BarcodingR: an integrated R package for species identification using DNA barcodes A Tutorial

Ai-bing ZHANG
Zhangab2008@mail.cnu.edu.cn
Meng-di HAO
dreamdy521@163.com
Cai-qing YANG
ivyyang0416@gmail.com
Zhi-yong SHI
zhiyong-shi@163.com

Introduction

BarcodingR: is an integrated R package for species identification using DNA barcodes. BarcodingR includes implementations of BP-based (Zhang et al 2008,Zhang et al 2009a, fuzzy-set based (Zhang et al 2012b), and Bayesian species identification (Jin et al 2013), and new/updated functions variously useful in barcode study such as barcode evaluation, kmer optimization, DNA barcode gap analysis, species membership analysis, delimitation comparison analysis, consensus identification. The 'BarcodingR' package is written in R language and is released on Github (https://github.com/zhangab2008/BarcodingR). It relies on the R packages 'ape' (Paradis et al 2004),'adegenet' (Jombart et al 2008),'mclust' (Fraley et al 1999),'phyclust' (Chen et al 2010),'pegas' (Paradis et al 2010), 'seqinr' (Charif et al 2005),'nnet' (Ripley et al 1996),'e1071'(https://cran.r-project.org/),'spider' (Brown et al 2012).

This document is a tutorial on the use of the functions implemented in BarcodingR, and offers a guide to the interpretation of its output. This tutorial does not discuss the basic R syntax and data structures in detail, and the reader is referred to other manuals (https://www.rstudio.com/online-learning/#R) for this information.

Install

After installing all dependencies R packages, installing BarcodingR package could be done from the R command line: "install.packages("BarcodingR.zip",repos=NULL)" for windows or "install.packages("BarcodingR.tar.gz",repos=NULL)" for linux system. In the RStudio environment, installing could be conducted through the following menu: Tools− >Install Packages− >Install from (Package Archive File(.zip; .tar.gz))− >Install.

Package 'BarcodingR'

July 14, 2016

Title an integrated R package for species identification using DNA barcodes

Version 1.0-1

Description To perform species identification using DNA barcodes.

Depends R (>= 3.2.1),ape,adegenet,cluster,mclust,phyclust,pegas,compiler,seqinr,cclust,nnet,class,e1071,spider,sp,stats,utils

License GPL-2

LazyData true

ByteCompile true

RoxygenNote 5.0.1

NeedsCompilation no

Author Ai-bing ZHANG [aut, cre],
Meng-di HAO [aut],
Cai-qing YANG [aut],
Zhi-yong SHI [aut]

Maintainer Ai-bing ZHANG <Zhangab2008@mail.cnu.edu.cn>

R topics documented:

`barcodes.eval` *Barcodes Evaluation*

Description

Evaluate two barcodes using species identification success rate critera.

Usage

```
barcodes.eval(barcode1, barcode2, kmer1 = kmer1, kmer2 = kmer2)
```

Arguments

`barcode1`	object of class "DNAbin" based on barcode1, which contains taxon information.
`barcode2`	object of class "DNAbin" based on barcode2, which contains taxon information.
`kmer1`	a numeric to indicate the length of kmer1 for barcode1, the opitimal kmer could be found by the function optimize.kmer() before running this function.
`kmer2`	a numeric to indicate the length of kmer2 for barcode2, see above.

Value

a list containing p_value of prop.test(), and so on.

Author(s)

Ai-bing ZHANG, PhD. CNU, Beijing, CHINA.

References

zhangab2008(at)mail.cnu.edu.cn

See Also

prop.test()

Examples

```
data(TibetanMoth)
barcode1<-as.DNAbin(as.character(TibetanMoth[1:30,]))
barcode2<-barcode1
b.eval<-barcodes.eval(barcode1,barcode2,kmer1=1,kmer2=3)
b.eval
```

barcoding.gap *Barcoding Gap Calculation*

Description

Calculation of DNA barcoding gap. Besides K2P distance, raw distance and euclidean could also be used for calculation DNA barcoding gap.

Usage

```
barcoding.gap(ref, dist = dist)
```

Arguments

ref	object of class "DNAbin" used as a reference dataset, which contains taxon information.
dist	a character string which takes one of ("raw","K80","euclidean").

Value

a list indicates the summary statistics of interspecific and intraspecific genetic distance, such as k2P distance.

