the location and file structure of the data that is generated by the C++ code. It is the script that users interact with to start simulation runs. The following parameters are adjustable using `controller.py`: landscape rows, landscape columns, number of tree replacements, number of species, size of the exclusion zone, the chance of a forced replacement occurring, and how many simulations will be run on a single thread. The program writes, at regular intervals, a Comma Separated Value (CSV) file that indicates the species at each position on the landscape at that time. The program also stores the number of extinctions at regular intervals and prints these at the end of a simulation. Each of these intervals can also be specified. `Controller.py` makes use of the following python libraries: `sys, os, shutil, getopt, and subprocess`. The C++ code is called by the controller, runs the simulation, and generates the CSV files and extinction output file. The C++ code relies on the following libraries: `ostream, stdio.h, stdlib.h, time.h, iomanip, fstream, string, sstream, string.h, and Math.h`. Once the data are generated by the C++ code, an R script utilizes the libraries `knitr` and `markdown` to generate HTML output to summarize model results, including species richness and extinctions over time and rank-abundance curves.

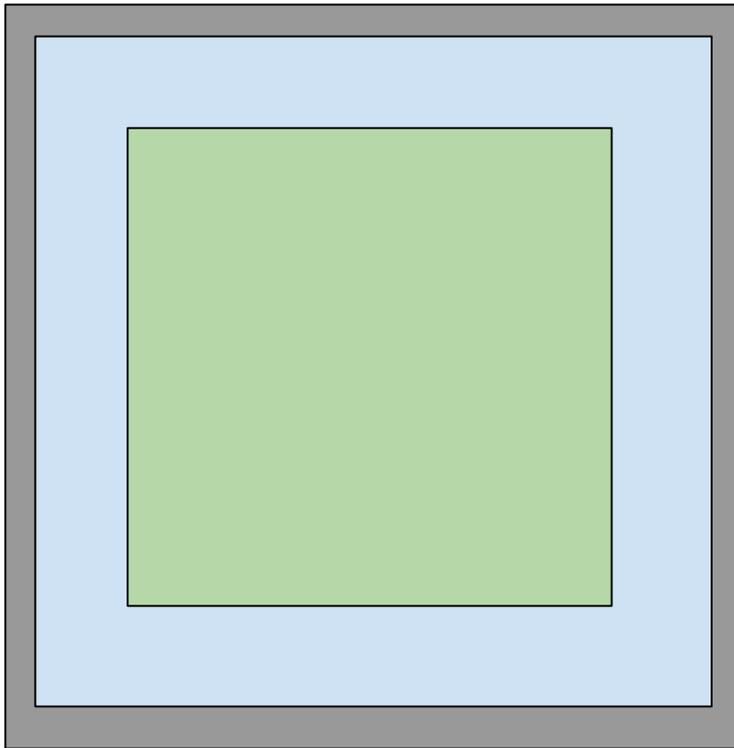
**Table S1: The libraries used by each component of FEZ**

<u>PYTHON</u>	<u>C++</u>	<u>R</u>
Sys	ostream	knitr
os	stdio.h	Markdown
shutil	stdlib.h	
getopt	time.h	
subprocess	iomanip	
	fstream	
	string	
	sstream	
	string.h	
	Math.h	

### *Algorithm*

The C++ code contains the entire FEZ algorithm and utilizes a two-dimensional matrix to model the forest landscape, which was a square lattice. For the simulations that we present, the landscapes were 200x200, 400x400, 800x800, and 1600x1600. To eliminate edge effects on our landscape, the FEZ algorithm adds 20% more rows and columns to create a functional buffer of 10% around all sides of the FEZ landscape. The cells in this buffer fully interact with the simulation but are not summarized for results in the CSV outputs or extinction snapshots. This extra space is used to minimize the influence of edge cases on the actual data generated from the simulation. Finally, another 20 rows and columns are added to produce a buffer of 10 spaces

encircling the landscape to prevent memory issues when interacting with exclusion zones. Unlike the previous functional buffer that is filled with trees, this buffer is filled with negative values that do not affect the simulation (Fig. S2).



