Processing Occurrence Data

Emily Bellis

8/31/2020

Description

This R markdown document describes cleaning of occurrence data and generation of background points for *Cuscuta* species described in Masanga *et al.* (submitted).

Load libraries

```r
library(raster)
```

```
## Loading required package: sp
```

```r
library(maps)
library(rgdal)
```

```
## rgdal: version: 1.4-6, (SVN revision 841)
##  Geospatial Data Abstraction Library extensions to R successfully loaded
##  Loaded GDAL runtime: GDAL 2.4.2, released 2019/06/28
##  Path to GDAL shared files: /Library/Frameworks/R.framework/Versions/3.6/Resources/library/rgdal/gda
##  GDAL binary built with GEOS: FALSE
##  Loaded PROJ.4 runtime: Rel. 5.2.0, September 15th, 2018, [PJ_VERSION: 520]
##  Path to PROJ.4 shared files: /Library/Frameworks/R.framework/Versions/3.6/Resources/library/rgdal/p
##  Linking to sp version: 1.3-1
```

```r
library(rgbif)
```

```
## Warning: package 'rgbif' was built under R version 3.6.2
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:raster':
##
##     intersect, select, union
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(dismo)
data(worldMapEnv)
```

Field data for Kenya collected by Runo lab

```
occ <- read.csv("Data/Cuscuta_georeferenced.csv")
cr <- subset(occ, Cuscuta.Species=="C. reflexa")[,1:2]

occ <- read.csv("Data/Cuscuta_georeferenced_v2.csv") # updated 9/28 from Joel w/ East African Herbarium
cc <- subset(occ, Species=="C. campestris")[,1:2]
ck <- subset(occ, Species=="C. kilimanjari")[,1:2]
```

C. reflexa

```
# no African occurrences in GBIF
len.cr <- nrow(unique(cr))

## Sample background following https://rspatial.org/raster/sdm/3_sdm_absence-background.html#
env <- raster("~/Desktop/Projects/StigaxSorghum/MacroecologyMS/env7.rs.tif")

write.table(cr, file="Data/Creflexa.occ.txt", quote=F, row.names = F, col.names = T)

coordinates(cr) <- ~Long+Lat
projection(cr) <- CRS('+proj=longlat +datum=WGS84')

# circles with a radius of 300 km
x <- circles(cr, d=300000, lonlat=TRUE)
pol <- polygons(x)

# sample randomly from all circles
samp1 <- spsample(pol, 1000, type='random', iter=25)

# get unique cells
cells <- cellFromXY(env, samp1)
cells <- unique(cells)
xy <- xyFromCell(env, cells)

map(database="world", regions=c("Kenya","Tanzania","Ethiopia","Uganda","Rwanda","Burundi","Sudan","South
points(xy, cex=0.75, pch=20, col='blue')
points(cr$Long, cr$Lat, col="red", pch=20)
```
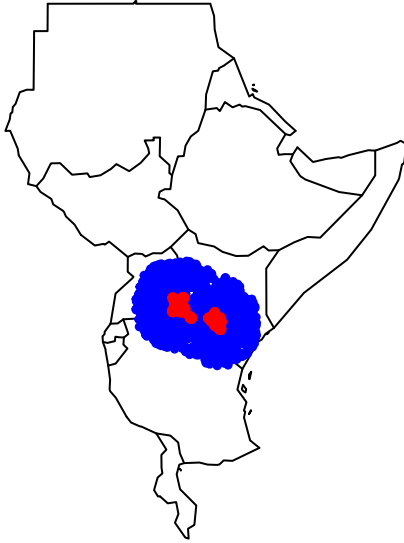
```
write.table(xy, file="Data/Creflexa.bg.txt", quote=F, row.names = F, col.names = T)
```