Note

the current version of the function can only be used for protein-coding barcodes, such as, COI. The futuren version may incorporate calculation for non-coding barcodes,for instance, ITS1, ITS2.

Author(s)

Ai-bing ZHANG, PhD. CNU, Beijing, CHINA, contact at zhangab2008(at)mail.cnu.edu.cn

References

Meyer, Christopher P., and Gustav Paulay. (2005). "DNA barcoding: error rates based on comprehensive sampling.".PLoS biology 3.12: e422.

F. Jiang, Q. Jin, L. Liang, A.B. Zhang,and Z.H. Li.(2014). Existence of Species Complex Largely Reduced Barcoding Success for Invasive Species of Tephritidae: A Case Study in Bactrocera spp. Mol Ecol Resour. 14(6):1114-1128 DOI: 10.1111/1755-0998.12259.

Examples

```
data(TibetanMoth)
b.gap<-barcoding.gap(ref=TibetanMoth,dist="K80")
b.gap
```

barcoding.spe.identify

Species Identification using Protein-coding Barcodes

Description

Species identification using protein-coding barcodes with different methods,including BP-based method (Zhang et al. 2008), fuzzy-set based method (Zhang et al. 2012), Bayesian-based method (Jin et al. 2013).

Usage

```
barcoding.spe.identify(ref, que, method = "bpNewTraining")
```

Arguments

ref	object of class "DNAbin" used as a reference dataset, which contains taxon information.
que	object of class "DNAbin", whose identities (species names) need to be inferred.
method	a character string indicating which method will be used to train model and/or infer species membership. One of these methods ("fuzzyId", "bpNewTraining","bpNewTrainingOnly","bpUseTrained","Bayesian") should be specified.

Value

a list containing model parameters used, species identification success rates using references, query sequences, species inferred, and corresponding confidence levels (bp probability for BP-based method / FMF values for fuzzy set theory based method) when available.

Note

functions fasta2DNAbin() from package:adegenet and read.dna() from package:ape were used to obtain DNAbin object in our package. The former is used to read large aligned coding DNA barcodes, the latter unaligned ones. ref and que should be aligned with identical sequence length. We provided a pipeline to perform fast sequences alignment for reference and query sequences. Windows users could contact zhangab2008(at)mail.cnu.edu.cn for an exec version of the package. For very large DNA dataset, read.fas() package:phyloch is strongly suggested instead of fasta2DNAbin() since the latter is very slow.

Author(s)

Ai-bing ZHANG, PhD. CNU, Beijing, CHINA, contact at zhangab2008(at)mail.cnu.edu.cn

References

Q. Jin, H.L. Han, X.M. Hu, X.H. Li,C.D. Zhu,S. Y. W. Ho, R. D. Ward, A.B. Zhang . (2013).Quantifying Species Diversity with a DNA Barcoding-Based Method: Tibetan Moth Species (Noctuidae) on the Qinghai-Tibetan Plateau. PloS One 8: e644.

Zhang, A. B., C. Muster, H.B. Liang, C.D. Zhu, R. Crozier, P. Wan, J. Feng, R. D. Ward.(2012). A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding. Molecular Ecology, 21(8):1848-63.

Zhang, A. B., D. S. Sikes, C. Muster, S. Q. Li. (2008). Inferring Species Membership using DNA sequences with Back-propagation Neural Networks. Systematic Biology, 57(2):202-215.

Examples

```
data(TibetanMoth)
ref<-as.DNAbin(as.character(TibetanMoth[1:50,]))
que<-as.DNAbin(as.character(TibetanMoth[50:60,]))
bsi<-barcoding.spe.identify(ref, que, method = "fuzzyId")
bsi
bsi<-barcoding.spe.identify(ref, que, method = "bpNewTraining")
bsi
bsi<-barcoding.spe.identify(ref, que, method = "bpUseTrained")
bsi
bsi<-barcoding.spe.identify(ref, que, method = "Bayesian")
bsi
```

---

barcoding.spe.identify2

Species Identification Based on Fuzzy-set Method and kmer

Description

Species identification based on fuzzy-set method (Zhang et al. 2012)and kmer.