**Figure S2: The green square represents the landscape size provided by the user, rows times columns. The blue border is a functional buffer that is 10% larger on each side, engages with our simulation, but is not included in the CSV files or snapshots. The grey border is a deadzone of size 10 on each side filled with no tree species that acts as error prevention.**

At the start of the simulation the landscape is randomly filled with tree species from the information provided by the user in the “number of species” variable. For example, if FEZ is passed 1000 initial tree species, every matrix cell is randomly filled with a single integer from the range 0-999, each representing a unique tree species. The simulation then generates a CSV file to represent this initial landscape. The extinction tracker then counts the number of existing tree species in the initial landscape.

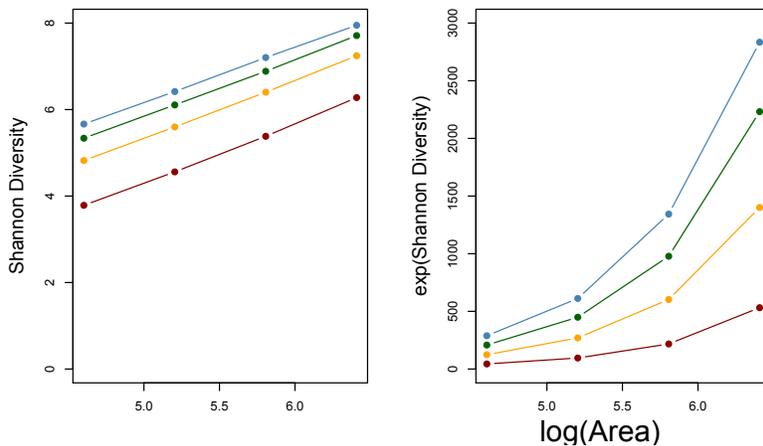
Once the initial landscape CSV file is created, the tree replacement process begins. The simulation randomly picks two matrix locations to select a *target tree* that will die and be replaced and a *sample tree* that the simulation will use to attempt to replace the target tree. Once the two trees are selected, the simulation checks the square area given by the exclusion zone surrounding the target tree to see if a tree of the sample tree’s species already exists within the exclusion zone. If there is no conspecific to the sample tree in the exclusion zone, then a tree of the sample tree’s species replaces the target tree and the next tree replacement begins. If a conspecific to the sample tree is present in the exclusion zone, then the sample tree’s species would generally be excluded, but can occasionally replace the target tree, with a probability specified by the user (the force rate). In this experiment, we used a force rate of 1/20 for a 5%

chance that the target tree replaces a sample tree even if a conspecific is present within the exclusion zone, and we additionally checked the sensitivity of our results using a 20% force rate, and 50% force rate. If the sample tree cannot replace the target tree, the simulation selects a new sample tree until a valid move is possible. When a tree is successfully replaced, the extinction tracker device updates the new tree abundances and the simulation continues. This process continues until a fixed number of tree replacements occur, which in this experiment was 200 billion.

The FEZ algorithm uses the parameters passed for number of CSVs and number of snapshots to create these files and data entries at equidistant points in time based on the number of tree replacements. This experiment used 200 CSV files and 20,000 extinction snapshots which caused CSVs of the landscape configuration to be generated every one billion tree replacements and snapshots of the number of extinctions to be stored every ten million tree replacements. The snapshotting mechanism counts how many tree species have zero abundance and stores that information in a data structure. Once the simulation has completed all tree replacements, the simulation then writes the number of extinctions over time to the file *extinction.txt*.

## S5. Results using other diversity metrics

Results from our simulations using Shannon Diversity and effective species richness (i.e. exponential of Shannon Diversity) closely mirror the results in the main text using species richness (Fig. S3).



**Figure S3. Shannon diversity and effective species richness (i.e. exponential of Shannon Diversity) after 200 billion iterations of the simulation model for the four landscape sizes (200x200, 400x400, 800x800, and 1600x1600) and four exclusion zones. Red, yellow, green, and blue refer to exclusion of dead tree only, one, two, and three tree crowns of exclusion ( $E = 1, 9, 25, \text{ and } 49$ ).**