

# Package: NicheBarcoding (via r-universe)

November 6, 2024

**Title** Niche-model-Based Species Identification

**Version** 1.8

**Description** Species Identification using DNA Barcodes Integrated with Environmental Niche Models.

**License** GPL (>= 3)

**Encoding** UTF-8

**Language** es

**LazyData** true

**LazyDataCompression** bzip2

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.1.2

**Imports** ape, dismo, e1071 (>= 1.7-7), maps, pROC, randomForest, raster, rJava, spider, vegan

**Depends** R (>= 2.10)

**VignetteBuilder** knitr

**Author** Cai-qing YANG [aut, cre], Ying WANG [aut], Xin-hai LI [aut], Jing LI [aut], Bing YANG [aut], Michael christopher ORR [aut], Ai-bing ZHANG [aut]

**Maintainer** Cai-qing YANG <yangcq@cnu.edu.cn>

**URL** <https://github.com/Yangcq-Ivy/NicheBarcoding>

**BugReports** <https://github.com/Yangcq-Ivy/NicheBarcoding>

**NeedsCompilation** no

**Suggests** rmarkdown, covr, testthat (>= 3.0.0), knitr

**Config/testthat/edition** 3

**Config/pak/sysreqs** libgdal-dev gdal-bin libgeos-dev make default-jdk libproj-dev libsqlite3-dev

**Repository** <https://yangcq-ivy.r-universe.dev>

**RemoteUrl** <https://github.com/yangcq-ivy/nichebarcoding>

**RemoteRef** HEAD

**RemoteSha** 40c0d6c43f9b632018501a59ece72f7ec8edd411

## Contents

bak.vir . . . . .	2
en.vir . . . . .	2
extractSpeInfo . . . . .	3
LappetMoths . . . . .	4
monophyly.prop . . . . .	4
NBSI . . . . .	5
NBSI2 . . . . .	7
niche.Model.Build . . . . .	9
niche.PCA . . . . .	11
pseudo.absent.points . . . . .	12
pseudo.present.points . . . . .	13
spe.mantel.test . . . . .	14
<b>Index</b>	<b>16</b>

---

bak.vir	<i>bak.vir data set, a class of matrix.</i>
---------	---

---

### Description

A dataset containing 5 of the 19 bioclimatic variables randomly generated as background points.

### Usage

bak.vir

### Format

a class of matrix.

**bak.vir** 5000\*5 matrix.

---

en.vir	<i>en.vir data set, a class of RasterBrick.</i>
--------	---

---

### Description

A dataset containing 5 of the 19 bioclimatic variables downloaded from WorldClim (version 1.4 with 10 arc minute resolution; Hijmans et al. 2005).

### Usage

en.vir

**Format**

a class of RasterBrick.

**en.vir** class: RasterBrick; dimensions : 360, 720, 259200, 5 (nrow, ncol, ncell, nlayers); resolution : 0.5, 0.5 (x, y); extent: -180, 180, -90, 90 (xmin, xmax, ymin, ymax); crs: +proj=longlat +datum=WGS84 +no\_defs; source: memory; names: bio1,bio4,bio7,bio12,bio15.

**Source**

<https://www.worldclim.org/>

---

extractSpeInfo

*Extraction of taxon/species and distribution information*

---

**Description**

Split comma-separated sample information into different columns of a data frame.

**Usage**

```
extractSpeInfo(seqID.full)
```

**Arguments**

seqID.full      Character, sample ID, taxon information and longitude and latitude data that splitted by comma in class character.

**Value**

A data frame of splitted sample ID, taxon information and longitude and latitude data for further analysis.

**Author(s)**

Cai-qing YANG (Email: yangcq(at)cnu.edu.cn) and Ai-bing ZHANG (Email:zhangab2008(at)cnu.edu.cn), Capital Normal University (CNU), Beijing, CHINA.

**Examples**

```
data(LappetMoths)
ref.seq<-LappetMoths$ref.seq
seqID.full<-rownames(ref.seq)

infor<-extractSpeInfo(seqID.full)
head(infor)
```

---

LappetMoths

*LappetMoths data set, a list of 8 data frames.*


---

### Description

A dataset containing the sequences IDs of species, coordinates of species sampled, and other attributes

### Usage

LappetMoths

### Format

list of 8 data frames.

**barcode.identi.result** data frame, species identifications by other methods or barcodes, containing query IDs, species identified, and corresponding probabilities.

**que.env** data frame, containing query sampleIDs, and a set of corresponding environmental variables collected by users.