Usage

```
barcoding.spe.identify2(ref, que, kmer = kmer, optimization = TRUE)
```

Arguments

ref	object of class "DNAbin" used as a reference dataset, which contains taxon information.
que	object of class "DNAbin", whose identities (species names) need to be inferred.
kmer	a numeric to indicate the length of maximum kmer to try in the range of 1 to kmer in case of optimization = TRUE, otherwise, only a certain length of kmer is used.
optimization	a character string, indicating whether different length of kmer (up to kmer) will be used or just a specified length of kmer will be used.

Value

a list indicating the identified species.

Note

read.dna() from package ape was used to obtain DNAbin object for unaligned non-coding barcodes.

Author(s)

Ai-bing ZHANG, Cai-qing YANG, Meng-di HAO, CNU, Beijing, CHINA, contact at zhangab2008(at)mail.cnu.edu.cn

References

Q. Jin, H.L. Han, X.M. Hu, X.H. Li,C.D. Zhu,S. Y. W. Ho, R. D. Ward, A.B. Zhang . (2013).Quantifying Species Diversity with a DNA Barcoding-Based Method: Tibetan Moth Species (Noctuidae) on the Qinghai-Tibetan Plateau. PloS One 8: e644.

Zhang, A. B., C. Muster, H.B. Liang, C.D. Zhu, R. Crozier, P. Wan, J. Feng, R. D. Ward.(2012). A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding. Molecular Ecology, 21(8):1848-63.

Zhang, A. B., D. S. Sikes, C. Muster, S. Q. Li. (2008). Inferring Species Membership using DNA sequences with Back-propagation Neural Networks. Systematic Biology, 57(2):202-215.

Examples

```
data(pineMothITS2)
ref<-pineMothITS2
que<-ref
spe.id<-barcoding.spe.identify2(ref,que, kmer = 10, optimization = TRUE)
spe.id
spe.id<-barcoding.spe.identify2(ref,que, kmer = 2, optimization = FALSE)
spe.id
```

---

bbsik *Bp Barcoding Species Identify using Kmer*

Description

Species identification using BP-based method for both protein-coding barcodes, for instance, COI, and non-coding barcodes, such as, ITS, using kmer statistics.

Usage

```
bbsik(ref, que, kmer = kmer, UseBuiltModel = FALSE, lr = 5e-05,
  maxit = 1e+06)
```

Arguments

ref	object of class "DNAbin" used as a reference dataset, which contains taxon information.
que	object of class "DNAbin", which needs to be inferred.
kmer	a numeric indicating the length of kmer used.
UseBuiltModel	logic value to indicate whether a built model is used or not.
lr	parameter for weight decay. Default 5e-5.
maxit	maximum number of iterations. Default 1e+6.

Value

a list containing model parameters used, species identification success rates using references, query sequences, species inferred, and corresponding confidence levels (bp probability for BP-based method).

Author(s)

Ai-bing ZHANG, Meng-di HAO, Cai-qing YANG, CNU, Beijing, CHINA. zhangab2008(at)mail.cnu.edu.cn

References

Zhang, A. B., D. S. Sikes, C. Muster, S. Q. Li. (2008). Inferring Species Membership using DNA sequences with Back-propagation Neural Networks. Systematic Biology, 57(2):202-215.

Examples

```
data(TibetanMoth)
ref<-as.DNAbin(as.character(TibetanMoth[1:50,]))
que<-as.DNAbin(as.character(TibetanMoth[51:60,]))
out<-bbsik(ref, que, kmer = 1, UseBuiltModel = FALSE)
out
out$convergence
out$success.rates.ref

data(pineMothITS2)
ref<-pineMothITS2
que<-pineMothITS2
out<-bbsik(ref, que, kmer = 1, UseBuiltModel = FALSE)
out
out$convergence
out$success.rates.ref
```

---

char2NumVector *Character to Integer Vector*

Description

Conversion from a character vector to an integer vector.

Usage

```
char2NumVector(c)
```

Arguments

c	character vector.

Value

an integer vector.

Author(s)

Ai-bing ZHANG, PhD. CNU, Beijing, CHINA. zhangab2008(at)mail.cnu.edu.cn

References

zhangab2008(at)mail.cnu.edu.cn

Examples

```
c<-c("a","a","b")
num<-char2NumVector(c)
num
```

`compare2delimitations` *Comparision between two Delimitations*

Description

Comparision between two delimitations of a group of samples, for instance, traditionally morphological delimitation and molecular delimitation (MOTU).

