

# Phylometh Project

*Patrick McKenzie & Rachel Swenie*

*10 May, 2017*

## Continuous Probability Biogeography Model for Range Evolution

Dispersal-extinction-cladogenesis (DEC) models are popular in phylogenetic research for reconstructing ancestral ranges (Ree and Smith 2008). DEC models describe occupancy among patches of possible habitat (e.g. islands) as discrete character states, and they employ transition matrices to describe variation in range along the branches of phylogenetic trees. However, discrete transitions might be impractical for modeling range evolution across landscapes that are fragmented by geographical or ecological barriers of varying restrictive qualities. For examples, while neither environment is ideal, a seed dispersed to a too-cold environment might be more likely to survive and reproduce than a seed dispersed into the ocean.

The model presented here uses data matrices representing extant species ranges, environmental matrices that describe the ability of species to disperse to a given cell, a movement function to describe range change within a lineage, and a speciation function to reconstruct ancestral ranges. This allows for range reconstruction without discrete categorization of species ranges, and for incorporation of geographical or ecological barriers that vary in their effects on species ranges.

## Movement and speciation

For each step in time, the following equation is applied to each cell to calculate its new “probability” value.

- These are normalized at each node and no longer represent true probabilities when moving up the tree.

$$P_{s,x,y}(t-1) = E_{x,y}(t-1)((1 - P_{s,x,y}(t)) * \alpha * \frac{\bar{N}}{8}) + (P_{s,x,y}(t) * \beta * \frac{\bar{N}}{8})$$

The same equation is written out in code below. Note that `get_Nbar()` is a separate function written to calculate  $\bar{N}$ .

```
get_prob_matrix <- function(Ps,Es, alpha, beta) {
  probs_older <- Ps
  for (i in 1:nrow(Ps)) { # rows
    for (q in 1:ncol(Ps)) { # columns
      Pcell <- Ps[i,q]
      Ecell <- Es[i,q]
      Nbar <- get_Nbar(Ps,i,q)
      Polder <- Ecell * ( ( (1-Pcell) * (Nbar/8) * alpha) + (Pcell * (Nbar/8) * beta) )
      probs_older[i,q] <- Polder
    }
  }
  probs_older
}
```

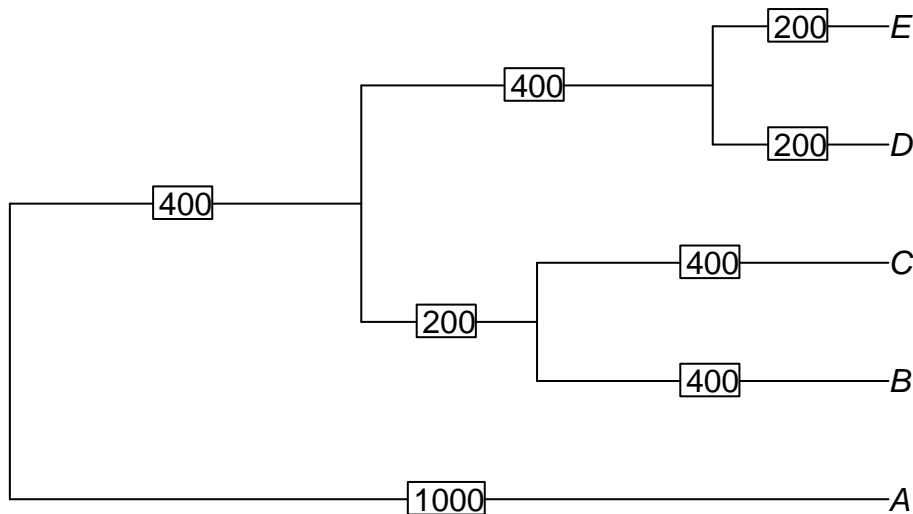
## Speciation

When a node is reached (moving from tips to root), the probability matrix for each lineage is normalized, and the two matrices are multiplied together.

## Sample tree

The tree below was generated for testing the continuous model of range evolution.

(A:1000,((B:400,C:400):200,(D:200,E:200):400):400);



## Generating random species ranges

This function accepts the number of matrix rows and columns as the first two arguments, and the number of cells to be occupied by the species as the second argument. To center a species' range around a particular point, designate the point using a two-element vector as the argument for "startingcell" – otherwise, the range will be centered around a random point.

```
get_random_species_range <- function(nrow, ncol, numbercells, startingcell = NULL) {
  rows <- nrow
  cols <- ncol
  randommatrix <- matrix(rep(0,rows*cols), nrow = rows, ncol = cols)
  if (is.null(startingcell)) {
    index <- sample(rows*cols,1)
    randommatrix[index] <- 1
    pointindices <- arrayInd(index,c(rows,cols))
  } else {
    pointindices <- startingcell
    dim(pointindices) <- c(1,2)
  }

  while(sum(randommatrix) < numbercells) {
    workingcell <- pointindices[sample(nrow(pointindices),1),]
```

```

random_adjacent <- sample(c(-1:1),2,replace = TRUE) #new cell row and col
check_occupancy <- workingcell + random_adjacent
if (sum(check_occupancy > 0) == 2 && sum(check_occupancy <= c(rows,cols)) == 2) {
  if (randommatrix[check_occupancy[1],check_occupancy[2]] == 0) {
    randommatrix[check_occupancy[1],check_occupancy[2]] <- 1
    pointindices <- rbind(pointindices,check_occupancy)
  }
}
}
}
randommatrix
}