**que.infor** data frame, query samples, containing sample IDs, longitude and latitude of each sample.

**que.seq** query sequences in binary format stored in a matrix

**ref.env** data frame, containing reference sampleIDs, species names, and a set of environmental variables collected by users.

**ref.infor** data frame, reference dataset containing sample IDs, taxon information, longitude and latitude of each sample.

**ref.seq** reference sequences in binary format stored in a matrix

**ref.add** data frame, additional reference dataset containing taxon information, longitude and latitude of each species.

---

monophyly.prop

*Calculate the proportion of monophyletic group on a tree*


---

### Description

Calculate the proportion of monophyletic group on a tree given species vector and a tree.

### Usage

```
monophyly.prop(phy, sppVector, singletonsMono = TRUE)
```

**Arguments**

phy                    A tree of class phylo.  
sppVector            Species vector.  
singletonsMono    Logical. Should singletons (i.e. only a single specimen representing that species) be treated as monophyletic? Default of TRUE. Possible values of FALSE and NA.

**Value**

A list containing proportion and number of monophyly group.  
A set monophyly and of non-monophyly group names.

**Author(s)**

Cai-qing YANG (Email: yangcq(at)cnu.edu.cn) and Ai-bing ZHANG (Email:zhangab2008(at)cnu.edu.cn), Capital Normal University (CNU), Beijing, CHINA.

**Examples**

```
library(ape)
tree<-ape::rtree(20)
tree$tip.label<-sample(tree$tip.label[1:10],size=20,replace = TRUE)
plot(tree)
sppVector<-tree$tip.label

MP<-monophyly.prop(tree,sppVector,singletonsMono = TRUE)
MP
```

---

NBSI

*Niche-model-Based Species Identification (NBSI)*

---

**Description**

Species identification using DNA barcoding integrated with niche model.

**Usage**

```
NBSI(
  ref.seq,
  que.seq,
  model = "MAXENT",
  ref.add = NULL,
  en.vir = NULL,
  bak.vir = NULL
)
```



```

#                               lonlatCorrection=TRUE)
#bak.vir<-raster::extract(en.vir,back)

library(ape)
data(LappetMoths)
ref.seq<-LappetMoths$ref.seq[1:50,]
que.seq<-LappetMoths$que.seq[1:5,]
NBSI.out<-NBSI(ref.seq,que.seq,ref.add=NULL,
               model="RF",en.vir=en.vir,bak.vir=bak.vir)
NBSI.out

### Add a parameter when additional reference coordinates are available ###
#ref.add<-LappetMoths$ref.add
#NBSI.out2<-NBSI(ref.seq,que.seq,ref.add=ref.add,
#               model="RF",en.vir=en.vir,bak.vir=bak.vir)
#NBSI.out2

```

---

NBSI2

*Niche-model-Based Species Identification (NBSI) for a prior analysis*


---

## Description

If users already have species identified by other barcodes or methods, they can use this function given the identified species names and corresponding probabilities to make further confirm by environmental niche model.

## Usage

```

NBSI2(
  ref.infor = NULL,
  que.infor = NULL,
  ref.env = NULL,
  que.env = NULL,
  barcode.identi.result,
  model = "MAXENT",
  en.vir = NULL,
  bak.vir = NULL
)

```

## Arguments

ref.infor	Data frame, reference dataset containing sample IDs, taxon information, longitude and latitude of each sample.
que.infor	Data frame, query samples, containing sample IDs, longitude and latitude of each sample.
ref.env	Data frame, containing reference sample IDs, species names, and a set of environmental variables collected by users.

<code>que.env</code>	Data frame, containing query sample IDs, and a set of corresponding environmental variables collected by users.
<code>barcode.identi.result</code>	Data frame, species identifications by other methods or barcodes, containing query IDs, species identified, and corresponding probabilities.
<code>model</code>	Character, string indicating which niche model will be used. Must be one of "MAXENT" (default) or "RF". "MAXENT" can only be applied when the java program <code>paste(system.file(package="dismo"), "/java/maxent.jar", sep="")</code> exists.
<code>en.vir</code>	RasterBrick, the global bioclimate data output from "raster::getData" function.
<code>bak.vir</code>	Matrix, bioclimate variables of random background points.

### Value

A dataframe of identifications for query samples and their niche-based reliability.

### Author(s)

Cai-qing YANG (Email: yangcq(at)cnu.edu.cn) and Ai-bing ZHANG (Email: zhangab2008(at)cnu.edu.cn), Capital Normal University (CNU), Beijing, CHINA.

