**code1\_sdm\_bg**

**Summary**

This code takes the the distribution information for scolopendrid centipedes in peninsular India (including *Otostigmus* and *Cormocephalus* species, which are not a part of this study) and uses WorldClim layers along with a soil type layer to predict habitat suitability of sampled locations across species. This prediction map can be used as a sampling bias layer since it predicts habitat suitability based on environmental conditions that have been surveyed. 10,000 background points are randomly chosen from this sampling bias layer and used to build species-specific species distribution models in subsequent code.

**Input**

**Location data:**

Scolopendridae\_5May20.csv

**Environmental variables:**

WorldClim climate and elevation rasters

soil\_stype raster from the ATREE spatial archive – polygons were rasterized to obtain this layer

**Details**

Maxent defaults are used to run the species distribution model.

**Output**

Apart from other components of the model input and output that are saved to disk, the dataframe that is used in downstream analysis is biasBg.csv. This consists of background points that are randomly chosen probabilistically from the sampling bias layer for the extent of peninsular India.

**code2\_sdm\_enmeval**

**Summary**

This code takes the distribution information for each centipede species (three genera from Otostigminae – *Digitipes*, *Ethmostigmus* and *Rhysida*) in peninsular India and uses different environmental predictor datasets (preds, preds.bc2 and pred.eco) to predict species distributions. For each species and predictor dataset, models are run using six different feature class combinations and 10 different regularization multipliers.

**Input**

**Location data:**

gps\_9Jun20.csv derived from Scolopendridae\_5May20.csv

**Background locations:**

biasBg.csv

**Environmental variables:**

A combination of different climate layers and the elevation layer available from WorldClim and soil\_stype layer from the ATREE spatial archive. The variables included in each predictor dataset are listed in Appendix S2.

**Details**

For species with more than 19 occurrence locations, geographically masked partitions are created for assigning test and training points. For species with fewer occurrences, jackknifing is done over the occurrence locations. The predictions are clamped for environmental values falling outside the observed range based on presence locations.

**Output**

The model outputs are stored in respective folders named using species names that are automatically generated. For each species, the output consists of -

[species]\_[predictor\_dataset]: the output of the ENMevaluate function for each species and predictor dataset

[species]\_[predictor\_dataset]\_occ.csv: presence locations used in the model and partition ids

[species]\_[predictor\_dataset]\_bg.csv: background locations used in the model and partition ids

[species]\_[predictor\_dataset]\_results.csv: ENMevaluate output for each species and predictor dataset across feature class and regularization multiplier combinations

**code3\_sdm\_enmeval\_results**

**Summary**

This code takes ENMevaluate results (obtained from the previous script) for each species and processes them to select models within each predictor dataset (across different feature class and regularization multiplier combinations) based on evaluation criteria defined by us. We use two methods of model evaluation – one which uses indices of model transferability and performance, and another which selects models based on AICc values. Different outputs associated with the selected model are saved to disk.

**Input**

[species]\_[predictor\_dataset]: the output of the ENMevaluate function for each species and predictor dataset

**Details**

Nothing much to add to the summary – this bit of code accesses different part of the model results within the ENMevaluate object, subsets models for each species and prediction dataset based on model evaluation criteria and provides summaries of these selected models.

**Output**

[species]\_[predictor\_dataset]\_[model\_selection\_method]\_results.csv: Maxent results for the best model selected based on evaluation criteria

[species]\_[predictor\_dataset]\_[model\_selection\_method]\_model: Maxent model object corresponding to the best model selected based on evaluation criteria

[species]\_[predictor\_dataset]\_[model\_selection\_method]\_map.tif: continuous Maxent prediction raster

[species]\_[predictor\_dataset]\_[model\_selection\_method]\_map.jpeg: continuous Maxent prediction map

[species]\_[predictor\_dataset]\_[model\_selection\_method]\_varCon.jpeg: percent contribution of predictor variables to the model

[species]\_[predictor\_dataset]\_[model\_selection\_method]\_perImp.jpeg: permutation importance of predictor variables to the model

[species]\_[predictor\_dataset]\_[model\_selection\_method]\_response.jpeg: response curves of each of the predictor variables

[species]\_[predictor\_dataset]\_[model\_selection\_method]\_[model\_threshold]\_map.pdf: predicted distribution map aggregated to 100\*100 km grid cell and thresholded (using one of three threshold values – maximum sum of sensitivity and specificity, minimum training presence and 10th percentile of training presence) – cropped only to represent the Western Ghats

[species]\_[predictor\_dataset]\_[model\_selection\_method]\_[model\_threshold]\_biodiverse.csv: saves aggregated and thresholded prediction map into a dataframe, where presences are assigned a value of 1 and absences a value of 0.