```

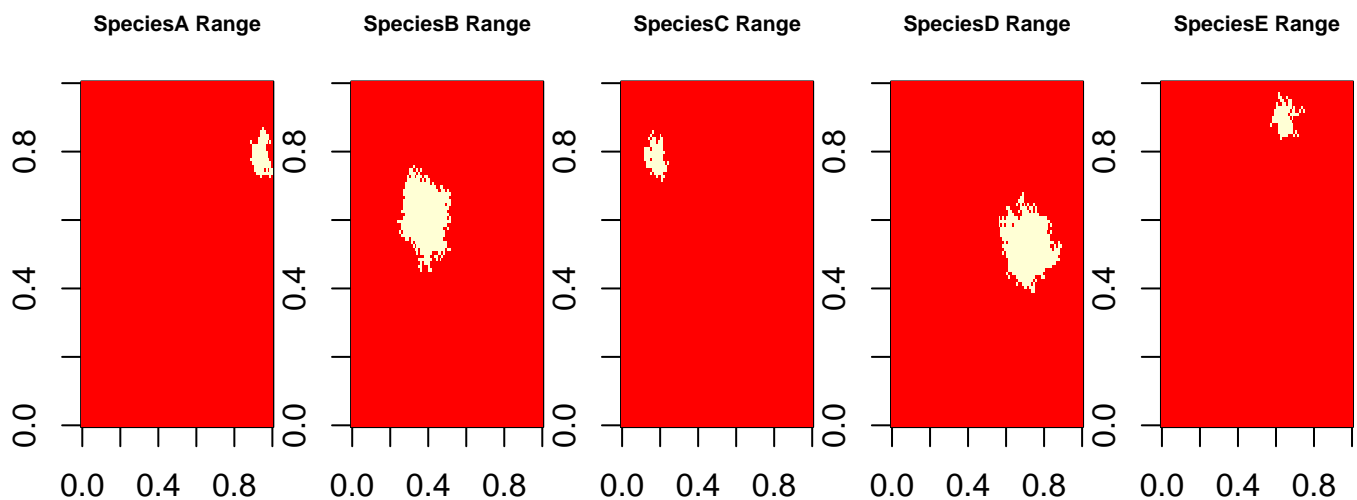
## Species range data

We set ranges for the extant lineages on the phylogeny above. We used random ranges generated by the `get_random_species_range()` function, having arbitrarily chosen occupancy of 100 cells for species A, C, and E and 500 cells for species B and D.

```

speciesA <- get_random_species_range(100,100, 100)
image(speciesA,main="SpeciesA Range")
speciesB <- get_random_species_range(100,100, 500)
image(speciesB,main = "SpeciesB Range")
speciesC <- get_random_species_range(100,100, 100)
image(speciesC,main = "SpeciesC Range")
speciesD <- get_random_species_range(100,100, 500)
image(speciesD,main="SpeciesD Range")
speciesE <- get_random_species_range(100,100, 100)
image(speciesE,main="SpeciesE Range")

```



Here, we can see the random ranges generated for each species. Now we can plot them beside our phylogeny:

```

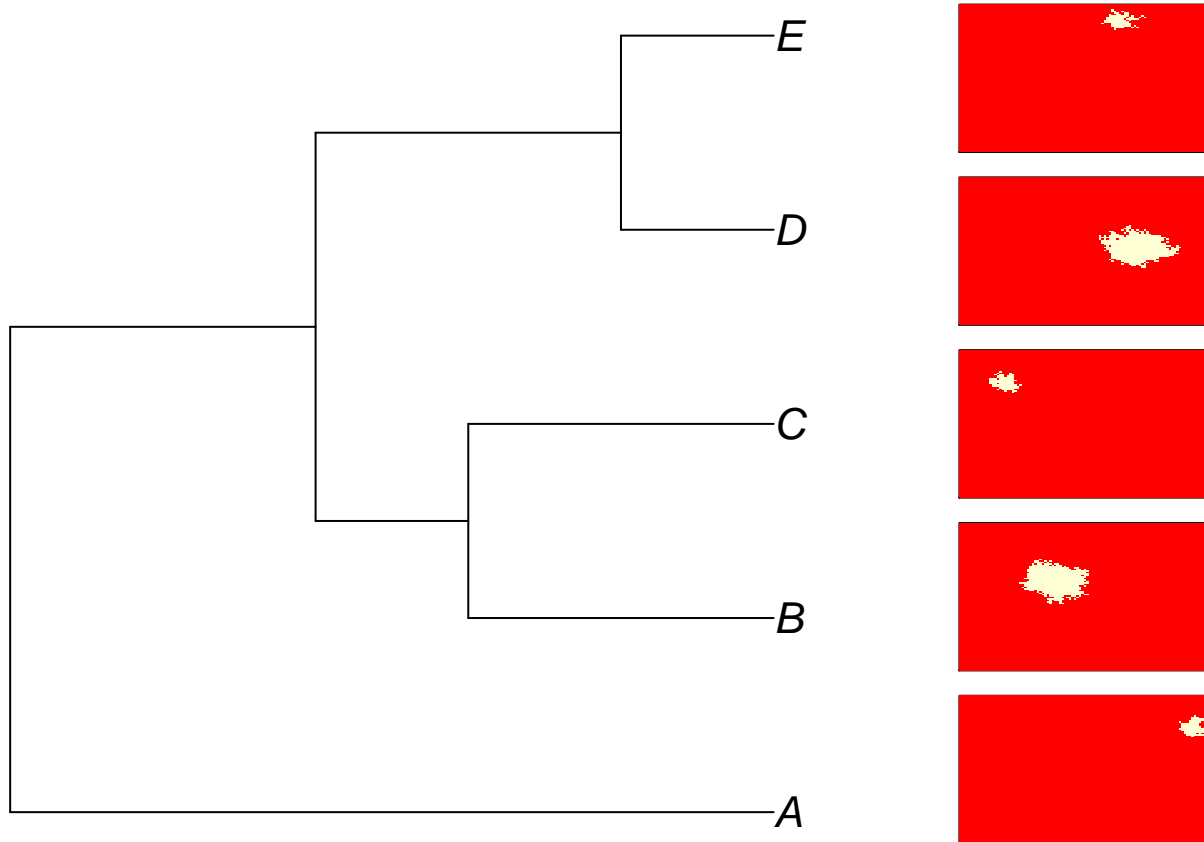
Ntip <- length(tree$tip.label)
layoutmatrix <- matrix(c(rep(1,Ntip),2:(Ntip+1)),ncol=2)