### References

- Breiman, L. 2001. Random forests. *Machine Learning* 45(1):5-32.
- Liaw, A. and M. Wiener. 2002. Classification and regression by randomForest. *R News*, 2/3:18-22.
- Phillips, S.J., R.P. Anderson and R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190:231-259.
- Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15):1965-1978.

### Examples

```
data(en.vir)
data(bak.vir)
#envir<-raster::getData("worldclim",download=FALSE,var="bio",res=2.5)
#en.vir<-raster::brick(envir)
#back<-dismo::randomPoints(mask=en.vir,n=5000,ext=NULL,extf=1.1,
#                           excludep=TRUE,prob=FALSE,
#                           cellnumbers=FALSE,tryf=3,warn=2,
#                           lonlatCorrection=TRUE)
#bak.vir<-raster::extract(en.vir,back)

data(LappetMoths)
barcode.identi.result<-LappetMoths$barcode.identi.result
ref.infor<-LappetMoths$ref.infor
que.infor<-LappetMoths$que.infor

if(class(en.vir) == "NULL"){
  NBSI2.out<-NBSI2(ref.infor=ref.infor,que.infor=que.infor,
```



```

        barcode.identi.result=barcode.identi.result,
        model="MAXENT",en.vir=NULL,bak.vir=NULL)
}else{
  NBSI2.out<-NBSI2(ref.infor=ref.infor,que.infor=que.infor,
        barcode.identi.result=barcode.identi.result,
        model="maxent",en.vir=en.vir,bak.vir=bak.vir)
}
NBSI2.out

ref.env<-LappetMoths$ref.env
que.env<-LappetMoths$que.env

NBSI2.out2<-NBSI2(ref.env=ref.env,que.env=que.env,
        barcode.identi.result=barcode.identi.result,
        model="MAXENT",en.vir=en.vir,bak.vir=bak.vir)
NBSI2.out2

```

---

niche.Model.Build

*Ecological niche model building using the randomForest classifier*


---

### Description

Build a niche model for a given species according to its distribution data.

### Usage

```

niche.Model.Build(
  prese = NULL,
  absen = NULL,
  prese.env = NULL,
  absen.env = NULL,
  model = "MAXENT",
  en.vir = NULL,
  bak.vir = NULL
)

```

### Arguments

prese	Data frame, longitude and latitude of the present data of a species (can be absent when providing prese.env parameter).
absen	Data frame, longitude and latitude of the absent data of a species.(can be absent when providing absen.env or back parameter).
prese.env	Data frame, bioclimate variables of present data. (can be absent when providing prese parameter).
absen.env	Data frame, bioclimate variables of absent data. (can be absent when providing absen or back parameter).

model	Character, string indicating which niche model will be used. Must be one of "MAXENT" (default) or "RF". "MAXENT" can only be applied when the java program <code>paste(system.file(package="dismo"), "/java/maxent.jar", sep="")</code> exists.
en.vir	RasterBrick, the global bioclimate data output from "raster::getData" function.
bak.vir	Matrix, bioclimate variables of random background points.

**Value**

randomForest/MaxEnt, a trained niche model object.

A vector including the specificity, sensitivity and threshold of the trained model.

**Author(s)**

Cai-qing YANG (Email: yangcq(at)cnu.edu.cn) and Ai-bing ZHANG (Email:zhangab2008(at)cnu.edu.cn), Capital Normal University (CNU), Beijing, CHINA.

**References**

Breiman, L. 2001. Random forests. *Machine Learning* 45(1):5-32.

Liaw, A. and M. Wiener. 2002. Classification and regression by randomForest. *R News*, 2/3:18-22.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15):1965-1978.

**Examples**

```
data(en.vir)
data(bak.vir)
#envir<-raster::getData("worldclim",download=FALSE,var="bio",res=2.5)
#en.vir<-raster::brick(envir)
#back<-dismo::randomPoints(mask=en.vir,n=5000,ext=NULL,extf=1.1,
#                           excludep=TRUE,prob=FALSE,
#                           cellnumbers=FALSE,tryf=3,warn=2,
#                           lonlatCorrection=TRUE)
#bak.vir<-raster::extract(en.vir,back)
```

```
data<-data.frame(species=rep("Acosmeryx anceus",3),
                 Lon=c(145.380,145.270,135.461),
                 Lat=c(-16.4800,-5.2500,-16.0810))
present.points<-pseudo.present.points(data,10,2,1,en.vir)
NMB.out<-niche.Model.Build(prese=present.points,absen=NULL,
                           prese.env=NULL,absen.env=NULL,
                           model="MAXENT",
                           en.vir=en.vir,bak.vir=bak.vir)
```

NMB.out

```
prese.env<-raster::extract(en.vir,present.points[,2:3])
prese.env<-as.data.frame(prese.env)
NMB.out2<-niche.Model.Build(prese=NULL,absen=NULL,
```

```

prese.env=prese.env,absen.env=NULL,
model="MAXENT",
en.vir=en.vir,bak.vir=bak.vir)
NMB.out2

```

---

niche.PCA	<i>Principal component analysis of ecological niche among unknown species and the potential species to which they may belong</i>
-----------	--

---

## Description

Determine whether unknown species belong to a known species through principal component analysis of ecological niche according to their distribution information.