Usage

```
compare2delimitations(deli1, deli2)
```

Arguments

`deli1`	a character array (vector),containing a set of, for example, morphological identification (species names), to compare with
`deli2`	a character array (vector),containing a set of, molecular delimitation (MOTU).

Value

a list containing the adjusted Rand index comparing the two partitions (a scalar). This index has zero expected value in the case of random partition, and it is bounded above by 1 in the case of perfect agreement between two partitions; the numbers of matches, splits,merges, and corresponding percentage.

Note

This is for the same set of samples with two partitions/delimitations.

Author(s)

Ai-bing ZHANG, PhD. CNU, Beijing, CHINA.

References

L. Hubert and P. Arabie (1985) Comparing Partitions, Journal of the Classification 2:193-218.

Examples

```
deli1<-c(1,1,1,1,1,1)
deli2<-c(1,1,2,1,1,3)
out<-compare2delimitations(deli1,deli2)
out
```

`consensus.identify` *Consensus Identification*

Description

Make consensus for identifications from two or more methods, usually for a set of query sequences.

Usage

```
consensus.identify(identifiedBy2orMore)
```

Arguments

`identifiedBy2orMore`
: an object of class "data.frame", containing (queIDs, as rownames), identifiedByMethod1,identifiedByMethod2,and so on.

Value

a data frame with concensus.identification, and corresponding votes.

Note

Suitable for case where a set of queries were identified by more than two methods.

Author(s)

Ai-bing ZHANG, PhD. CNU, Beijing, CHINA, contact at zhangab2008(at)mail.cnu.edu.cn

Examples

```
queIDs<-c("q1","q2","q3")

bp<-c("sp1","sp1","sp1")
bpk<-c("sp1","sp1","sp2")
bayes<-c("sp2","sp1","sp3")
fuzzyID<-c("sp1","sp1","sp2")
identifiedBy2orMore<-data.frame(bp=bp,bpk=bpk,bayes=bayes,fuzzyID=fuzzyID)
rownames(identifiedBy2orMore)<-queIDs<-c("q1","q2","q3")
ccs<-consensus.identify(identifiedBy2orMore)
```

`digitize.DNA` *Digitize DNAbin*

Description

Digitize an object of DNAbin.

Usage

```
digitize.DNA(seqs)
```

Arguments

`seqs`	an object of DNAbin.

Value

a numeric matrix of DNA sequences digitized.

Author(s)

Ai-bing ZHANG, PhD. CNU, Beijing, CHINA.

References

zhangab2008(at)mail.cnu.edu.cn

Examples

```
data(woodmouse)
digitized.DNA<-digitize.DNA(seqs=woodmouse)
digitized.DNA
```

`DNAbin2kmerFreqMatrix` *Calculation of Kmer Frequency Matrix from DNAbin for Both Reference and Query Sequences*

Description

Calculation of kmer frequency matricies from DNAbin for both reference and query sequences.

Usage

```
DNAbin2kmerFreqMatrix(ref, que, kmer = kmer)
```

Arguments

ref	Object of class "DNAbin" used as a reference dataset, which contains taxon information.
que	Object of class "DNAbin", whichs need to be inferred.
kmer	a numeric to indicate the length of kmer used.

Value

kmer frequency matricies for both ref and que sequences, but only based on kmers found in ref!!! new kmers in que will be ignored.

Author(s)

Ai-bing ZHANG, PhD. CNU, Beijing, CHINA. zhangab2008(at)mail.cnu.edu.cn

References

zhangab2008(at)mail.cnu.edu.cn

Examples

```
data(TibetanMoth)
ref<-as.DNAbin(as.character(TibetanMoth[1:50,]))
que<-as.DNAbin(as.character(TibetanMoth[51:60,]))
out<-DNAbin2kmerFreqMatrix(ref,que,kmer=3)
out
```

FMF *Fuzzy Membership Function Value*

Description

Calculation fuzzy membership function value given a distance from query to a potenial species, maximual intraspecific variation of the potential species theta1, and minimal interspecific distance (here, the distance between the potential species and its nearest neighbor theta2) (fuzzy-set based method, Zhang et al. 2012), different definition of distances could also be used.

Usage

```
FMF(xtheta12)
```

Arguments

xtheta12	a numerical vector containing three elements, a distance from query to a potenial species, maximual or sd of intraspecific variation of the potential species theta1,minimal or mean interspecific distance.

Value

a numeric between 0 and 1.

Note

different definitions of distances could also be used.