```

```

layout(layoutmatrix,widths = c(.8,.2))
par(mar = c(.5,.5,.5,0))
plot.phylo(tree,cex = 2)
par(mar = c(0.5,0,0.5,0))
image(speciesE,xaxt='n', yaxt='n', ann=FALSE)
image(speciesD,xaxt='n', yaxt='n', ann=FALSE)
image(speciesC,xaxt='n', yaxt='n', ann=FALSE)
image(speciesB,xaxt='n', yaxt='n', ann=FALSE)
image(speciesA,xaxt='n', yaxt='n', ann=FALSE)

```



## Scenario 1: Homogeneous environment

Below is an environmental matrix that is completely homogeneous. Every cell is equally colonizable by any species.

```

Es <- matrix(nrow = 100,ncol = 100)
Es[1:10000] <- 1

```

We ran our `get_prob_matrix()` function up each branch of the phylogeny to reconstruct the ancestral range at each node of the phylogeny.

```

ancE <- speciesE
ancD <- speciesD
for (w in 1:200) {
  probs <- get_prob_matrix(ancD,Es,alpha = .5,beta = 0.8)
  ancD <- probs
  probs <- get_prob_matrix(ancE,Es,alpha = .5,beta = 0.8)
  ancE <- probs
  print(w)
}
ancD <- ancD/max(ancD)
ancE <- ancE/max(ancE)
ancDE <- ancD*ancE

ancC <- speciesC
ancB <- speciesB
for (w in 1:400) {
  probs <- get_prob_matrix(ancB,Es,alpha = .5,beta = 0.8)
  ancB <- probs
  probs <- get_prob_matrix(ancC,Es,alpha = .5,beta = 0.8)
  ancC <- probs
  print(w)
}
ancB <- ancB/max(ancB)
ancC <- ancC/max(ancC)
ancBC <- ancB*ancC

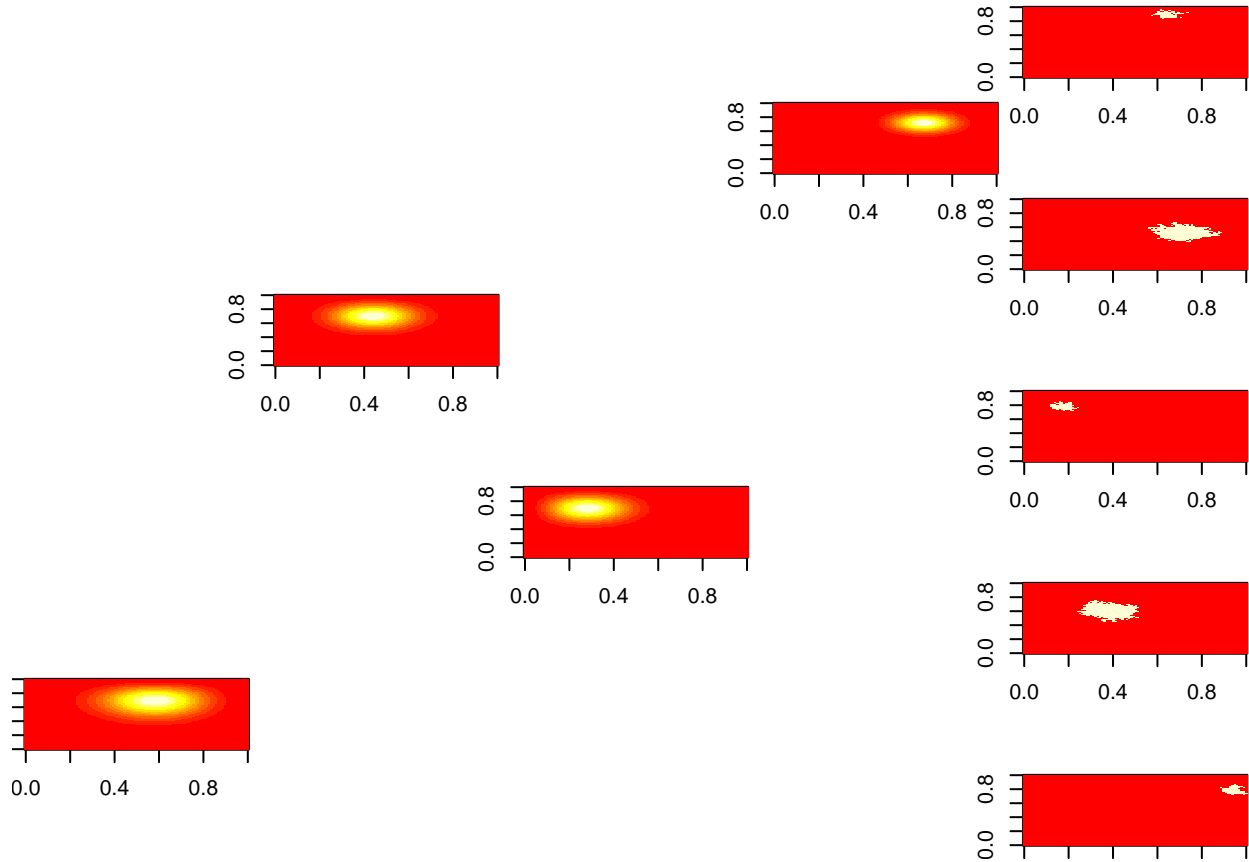
for (w in 1:400) {
  probs <- get_prob_matrix(ancDE,Es,alpha = .5,beta = 0.8)
  ancDE <- probs
  print(w)
}
for (w in 1:200) {
  probs <- get_prob_matrix(ancBC,Es,alpha = .5,beta = 0.8)
  ancBC <- probs
  print(w)
}
ancDE <- ancDE/max(ancDE)
ancBC <- ancBC/max(ancBC)
ancBCDE <- ancBC*ancDE

for (w in 1:400) {
  probs <- get_prob_matrix(ancBCDE,Es,alpha = .5,beta = 0.8)
  ancBCDE <- probs
  print(w)
}
ancA <- speciesA
for (w in 1:1000) {
  probs <- get_prob_matrix(ancA,Es,alpha = .5,beta = 0.8)
  ancA <- probs
  print(w)
}
ancBCDE <- ancBCDE/max(ancBCDE)
ancA <- ancA/max(ancA)