## Usage

```
niche.PCA(ref.lonlat, que.lonlat, en.vir = NULL)
```

## Arguments

ref.lonlat	Data frame, longitude and latitude of the known species.
que.lonlat	Data frame, longitude and latitude of unknown species.
en.vir	RasterBrick, the globe bioclimate data obtained from "raster::getData" function.

## Value

A list containing importance and loadings of the components.

A dataframe of points that within the 95% confidence interval of the reference dataset ecological space.

A figure shows whether the query points (blue solid circles) are located in the 95%CI range of the niche space of reference species.

## Author(s)

Cai-qing YANG (Email: yangcq(at)cnu.edu.cn) and Ai-bing ZHANG (Email:zhangab2008(at)cnu.edu.cn), Capital Normal University (CNU), Beijing, CHINA.

## Examples

```

data(en.vir)
#envir<-raster::getData("worldclim",download=FALSE,var="bio",res=2.5)
#en.vir<-raster::brick(envir)

```

```

data(LappetMoths)
ref.infor<-LappetMoths$ref.infor
que.infor<-LappetMoths$que.infor

```

```

#windows() # open a new plot window when the image format is abnormal
nPCA<-niche.PCA(ref.lonlat=ref.infor[,3:5],
               que.lonlat=que.infor[,c(2,4:5)],
               en.vir=en.vir)

nPCA$summary
nPCA$que.CI

data<-data.frame(species=rep("Acosmeryx anceus",3),
                 Lon=c(145.380,145.270,135.461),
                 Lat=c(-16.4800,-5.2500,-16.0810))
simuSites<-pseudo.present.points(data,500,4,2,en.vir)
ref.lonlat<-simuSites[1:480,]
que.lonlat<-simuSites[481:500,]

#windows() # open a new plot window when the image format is abnormal
nPCA2<-niche.PCA(ref.lonlat,que.lonlat,en.vir=en.vir)
nPCA2$summary
nPCA2$que.CI

```

---

pseudo.absent.points    *Generation of pseudo absent points for niche model building*

---

## Description

Randomly generate pseudo points outside the 95%CI of the ecological space of the present data when there is no absent data for building a niche model.

## Usage

```
pseudo.absent.points(data, outputNum = 500, en.vir = NULL, map = TRUE)
```

## Arguments

data	Data frame, longitude and latitude of a single species.
outputNum	Numeric, the expected number of points.
en.vir	RasterBrick, the globe bioclimate data obtained from "raster::getData" function.
map	Logical. Should a map be drawn?

## Value

A data frame of simulated pseudo points.

A data frame of bioclimate variables of each pseudo points.

**Author(s)**

Cai-qing YANG (Email: yangcq(at)cnu.edu.cn) and Ai-bing ZHANG (Email:zhangab2008(at)cnu.edu.cn), Capital Normal University (CNU), Beijing, CHINA.

**Examples**

```
data(en.vir)
#envir<-raster::getData("worldclim",download=FALSE,var="bio",res=2.5)
#en.vir<-raster::brick(envir)

data<-data.frame(species=rep("Acosmeryx anceus",3),
                 Lon=c(145.380,145.270,135.461),
                 Lat=c(-16.4800,-5.2500,-16.0810))

absent.points<-pseudo.absent.points(data,en.vir=en.vir,outputNum=100)
head(absent.points$lonlat)
head(absent.points$envir)
```

---

pseudo.present.points *Generation of pseudo present points for niche model building*

---

**Description**

Randomly generate pseudo points around actual present distribution site when the number of present points is inadequate for building a niche model.

**Usage**

```
pseudo.present.points(
  data,
  outputNum = 50,
  lonRange = 2,
  latRange = 1,
  en.vir = NULL,
  map = TRUE
)
```

**Arguments**

data	Data frame, longitude and latitude of a single species.
outputNum	Numeric, the expected number of points.
lonRange	Range of the longitude of the points generated.
latRange	Range of the latitude of the points generated.
en.vir	RasterBrick, the globe bioclimate data obtained from "raster::getData" function.
map	Logical. Should a map be drawn?