Author(s)

Ai-bing ZHANG, Zhi-yong SHI. CNU, Beijing, CHINA, contact at zhangab2008(at)mail.cnu.edu.cn

References

Zhang, A. B., C. Muster, H.B. Liang, C.D. Zhu, R. Crozier, P. Wan, J. Feng, R. D. Ward.(2012). A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding. Molecular Ecology, 21(8):1848-63.

Examples

```
xtheta12<-c(0.6289163,0.1465522,0.6379375)
FMF.out<-FMF(xtheta12)
FMF.out
```

FMFtheta12 *Calculate Intraspecific and Interspecific Variation*

Description

Calculation intraspecific variation (sd) of the potential species theta1, and mean interspecific distance (here, the mean distance between the potential species and its nearest neighbor theta2) (fuzzy-set based method,slightly modified from Zhang et al. 2012). The calculation was done for all species in the reference dataset.

Usage

```
FMFtheta12(ref)
```

Arguments

ref	object of class "DNAbin" used as a reference dataset, which contains taxon information.

Value

a data frame containing intraspecific (sd, theta1) and interspefic variation (mean) of all species, and their corresponding nearest neighbor (NN).

Author(s)

Ai-bing ZHANG, PhD. CNU, Beijing, CHINA, contact at zhangab2008(at)mail.cnu.edu.cn

References

Zhang, A. B., C. Muster, H.B. Liang, C.D. Zhu, R. Crozier, P. Wan, J. Feng, R. D. Ward.(2012). A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding. Molecular Ecology, 21(8):1848-63.

Examples

```
data(TibetanMoth)
ref<-as.DNAbin(as.character(TibetanMoth[1:50,]))
FMF.theta12<-FMFtheta12(ref)
FMF.theta12
```

NAMES *Extracts Lables of Samples*

Description

Extract sequence names from different objects of DNAbin, including generated from fasta2DNAbin() (package:adegenet), and read.dna() (package:ape).

Usage

```
NAMES(seqs)
```

Arguments

seqs	object of class "DNAbin", generated from fasta2DNAbin() (package:adegenet), and read.dna() (package:ape).

Value

a character string array/vector.

Author(s)

Ai-bing ZHANG, PhD. CNU, Beijing, CHINA.

References

zhangab2008(at)mail.cnu.edu.cn

Examples

```
data(TibetanMoth)
seqNames<-NAMES(TibetanMoth)
seqNames
```

optimize.kmer *Optimize kmer Length*

Description

Optimize kmer length by trying kmers which length is in the range from 1 to max.kmer. The optimal kmer will have maximumal species identification success rate.

Usage

```
optimize.kmer(ref, max.kmer = max.kmer)
```

Arguments

ref	object of class "DNAbin" used as a reference dataset, which contains taxon information.
max.kmer	a numeric to indicate the length of maximal kmer.

Value

a numeric indicating the optimal kmer in the range examined.

Author(s)

Ai-bing ZHANG, Cai-qing YANG, Meng-di HAO, CNU, Beijing, CHINA.

References

zhangab2008(at)mail.cnu.edu.cn/zhangab2008(at)gmail.com.

Examples

```
data(TibetanMoth)
ref<-TibetanMoth[1:10,]
optimial.kmer<-optimize.kmer(ref,max.kmer=5)
```

pineMothCOI *pine Moth COI*

Description

COI DNA barcodes of Pine Moth in China.

Usage

```
data("pineMothCOI")
```

Format

The format is: 'DNAbin' raw [1:140, 1:652] a t a a ... - attr(*, "dimnames")=List of 2 ..$: chr [1:140] "A111,Lasiocampidae_Dendrolimus_punctatus" "A22,Lasiocampidae_Dendrolimus_punctatus" "A23,Lasiocampidae_Dendrolimus_punctatus" "A27,Lasiocampidae_Dendrolimus_punctatus" ... ..$: NULL

Details

COI DNA barcodes of Pine Moth (Lasiocampidae) in China.

Source

http://dx.plos.org/10.1371/journal.pone.0064428.

References

Dai Q-Y, Gao Q, Wu C-S, Chesters D, Zhu C-D, and A.B. Zhang*. (2012) Phylogenetic Reconstruction and DNA Barcoding for Closely Related Pine Moth Species (Dendrolimus) in China with Multiple Gene Markers. PLoS ONE 7(4): e32544.