```

```
ancABCDE <- ancA*ancBCDE
```

We can look at the ancestral states plotted as nodes on the phylogenetic tree:

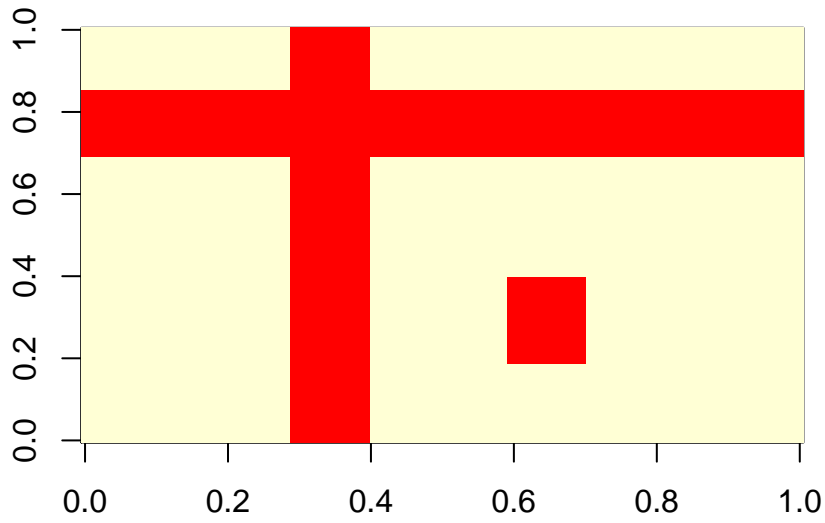


## Scenario: Adding bands of poorer-quality environment

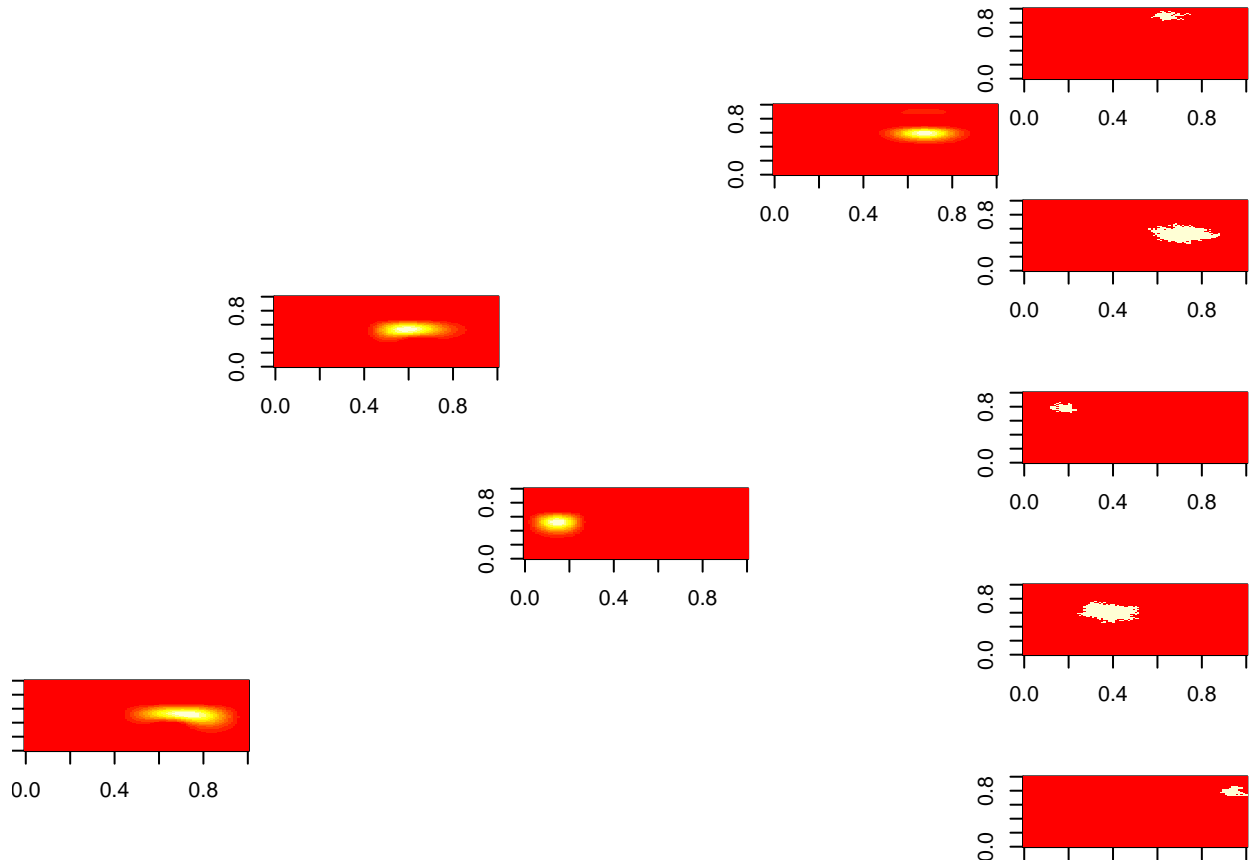
Heterogeneity was added in the environmental matrix to test its effects on ancestral state outcomes. Bands of less-hospitable environment are added below.

```
Es <- matrix(nrow = 100, ncol = 100)
Es[1:10000] <- 1

Es[, 70:85] <- .8
Es[30:40,] <- .8
Es[60:70, 20:40] <- .8
image(Es)
```