**code4\_biodiverse\_create\_input**

**Summary**

This code uses the best model for the predictor dataset bc2 selected using evaluation criteria related to model transferability and performance (named “rank” in the code) and generates predicted thresholded distribution maps at the native resolution (1x1 km) and at the aggregated resolution (100x100 km) for each species. The aggregated thresholded map is then converted into a dataframe where presence is given a value of 1 and absence a value of 0. This forms the input for the Biodiverse program. The phylogenetic tree with relationships between all species is pruned and renamed so that tip labels match the species names in the Biodiverse input file.

**Input**

[species]\_bc2\_rank\_model

[species]\_bc2\_occ.csv

[species]\_bc2\_rank\_map.tif

threegenera.tre

tiplabels.csv

**Details**

The predicted binary maps are thresholded based on maximum sum of specificity and sensitivity. Apart from predicted presences, species occurrence locations are also given a value of 1 – because they may not have been predicted accurately by the model. For *Rhysida sp. 1* and *R. sada*, since there was a lot of variation between predictions across predictor data-sets and model evaluation methods, only observed presence locations were considered and model predictions were ignored. There were no presence locations for *R. konda* (an Eastern Ghats species) in the Western Ghats and thus it is not included in the Biodiverse input file. The predicted distributions for *Ethmostigmus conooranus* and *E. sahyadrensis* were modified to remove small areas of predicted distribution considerably distant from known occurrence locations. For the former predictions above 16 degrees N were removed, while for the latter predictions south of 14 degrees N were removed.

The tip labels in the phylogenetic tree are relabelled to match the Biodiverse input file. The tree is also pruned to remove possible sus-species level divisions (for *Digitipes jonesii*, *D. coonoorensis*, *D. barnabasi* – the branches which were dropped were suggested by Jahnavi – the longest branch was retained as a representative of the species as a whole.

**Output**

[species]\_[predictor\_dataset]\_[model\_selection\_method]\_map\_1km.pdf: predicted thresholded map at native resolution, overlaid with species occurrence locations

[species]\_[predictor\_dataset]\_[model\_selection\_method]\_map\_100km.pdf: predicted thresholded map at 100 km resolution

biodiverse\_[date].csv: input file for biodiverse program where the first two columns give the coordinates for the centroids of the cells within WG and the following column indicates presence (1) or absence (0) for a species (each column is occurrence data for a specific species)

tiplabels\_biodiverse.csv: mapping of species names in the biodiverse input file with the tip labels of the phylogenetic tree

threegenera\_biodiverse.tre: pruned phylogenetic tree with matching tip labels

**code5\_biodiverse\_run\_instructions**

Instructions are using GUI on a ubuntu system – details as in the file

**code6\_biodiverse\_output\_analysis**

**Summary**

This code takes the results from Biodiverse analysis and uses the observed values of diversity and endemism and their percentile values within the randomised distribution to generate plots. Taxonomic and phylogenetic turnover are calculated using the betapart package and cluster analysis is used to classify grid cells in the Western Ghats. Turnover is plotted in colorspace using instructions from the gdm vignette.

**Input**

biodiverse\_16Mar21.csv

threegenera\_results\_16Mar21.csv

threegenera\_struc\_randomization\_results\_16Mar21.csv

threegenera\_biodiverse.tre

**Output**

div\_end\_[number].pdf

simpson\_upgma\_trees.pdf

simpson\_map.pdf

simpson\_map\_rgb.pdf

phylosor\_turn\_map\_rgb.pdf

**code7\_plots**

**Summary**

This code generates the maps in Figure 1 of the manuscript and some plots (Figure S2.1 and S2.2) available in the appendices.

**Input**

supp\_table\_locations.csv

biasMap.tif

biasBg.csv

results\_summary\_5Apr21.csv

**Output**

sampling\_locations\_wg\_outline.pdf

sampling\_locations\_wg.pdf

bias\_map.pdf

variable\_importance.pdf

**code8\_sdm\_calc\_PD\_PE**

**Summary**

This code calculates indices of diversity and endemism using continuous habitat suitability maps at 1x1 km resolution. These indices are calculated genus-wise and summed across the three genera. These indices are plotted to obtain Figure S3.4.

**Input**

calc\_div\_end\_22May20.R

[species]\_bc2\_rank\_map.tif

[genus].tre

wg\_boundary\_mod\_23Jun20

**Output**

SR.tif

WE.tif

PD.tif

PE.tif

diversity\_measures\_1km\_continuous.pdf