Examples

```
data(pineMothCOI)
pineMothCOI
```

---

pineMothITS1 *pine Moth ITS1*

Description

ITS1 sequences of seven closely related pine moths species sampled through China.

Usage

```
data("pineMothITS1")
```

Format

The format is: List of 69 $ A43,Lasiocampidae_Dendrolimus_punctatus: raw [1:698] 48 48 48 28 ... $ A54,Lasiocampidae_Dendrolimus_punctatus: raw [1:696] 48 18 28 18 ... - attr(*, "class")= chr "DNAbin"

Details

ITS1 used as DNA barcodes for closely related rine moth species (Dendrolimus) in China.

Source

http://dx.plos.org/10.1371/journal.pone.0064428.

References

Dai Q-Y, Gao Q, Wu C-S, Chesters D, Zhu C-D, and A.B. Zhang*. (2012) Phylogenetic Reconstruction and DNA Barcoding for Closely Related Pine Moth Species (Dendrolimus) in China with Multiple Gene Markers. PLoS ONE 7(4): e32544.

Examples

```
data(pineMothITS1)
pineMothITS1
```

`pineMothITS2` *pine Moth ITS2*

Description

ITS2 sequences of seven closely related pine moths species sampled in China.

Usage

```
data("pineMothITS2")
```

Format

The format is: List of 97 $ A22,Lasiocampidae_Dendrolimus_punctatus: raw [1:568] 48 48 48 28 ... $ A23,Lasiocampidae_Dendrolimus_punctatus: raw [1:574] 48 18 28 18 ... $ A29,Lasiocampidae_Dendrolimus_punctat raw [1:569] 88 48 48 48 ... $ A52,Lasiocampidae_Dendrolimus_punctatus: raw [1:570] 48 18 28 18 ... - attr(*, "class")= chr "DNAbin"

Details

ITS2 used as DNA barcodes for closely related rine moth species (Dendrolimus) in China.

Source

http://dx.plos.org/10.1371/journal.pone.0064428.

References

Dai Q-Y, Gao Q, Wu C-S, Chesters D, Zhu C-D, and A.B. Zhang*. (2012) Phylogenetic Reconstruction and DNA Barcoding for Closely Related Pine Moth Species (Dendrolimus) in China with Multiple Gene Markers. PLoS ONE 7(4): e32544.

Examples

```
data(pineMothITS2)
pineMothITS2
```

sample.ref *Sample Random Datasets from References (DNAbin)*

Description

Randomly sample reference data at different levels of taxon.

Usage

```
sample.ref(ref, sample.porp = 0.5, sample.level = "full")
```

Arguments

ref	Object of class "DNAbin" used as a reference dataset, which contains taxon information.
sample.porp	a numeric value between 0 and 1, indicating proportion of samples to draw at each level of taxon.
sample.level	a character string choosing from c("full","family","genus","species").

Value

a list containing the selected samples and the samples left, in DNAbin format stored in a matrix or a list.

Note

the ref must contain information on taxonomy, in format like, ">LS0909030M,Noctuidae_Himalaea_unica", i.e., "seqID,family_genus_species", or ">LS0909030M,Himalaea_unica"; in case there is only one sample/individual for a taxon level, this sample will be retained in ref.selected.

Author(s)

Ai-bing ZHANG, PhD. CNU, Beijing, CHINA.

References

zhangab2008(at)mail.cnu.edu.cn;

Examples

```
data(TibetanMoth)
data(pineMothITS2)
ref<-TibetanMoth
ref2<-pineMothITS2
out<-sample.ref(ref,sample.porp=0.5,sample.level="full")
out
out2<-sample.ref(ref2,sample.porp=0.5,sample.level="full")
out2
```

save.ids *Save Identifications*

Description

Output identified results to an outfile.

Usage

```
save.ids(outfile = "identified.txt", ids)
```

Arguments

outfile	character string to indicate outfile name.
ids	object of class "BarcodingR", which contains identified taxon information.

Value

no value returned,but an output file.

Author(s)

Ai-bing ZHANG, PhD. CNU, Beijing, CHINA.