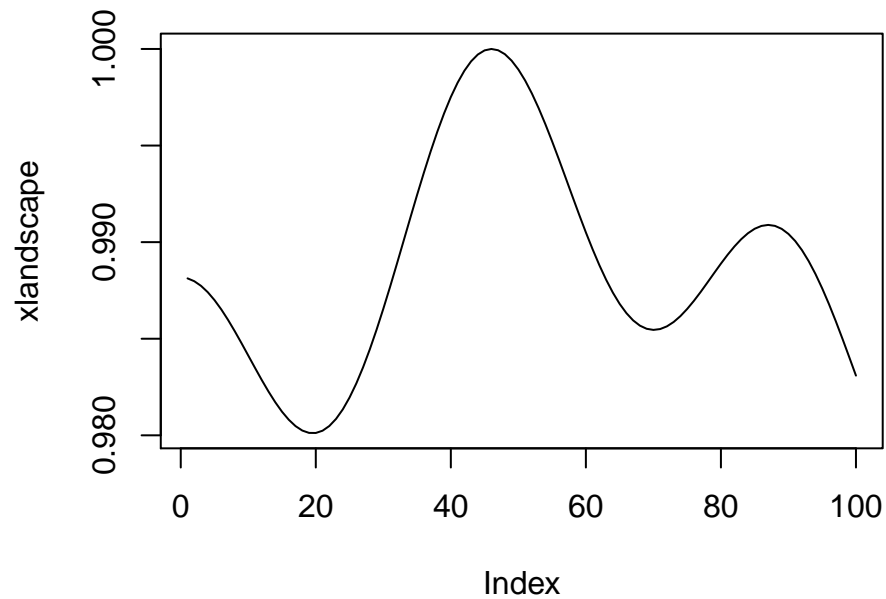
Now, again, we can see how ancestral states are reconstructed on the tree.



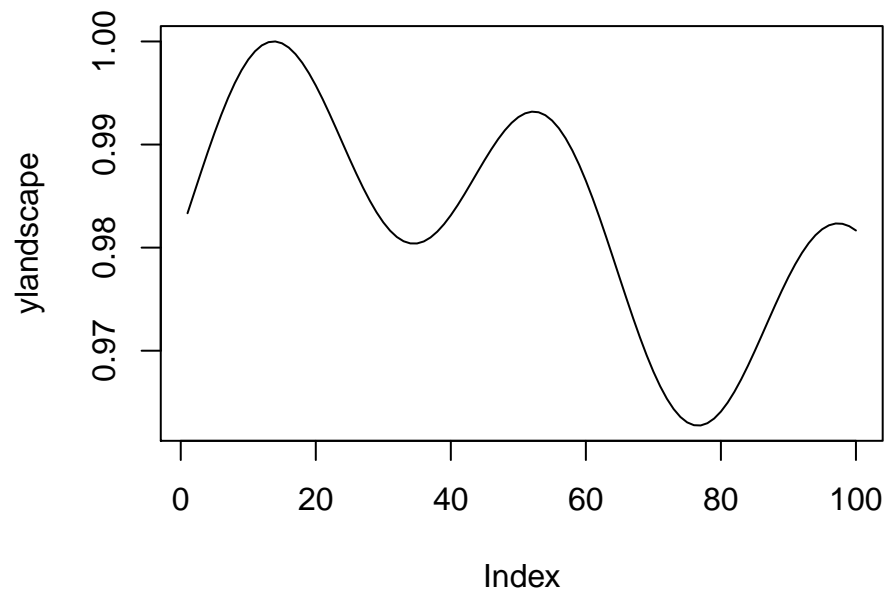
## Scenario: Hotspot Environment

A new environmental matrix with “hotspots” of hospitable regions was generated by two intervering trig functions.

```
x <- 1:100
xlandscape <- (sin(.04*x))/9 * (sin(.1*x))/9
xlandscape <- xlandscape-min(xlandscape)
xlandscape <- 1 - xlandscape
plot(xlandscape,type="l")
```



```
y <- 1:100
ylandscape <- sin(.05*y +.4*pi)/7 * sin(.1*y +pi)/7
ylandscape <- ylandscape-min(ylandscape)
ylandscape <- 1 - ylandscape
plot(ylandscape,type="l")
```



```
Es <- matrix(nrow = 100,ncol = 100)
Es[1:10000] <- 1

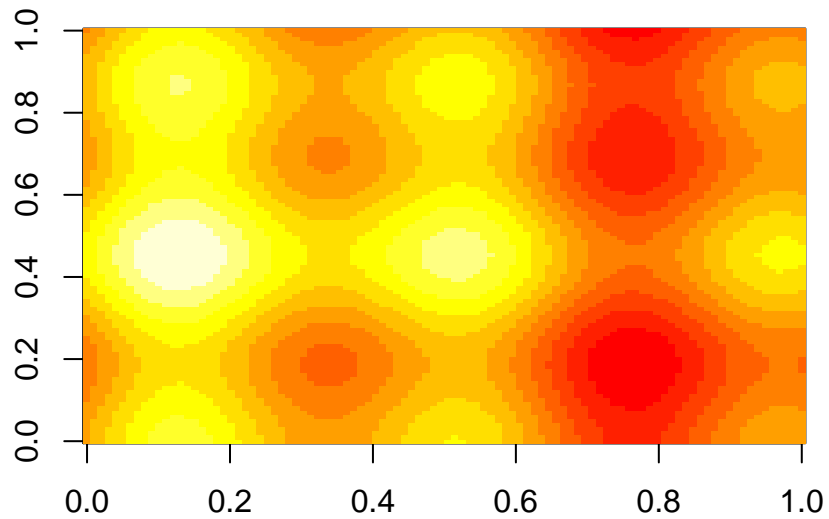
for(i in 1:100) {
```