**Value**

A data frame, containing actual present points and simulated pseudo points.

**Author(s)**

Cai-qing YANG (Email: yangcq(at)cnu.edu.cn) and Ai-bing ZHANG (Email:zhangab2008(at)cnu.edu.cn), Capital Normal University (CNU), Beijing, CHINA.

**Examples**

```
data(en.vir)
#envir<-raster::getData("worldclim",download=FALSE,var="bio",res=2.5)
#en.vir<-raster::brick(envir)

data<-data.frame(species=rep("Acosmeryx anceus",3),
                  Lon=c(145.380,145.270,135.461),
                  Lat=c(-16.4800,-5.2500,-16.0810))

present.points<-pseudo.present.points(data,10,2,1,en.vir=en.vir)
present.points
```

---

spe.mantel.test	<i>Mantel test between interspecific pairwise genetic distance and ecological distance</i>
-----------------	--

---

**Description**

Determine the independence between genetic distance and ecological distance for a reference dataset at the level of species.

**Usage**

```
spe.mantel.test(
  fas,
  dna.model = "raw",
  ecol.dist.method = "euclidean",
  mantel.method = "spearman",
  permutations = 999,
  en.vir = NULL
)
```

**Arguments**

fas DNABin, reference dataset containing sample IDs, taxon information, longitude and latitude, and barcode sequences of samples.

dna.model	Character, specifying the evolutionary model to be used; must be one of "raw" (default), "N", "TS", "TV", "JC69", "K80", "F81", "K81", "F84", "BH87", "T92", "TN93", "GG95", "logdet", "paralin", "indel", or "indelblock".
ecol.dist.method	Character, distance measure to be used; must be one of "euclidean" (default), "maximum", "manhattan", "canberra", "binary" or "minkowski".
mantel.method	Character, correlation method, as accepted by cor: "pearson", "spearman" (default) or "kendall".
permutations	Numeric, the number of permutations required.
en.vir	RasterBrick, the global bioclimate data output from "raster::getData" function.

**Value**

The Mantel statistic.  
 The empirical significance level from permutations.  
 A matrix of interspecific pairwise genetic distance.  
 A matrix of interspecific pairwise ecological distance.

**Author(s)**

Cai-qing YANG (Email: yangcq(at)cnu.edu.cn) and Ai-bing ZHANG (Email: zhangab2008(at)cnu.edu.cn), Capital Normal University (CNU), Beijing, CHINA.

**References**

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Can. Res.* 27:209-220.

Oksanen J., F.G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlenn, P.R. Minchin, R.B. O'Hara, G.L. Simpson, P. Solymos, M.H.H. Stevens, E. Szoecs and H Wagner. 2016. *vegan: Community Ecology Package* <https://CRAN.R-project.org/package=vegan>. r package version 2.5-6.

**Examples**

```
data(en.vir)
#envir<-raster::getData("worldclim",download=FALSE,var="bio",res=2.5)
#en.vir<-raster::brick(envir)

library(ape)
data(LappetMoths)
ref.seq<-LappetMoths$ref.seq

spe.mantel<-spe.mantel.test(fas=ref.seq,en.vir=en.vir)
spe.mantel$MantelStat.r
spe.mantel$p.value
```

# Index

- \* **NBSI2**
    - NBSI2, [7](#)
  - \* **NBSI**
    - NBSI, [5](#)
  - \* **datasets**
    - bak.vir, [2](#)
    - en.vir, [2](#)
    - LappetMoths, [4](#)
  - \* **extractSpeInfo**
    - extractSpeInfo, [3](#)
  - \* **monophyly.prop**
    - monophyly.prop, [4](#)
  - \* **niche.Model.Build**
    - niche.Model.Build, [9](#)
  - \* **niche.PCA**
    - niche.PCA, [11](#)
  - \* **pseudo.absent.points**
    - pseudo.absent.points, [12](#)
  - \* **pseudo.present.points**
    - pseudo.present.points, [13](#)
  - \* **spe.mantel.test**
    - spe.mantel.test, [14](#)
- bak.vir, [2](#)
- en.vir, [2](#)
- extractSpeInfo, [3](#)
- LappetMoths, [4](#)
- monophyly.prop, [4](#)
- NBSI, [5](#)
- NBSI2, [7](#)
- niche.Model.Build, [9](#)
- niche.PCA, [11](#)
- pseudo.absent.points, [12](#)
- pseudo.present.points, [13](#)
- spe.mantel.test, [14](#)