C. kilimanjari

```r
# combine w/GBIF data
#citation: GBIF.org (31 August 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.muwe3t
gbif.ck <- read.csv("Data/C.kilimanjari_GBIF_JM.csv", header = T)
gbif.ck <- gbif.ck %>% select(Lat, Long)

ck.all <- rbind(ck, gbif.ck)
write.table(ck.all, file="Data/Ckilimanjari.occ.txt", quote=F, row.names = F, col.names = T)

len.ck <- nrow(unique(ck.all))

## Sample background
coordinates(ck.all) <- ~Long+Lat
projection(ck.all) <- CRS('+proj=longlat +datum=WGS84')

# circles with a radius of 50 km
x <- circles(ck.all, d=300000, lonlat=TRUE)
pol <- polygons(x)

# sample randomly from all circles
samp1 <- spsample(pol, 1000, type='random', iter=25)

# get unique cells
cells <- cellFromXY(env, samp1)
cells <- unique(cells)
xy <- xyFromCell(env, cells)

map(database="world", regions=c("Kenya","Tanzania","Ethiopia","Uganda","Rwanda","Burundi","Sudan","South
points(xy, cex=0.75, pch=20, col='blue')
points(ck.all$Long, ck.all$Lat, col="red", pch=20)
```
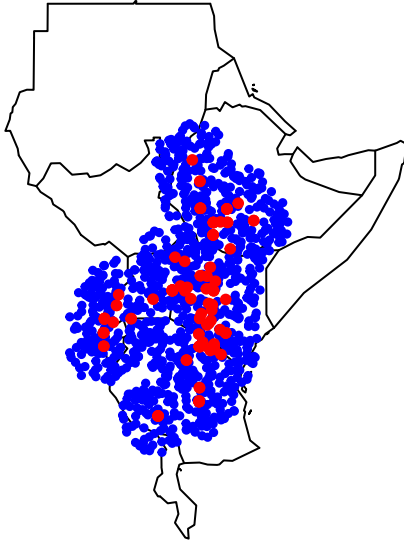
```r
write.table(xy, file="Data/Ckilimanjari.bg.txt", quote=F, row.names = F, col.names = T)
```

C. campestris

```r
# combine w/GBIF data
#citation: GBIF.org (31 August 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.y8rtg4
gbif.cc <- read.csv("Data/C.campestris_GBIF_JM.csv", header = T)
gbif.cc <- gbif.cc %>% select(Lat,Long)

cc.all <- rbind(cc, gbif.cc)
len.cc <- nrow(unique(cc.all))

write.table(cc.all, file="Data/Ccampestris.occ.txt", quote=F, row.names = F, col.names = T)

## Sample background
coordinates(cc.all) <- ~Long+Lat
projection(cc.all) <- CRS('+proj=longlat +datum=WGS84')

# circles with a radius of 50 km
x <- circles(cc.all, d=300000, lonlat=TRUE)
pol <- polygons(x)

# sample randomly from all circles
samp1 <- spsample(pol, 1000, type='random', iter=25)

# get unique cells
cells <- cellFromXY(env, samp1)
cells <- unique(cells)
xy <- xyFromCell(env, cells)

map(database="world", regions=c("Kenya","Tanzania","Ethiopia","Uganda","Rwanda","Burundi","Sudan","Sout
points(xy, cex=0.75, pch=20, col='blue')
points(cc.all$Long, cc.all$Lat, col="red", pch=20)
```
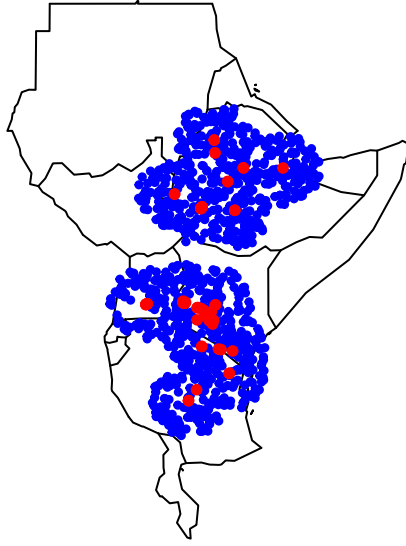
```
write.table(xy, file="Data/Ccampestris.bg.txt", quote=F, row.names = F, col.names = T)
```

Methods Summary

All occurrences for *C. reflexa* were from sampled localities ($n = 66$). Sampled localities were combined with occurrences from GBIF for *C. campestris* (total $n = 51$) and *C. kilimanjari* (total $n = 74$). Background points (n=1,000) were chosen from within a radius of 300 km of known occurrences.