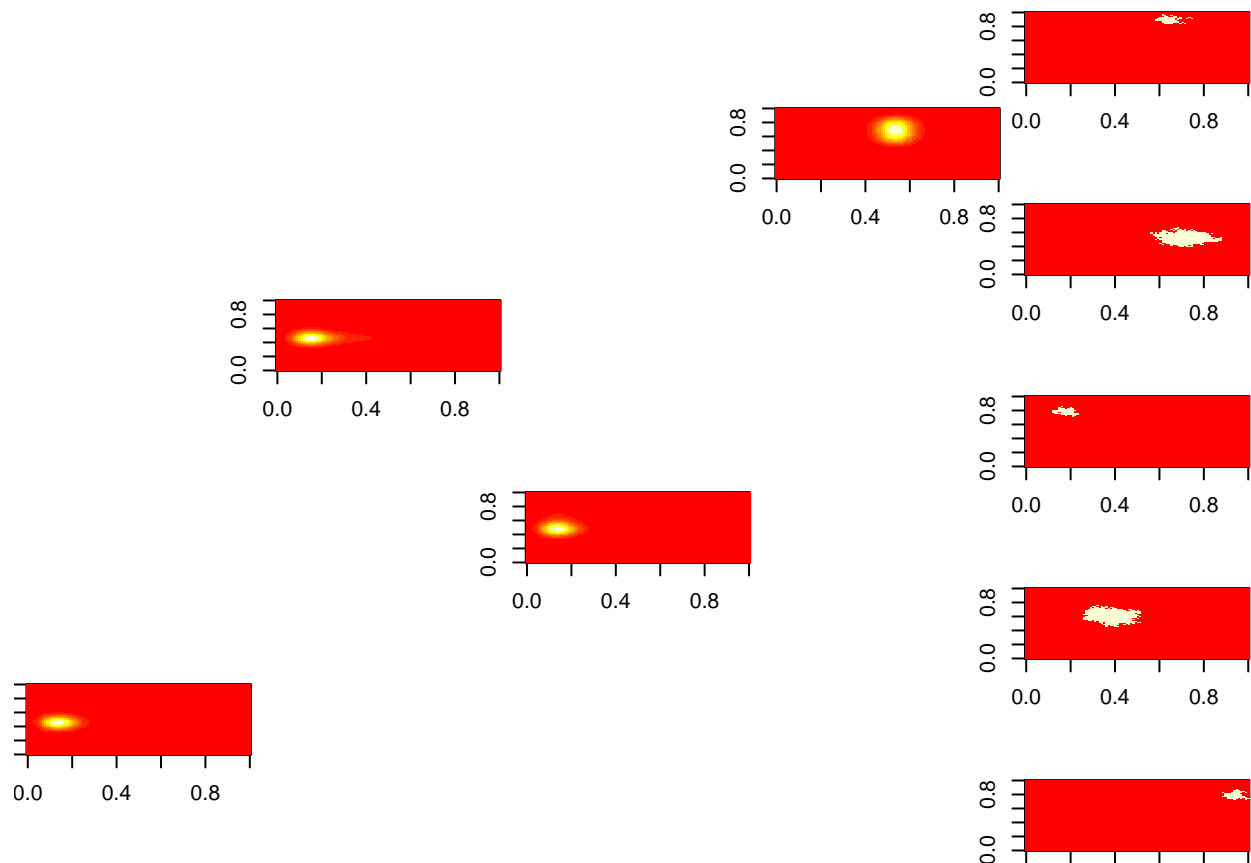
```

  Es[,i] <- Es[,i]*xlandscape[i]
}
for(i in 1:100) {
  Es[i,] <- Es[i,]*ylandscape[i]
}
image(Es)

```



Now we can see how this influences the ancestral state predictions.



