Package 'geneticRF'

February 28, 2015

Type Package

Title Genetic Random Forest

Description Random Forest Analysis For Genetic Data

Version 0.9.0

Date 2014-11-18

Author Eric Archer

Maintainer Eric Archer <eric.archer@noaa.gov>

Imports randomForest, strataG

License GNU General Public License

R topics documented:

`classify.by.hap.freq` *Classify By Haplotype Frequency*

Description

Classify samples by haplotype frequency.

Usage

```
classify.by.hap.freq(seq.df)
```

Arguments

seq.df	data.frame of stratified and aligned sequences from create.seq.df.
...	arguments passed to randomForest.

Value

a list containing a data.frame of summary statistics (smry), and the randomForest object (rf).

Author(s)

Eric Archer <eric.archer@noaa.gov>

---

collapse.to.haps *Collapse To Haplotypes*

Description

Collapse sequence data.frame to unique haplotypes.

Usage

```
collapse.to.haps(seq.df)
```

Arguments

seq.df	data.frame of stratified and aligned sequences from create.seq.df.

Author(s)

Eric Archer <eric.archer@noaa.gov>

---

create.seq.df *Create Sequence Data.Frame*

Description

Create data.frame of variable sites from gtypes object.

Usage

```
create.seq.df(g, label = NULL)
```

Arguments

g	haploid gtypes object with aligned sequences.
label	label to add to beginning of each stratum name.

Value

a data.frame where the first column lists the (strata) and every column column afterwards is a variable site. All columns are factors.

Author(s)

Eric Archer <eric.archer@noaa.gov>

---

gtype.rf *gtype Random Forest*

Description

Conduct Random Forest on a gtypes object.

Usage

```
gtype.rf(g, pairwise = FALSE, ...)
```

Arguments

g	haploid gtypes object with aligned sequences.
pairwise	do analysis on all pairwise combinations of strata?
...	arguments passed to randomForest.

Value

a list containing a data.frame of summary statistics (smry), and the randomForest object (rf). If pairwise is TRUE then the rf element is a list of randomForest results for each row in smry.

Author(s)

Eric Archer <eric.archer@noaa.gov>

---

make.haps *Make Haplotypes*

Description

Identify haplotypes in data.frame of aligned.

Usage

```
make.haps(x)
```

Arguments

x	data.frame of base pairs.

Author(s)

Eric Archer <eric.archer@noaa.gov>

min.votes *Minimum Votes*

Description

Calculate Random Forest confidence intervals

Usage

```
## S3 method for class 'votes'
min(rf, mv.vec)
```

Arguments

rf	a randomForest object.
mv.vec	a vector of of minimum vote values to return.

Author(s)

Eric Archer <eric.archer@noaa.gov>

pct.diag *Extract Percent Diagnosable*

Description

Calculate Random Forest confidence intervals

Usage

```
pct.diag(rf, pd.vec)
```

Arguments

rf	a randomForest object.
pd.vec	a vector of of minimum vote values to return.

Author(s)

Eric Archer <eric.archer@noaa.gov>

rf.conf.int *Confidence Intervals*

Description

Calculate Random Forest confidence intervals

Usage

```
rf.conf.int(rf, conf.level = 0.95)
```

Arguments

rf	a randomForest object
conf.level	confidence level for the binom.test confidence interval

Author(s)

Eric Archer <eric.archer@noaa.gov>

rf.species.id *Random Forest Species ID*

Description

Predict species for unknown samples based on reference samples using Random Forest.

Usage

```
rf.species.id(g, ref.strata = NULL, unk.strata = NULL, ...)
```

Arguments

g	haploid gtypes object with aligned sequences.
ref.strata	a character vector of 2 or more strata in g top use as reference samples.
unk.strata	a character vector of strata representing unknown samples.
...	arguments passed to randomForest.

Value

a list with the following elements:

pred	vector of species predictions.
prob	matrix of species assignment probabilities.
rf	the randomForest model object.

Author(s)

Eric Archer <eric.archer@noaa.gov>

---

seq.df.rf *Sequence Random Forest*

Description

Conduct Random Forest on stratified data.frame of sequences.

Usage

```
## S3 method for class 'df.rf'
seq(seq.df, ...)
```

Arguments

seq.df	data.frame of stratified and aligned sequences from create.seq.df.
...	arguments passed to randomForest.

Value

a list containing a data.frame of summary statistics (smry), and the randomForest object (rf).

Author(s)

Eric Archer <eric.archer@noaa.gov>

---

shared.haps.pd *Diagnosability Based on Shared Haplotypes*

Description

Maximum diagnosability based on shared haplotypes.

Usage

```
shared.haps.pd(seq.df)
```

Arguments

seq.df	data.frame of stratified and aligned sequences from create.seq.df.

Author(s)

Eric Archer <eric.archer@noaa.gov>

unique.hap.by.site *Make Haplotypes*

Description

Identify haplotypes in data.frame of aligned.

Usage

```
## S3 method for class 'hap.by.site'
unique(x)
```

Arguments

x	data.frame of base pairs.

Author(s)

Eric Archer <eric.archer@noaa.gov>

Index