Boreal and Neotropical Population Simulation Software

Nihshanka Debroy[1]: Work with Dr. Ganesh Ganapathy[2]

January 8, 2008

[1] Department of Computer Science, Duke University, Durham, NC 27708 and [2] National Evolutionary Synthesis Center, Durham NC 27705

1   What is it?

This code simulates two population systems, boreal and neotropical, with speciation, extinction, turnover and migration events following an exponential distribution. For a detailed description of the model, consult [1].

2   How do I use it?

At the console, just type *python bd_m.py* to run the code. You will be asked to enter the values of several input parameters:

1. Speciation rate

2. Extinction rate

3. Migration rate

4. Turnover rate

5. Number of events to generate STARTING population in boreal region

6. Number of events to generate STARTING population in neotropical region

7. Number of events for the simulation

After providing all the inputs, the code will first generate starting populations in the boreal and neotropical regions. Then, the simulation (migrations and turnovers for the boreal region or speciations and extinctions for the neotropical region) is carried out.

3   Interpreting the Output - Examples

The existing species and event history for each population region are saved.

1. **Species**: Each species has 3 pieces of information [Name, Flag1, Flag2]. Flag1 is set to -1 if the species has not migrated and 1 otherwise. Flag2 is set to 0 if the species was born in the boreal region and 1 otherwise.

   Examples: [0, -1, 0] means that species 0 is in the boreal region and has not migrated yet. [1, -1, 1] means that species 1 is in the neotropical region.

   Flag1 does not have any relevance for species born in the neotropical region, as they cannot migrate to the boreal region (as per the modeling assumptions). Also, if a species from the boreal region has already migrated, it is not allowed to migrate again.

2. **Events**: Each event has 5 pieces of information [Event no., Time of Event, Left Child, Right Child, Parent node]. In the building of the tree, the events are represented as nodes, and the species as edges. For a birth, the left child of the node denotes the species undergoing birth, and the right child indicates the species that has been born.

   Examples: [3, 1.5, 2, 4, 1] means that at event 3, which occurred at 1.5 seconds, species 2 went through a birth that gave rise to 4, and the parent node was event 1. If there is no left/right child, that corresponding entry is set to -1. So, [2, 1, -1, -1, -1] would mean that at event 2, which occurred after 1 second, a death took place. The code then looks for the most recent edge corresponding to the dying species and puts a -1 there.

After the starting populations in the boreal and neotropical regions are created, the species in existence and the event histories are printed out. Once the simulation is carried out, the user is asked if he/she would like to see the lineage of a particular living species. The user can pick the species from any of the two regions, by entering the index of the species of interest from the population array. Note that the indices begin from 0.

4   Sample Run

Here is an example of running the code, along with sample inputs:

ndebroy@lilac> python *bd_m.py*
Please enter the speciation rate:2
Please enter the extinction rate:.5
Please enter the migration rate:2
Please enter the turnover rate:.5
Please enter desired number of events to create STARTING population in boreal region:5
Please enter desired number of events to create STARTING population in neotropical region:5
Please enter desired number of events for simulation:3

After some explanatory text is provided, the starting populations are created as follows:

Generating starting population in boreal region...

0 undergoes birth.
0 undergoes birth.
2 undergoes birth.
1 undergoes death.
3 undergoes birth.

Generating starting population in neotropical region...
0 undergoes birth.
1 undergoes birth.
1 undergoes birth.
3 undergoes death.
1 undergoes birth.

Printing out the starting populations and event histories for the simulations...

Boreal:
Population: [[0, -1, 0], [2, -1, 0], [3, -1, 0], [5, -1, 0]]
Event history: [[1, 0.36797463768555516, 0, -1, -1], [2, 0.5558306930356709, 0, 2, 1], [3, 0.6946434445105969, 2, 3, 2], [4, 0.85136074542781581, -1, -1,-1], [5, 1.029406624970483, 3, 5, 3]]

Neotropical:
Population: [[0, -1, 1], [1, -1, 1], [2, -1, 1], [5, -1, 1]]
Event history: [[1, 1.2596693554706782, 0, 1, -1], [2, 1.3503309307739648, 1, 2, 1], [3, 1.446244088837104, 1, -1, 2], [4, 1.4907685652575475, -1, -1, -1], [5, 1.5631995304672572, 1, 5, 3]]

At this point, the simulation starts:

SIMULATION...

Migration event:
Species 2 moves over from the boreal region
Speciation event:
1 undergoes birth.
Extinction event:
2 undergoes death.

Final population in boreal region: [[0, -1, 0], [2, 1, 0], [3, -1, 0], [5, -1, 0]]
Final population in neotropical region: [[0, -1, 1], [1, -1, 1], [5, -1, 1], [2, -1, 0], [7, -1, 1]]

The user is now asked if he/she would like to print out the lineage of a particular species. This continues as long as the user wishes to. Here is an example:

Do you wish to print out the lineage of a species (y/n)?y
From boreal or neotropical region (b/n)?n
You chose the NEOTROPICAL region
Enter index of species of interest in neotropical pop. array (indices start from 0):1

You picked species 1
Species: 1
1 was born from 0 at time 1.25966935547
Species: 0 No ancestor in current population.
Do you wish to print out the lineage of a species (y/n)?n
ndebroy@lilac>

5   Known Issues and Scope for Improvement

1. **Names of Species:** Currently, species are named with numbers, and an organism born later on is named a higher number. A potential problem could have come up when creating starting populations for the boreal and neotropical regions - For example, what if there is a species 0 in both populations? How does one differentiate between the two? To solve this, in the current implementation, there is a separate flag (Flag2, described earlier in Section 3) for each species - it is set to 0 if the species is in the boreal region and 1 otherwise. But a better way of assigning unique names to species would simplify matters.

2. **Special Cases:** Suppose at the first event, the only species in the population dies out. At the second event, let's say a species is born. A node corresponding to the event is added to the tree, and both the node's edges are labeled with the species name. Normally, the left and right edges of a node are not identical. But for this rare case, they are left as identical. It has no effect on the execution - the results are unchanged. Yet, having a duplicate edge is unnecessary.

3. **Ancestry of Extinct Species**: The current implementation does not allow the user to see the ancestry of extinct species, although the information is saved. It may be of interest in some cases.

4. If species A's parent is extinct, then no ancestry is printed out for A.

References

[1] G. Ganapathy, M. Uyenoyama. The Effect of Geography on Species Diversification Rates: A Coalescent-based Approach, 2007.