## Scenario: Variation in $\alpha$ and $\beta$ values

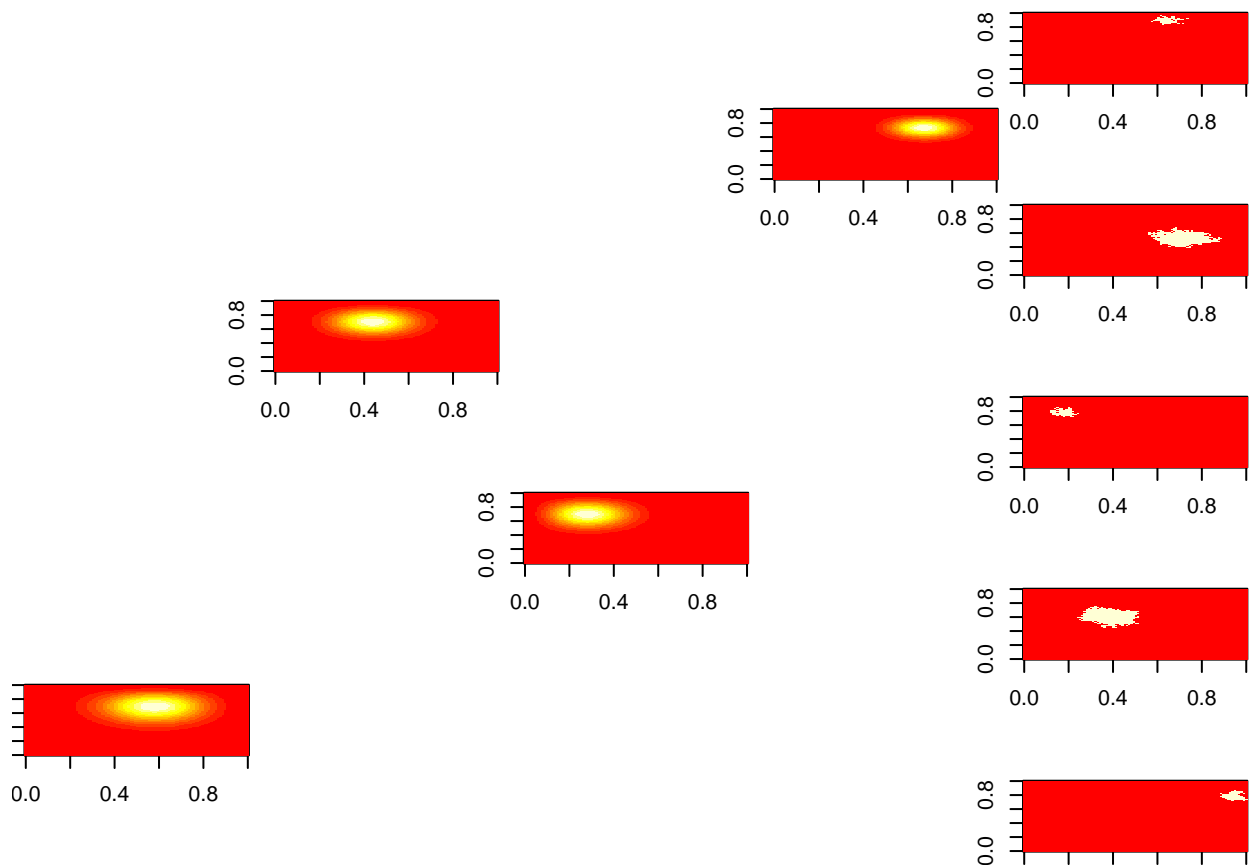
The sensitivity of ancestral range reconstruction to variation in alpha and beta values was tested in the following section.

Three tests are included:

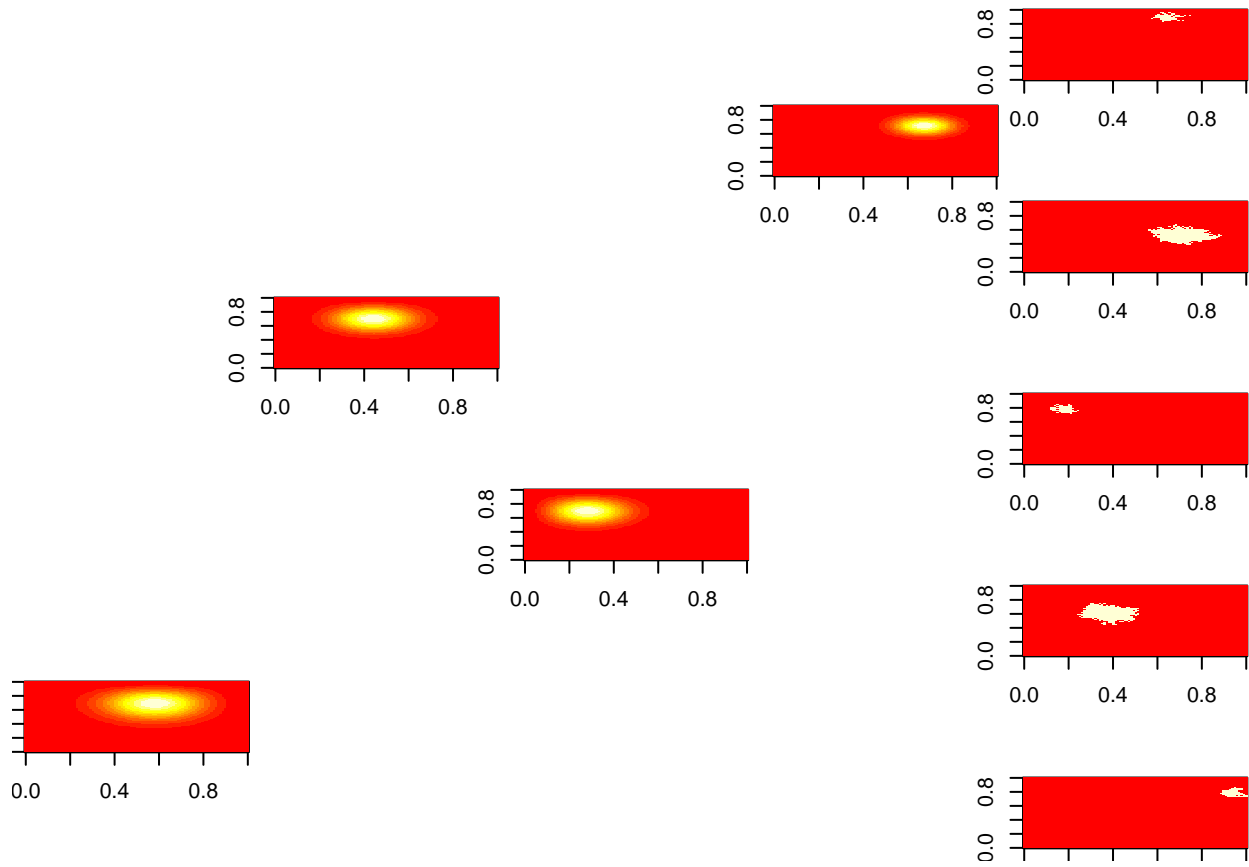
- 1)  $\alpha = 0.8$  and  $\beta = 0.2$
- 2)  $\alpha = 0.2$  and  $\beta = 0.8$
- 3)  $\alpha = 0.5$  and  $\beta = 0.5$

Similarity among the results shows that the resulting ancestral state reconstructions are not sensitive to variation in alpha and beta values.