References

zhangab2008(at)mail.cnu.edu.cn

See Also

barcoding.spe.identify()

Examples

```
data(TibetanMoth)
ref<-as.DNAbin(as.character(TibetanMoth[1:50,]))
que<-as.DNAbin(as.character(TibetanMoth[50:60,]))
bsi<-barcoding.spe.identify(ref, que, method = "fuzzyId")
bsi
save.ids(outfile="identified.txt",bsi)
```

summarize.ref *Summarize Reference Data*

Description

Summarize taxon information, sequence statistics,barcodes numbers per species for reference dataset.

Usage

```
summarize.ref(ref, taxonStat = TRUE, seqStat = TRUE, barcodeStat = TRUE)
```

Arguments

ref	object of class "DNAbin" used as a reference dataset, which contains taxon information
taxonStat	logic value to indicate whether the item is calculated
seqStat	logic value to indicate whether the item is calculated
barcodeStat	logic value to indicate whether the item is calculated

Value

a list containing taxon statistics, sequence statistics, population parameters,barcoding statistics ()

Author(s)

Ai-bing ZHANG, Meng-di HAO, CNU, Beijing, CHINA.

References

zhangab2008(at)mail.cnu.edu.cn./zhangab2008(at)gmail.com.

Examples

```
data(TibetanMoth)
s.r<-summarize.ref(TibetanMoth,taxonStat=TRUE,seqStat=TRUE,barcodeStat=TRUE)
s.r
```

TDR2 *TDR2 Species Membership Value*

Description

To calculate TDR value for a set of queries and one potential species. Its value is in the range of [0,1], 0 indicates extremly weak species membership, values close 1 indicating strong species membership.

Usage

```
TDR2(oneSpe, que, boot, boot2)
```

Arguments

oneSpe	object of class "DNAbin" which contains DNA squences from one speices
que	object of class "DNAbin" which contains DNA squences different samples
boot	a numeric value indicating times of resampling along sequence columns
boot2	a numeric value indicating times of resampling along sequence rows (different samples)

Value

a numeric vector represents TDR values for each query against the species

Note

oneSpe and que should be the same in sequence length, i.e., they should be aligned in prior. It's strongly recommended that oneSpe should have large enough sample size,e.g., 20.

Author(s)

Ai-bing ZHANG, PhD. CNU, Beijing, CHINA, contact at zhangab2008(at)mail.cnu.edu.cn

References

Jin Q, L,J.He, A.B. Zhang* (2012). A Simple 2D Non-Parametric Resampling Statistical Approach to Assess Confidence in Species Identification in DNA Barcoding-An Alternative to Likelihood and Bayesian Approaches. PLoS ONE 7(12): e50831. doi:10.1371/journal.pone.0050831.http://dx.plos.org/10.1371/journal.

Examples

```
data(TibetanMoth)
sampleSpeNames<-NAMES(TibetanMoth)
Spp<-gsub(".+,","",sampleSpeNames)
oneSpe<-TibetanMoth[grep("Macdunnoughia_crassisigna", Spp, value = FALSE,fixed = TRUE),]
que<-TibetanMoth[grep("Agrotis_justa", Spp, value = FALSE,fixed = TRUE),]
que2<-oneSpe[1:2,]
out<-TDR2(oneSpe,que, boot=10,boot2=10) ### true false identification
out2<-TDR2(oneSpe,que2, boot=10,boot2=10) ### true positive identification
```

TibetanMoth *Tibetan Moth*

Description

COI DNA barcodes of Tibetan Moth in China.

Usage

```
data("TibetanMoth")
```

Format

The format is: 'DNAbin' raw [1:319, 1:630] a t a a ... - attr(*, "dimnames")=List of 2 ..$: chr [1:319] "LS0909030M,Noctuidae_Himalaea_unica" "LZ0827026M,Noctuidae_Amphipyra_pyramidea" "ML0829010M,Noctuidae_Auchmis_saga" "BM0830055M,Noctuidae_Auchmis_saga" ... ..$: NULL

Details

COI DNA barcodes of Tibetan Moth (Noctuidae) in China.

Source

http://dx.plos.org/10.1371/journal.pone.0064428.

References

Q. Jin, H.L. Han, X.M. Hu, X.H. Li,C.D. Zhu,S. Y. W. Ho, R. D. Ward, A.B. Zhang* . (2013).Quantifying Species Diversity with a DNA Barcoding-Based Method: Tibetan Moth Species (Noctuidae) on the Qinghai-Tibetan Plateau. PloS One 8: e644.

Examples

```
data(TibetanMoth)
TibetanMoth
```

Index