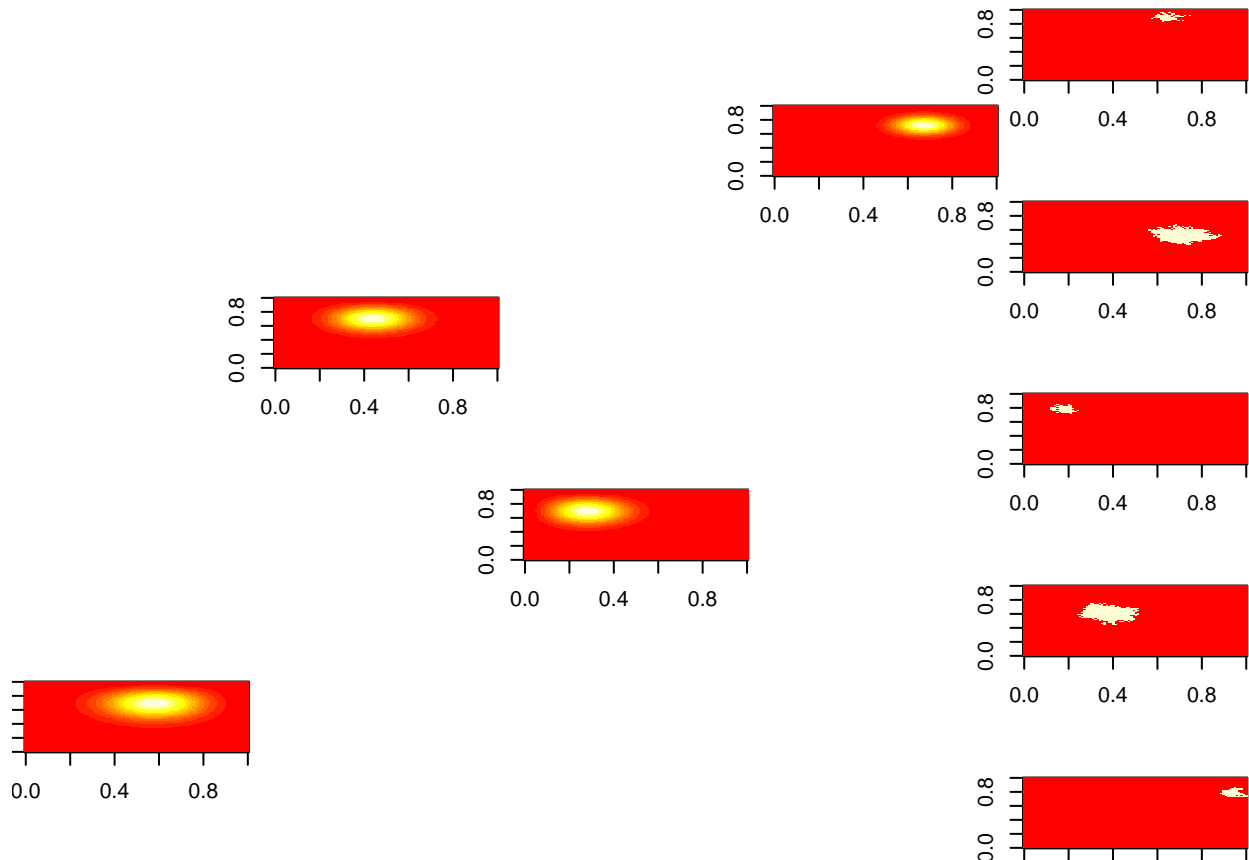
$\alpha = 0.8, \beta = 0.2$ , homogeneous environment



$\alpha = 0.2, \beta = 0.8$ , homogeneous environment



$\alpha = 0.5, \beta = 0.5$ , homogeneous environment



## Discussion

As expected, the model responds to heterogeneous environments by predicting ancestral ranges to fall among the better-quality habitats. Given a homogeneous environment, the ancestral ranges are not sensitive to changes in the alpha and beta parameters. This is likely due to normalization before the speciation function is applied at each node, which is performed to prevent probabilities from becoming too small and to keep from weighting speciation in favor of the species with the largest starting ranges.

## Use of the model

### Environmental heterogeneity

This model is particularly useful when considering heterogeneous environments, where inhabitation of some inhospitable habitats might be more probable than the inhabitation of others.

### Temporal heterogeneity

This model allows environmental favorability of cells to change through time. Unlike DEC models, which only allow discrete “islands” to appear or disappear through time, the continuous range evolution model allows gradual changes in properties of the environmental matrix through time. For example, the favorability of specific peaks in a mountain range might be adjusted across time to account for changing climate.

## Some Remaining Problems

### Probabilities of cell occupancy

Ideally, the matrices at ancestral states would represent probabilities of cell occupancies. Perhaps it would be best to represent the surface created by the ancestral matrices as a cumulative density function, so that integration over a specific region would return a probability of occupancy. The movement function will need to be adjusted to allow this.

### Speciation

Ideally, different modes of speciation could be represented by different speciation functions. As written, the speciation function favors an ancestral range where the ranges of the two descendent lineages overlap. Also, because both lineages are normalized before the speciation function is applied, the uncertainty associated with longer branches is ignored.

### Environmental matrix for ancestral lineages

The environmental matrix can be adjusted to allow the colonizability of specific cells to change through time. However, the abilities of ancestral lineages to tolerate specific conditions might not be known, so the uncertainty associated with the effect of some environmental conditions (e.g. high temperatures) on cell occupancy will increase toward the root of the tree.

## References

Ree, Richard H, and Stephen A Smith. 2008. "Maximum Likelihood Inference of Geographic Range Evolution by Dispersal, Local Extinction, and Cladogenesis." *Systematic Biology* 57 (1). Oxford University Press: 4–14.