Package 'NeON'

October 24, 2013

Type Package

Title Package for the effective population size estimation

Version 1.0

Depends R (>= 2.13.0), psych

Date 2013-10-11

Author Massimo Mezzavilla and Silvia Ghirotto

Maintainer Massimo Mezzavilla <`massimo.mezzavilla@burlo.trieste.it`>

Description The effective population size (Ne) is one of the most
interesting population parameter. Linkage disequilibrium (LD)
patterns contain information about demographic changes of a
population, can be used to monitor variation in population size
through time. This package implements several functions to
estimate the effective population size from SNP data and
estimate Time of divergence between population.

License GPL (>= 2)

LazyLoad yes

R topics documented:

NeON-package   *Package for the effective population size estimation*

Description

The effective population size (Ne) is one of the most interesting population parameter, Linkage disequilibrium (LD) patterns contain information about demographic chnages of a popualtion, can be used to monitor variation in population size through time. This package implements several functions to estimate the effective population size from SNP data and estimate Time of divergence between population.

Details

Package:	NeON
Type:	Package
Version:	1.0
Date:	2013-05-17
License:	GPL (>=2)
LazyLoad:	yes

NeLD   *Estimation of linkage disequilibrium*

Description

This function use PLINK program to estimates the squared correlation coefficient of linkage disequilibrium (r2) between markers

Usage

```
NeLD(plink.file, geno = 0.02, mind = 0.9, ld.window.kb = 500, ld.window = 9999,
outfile = "output.ld")
```

Arguments

`plink.file`	prefix of your PLINK .bed .bim .fam files
`geno`	markers missing genotyping rate allowed
`mind`	individuals missing genotype rates allowed
`ld.window.kb`	sliding window of estimate linkage disequilibrium along each chromosome
`ld.window`	number of max confrontations in each window
`outfile`	name of your output file

Value

return a PLINK output file output.ld with all the comparison between each marker and their r^2 values

References

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics, 81(3), 559-575

See Also

'Nemap','NeUpdate', 'Nestimate'

Nemap   *function to collect the genetic map information*

Description

This function prepares the file to update the genetic map information of the markers in your dataset, basing on the recombination rates and hotspots compiled file present in HapMap website

Usage

```
Nemap(bim.file, map.file)
```

Arguments

bim.file	PLINK formatted .bim file
map.file	genetic path where the genetic maps are present and prefix of each map file (one for each chromosome)

Details

The format of the genetic maps must be like that: into the same folder one file for each chromosome, each file must be structured with five columns with an header ("Chromosome" "Position(bp)" "Rate(cM/Mb)" "Map(cM)")

Value

a list of SNP identifiers which corresponds to the match of genetic maps physical position and the physical position of your .bim file

References

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics, 81(3), 559-575

See Also

'NeUpdate'

Neplot   *Function to plot the demography of a population*

Description

This function take the output of NeMed function and plot the demography of the population,

Usage

```
Neplot(Ne.file, approx = TRUE, ylim = c(0, 15000), xlim = c(200, 6000),
main = "Ne from linkage disequilibrium", xlab = "Generation ago",
ylab = "Ne", ci = TRUE)
```

Arguments

`Ne.file`	output from NeMed function
`approx`	if you wish to make and interpolation of each point, default TRUE
`ylim`	a vector with the upper and lower limits of Ne axis, default (0,15000)
`xlim`	a vector with upper and lower temporal interval of generations where you want to plot the demography, default(200, 6000)
`main`	title of the plot
`xlab`	label for the x axis
`ylab`	label for the y axis
`ci`	if you want to plot the confidence interval, only if previosly calculated with NeMed, default(TRUE)

Details

This function is useful to obtain a graphical representation of the changes in the effective population size of a population over time

See Also

'Ne_Med'

Nestimate   *Estimation of effective population size from linkage disequilibrium*

Description

This function estimates the effective population size from linkage disequilibrium

Usage

```
Nestimate(file.ld, sample.size, min.R2 = 0.001, max.R2 = 0.999,
method = 'MG', min.cfr = 5)
```

Arguments

`file.ld`	output file from PLINK linkage disequilibrium analysis
`sample.size`	number of individuals used to estimate r^2 into the population
`min.R2`	minimum value of r^2 allowed into each comparison, default(0.001)
`max.R2`	maximum value of r^2 allowed into each comparison default(0.999)
`method`	two possible method are allowed to estimate Ne trough generations: 'MG' and 'McEvoy', default 'MG'
`min.cfr`	minimum number of comparison in each bin that are allowed to estimate the mean Ne into each bin, (default 5)

Details

This function estimate the demography of the population using each chromosome as an independent replicates,it applies the well-known formula Ne= 1/(4c) * [(1/r2) - 2],where c is the distance between genetic markers in Morgan. Nestimate creates several categories of recombination distance ,with incremental upper boundaries of 0.005 centiMorgan (cM) up to 0.25 cM, and calculates the r2 for each pairs of markers in each recombination distance category. Two methods of binning are possible: the first one is identical to the method used into McEvoy et al. were 50 bins were created, each bin correspond to a specific intervals and are not overlapping, the second method 'MG' (Mezzavilla-Ghirotto) created 250 bins that are overlapping, in this way we have more bins and more temporal points in which we could estimate the effective population size this could provide a fine scale of demography reconstruction

Value

a dataframe with the values of the effective population size and the correspondent time in the past (in generation), for each bin, for each chromosome.

Note

We suggest to no decrease the min.cfr under 5, in order to have reliable estimates of Ne

References

Hayes, B. J., Visscher, P. M., McPartlan, H. C., & Goddard, M. E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Research, 13(4), 635-643.

Tenesa, A., Navarro, P., Hayes, B. J., Duffy, D. L., Clarke, G. M., Goddard, M. E., & Visscher, P. M. (2007). Recent human effective population size estimated from linkage disequilibrium. Genome research, 17(4), 520-526

McEvoy, B. P., Powell, J. E., Goddard, M. E., & Visscher, P. M. (2011). Human population dispersal ,'Out of Africa' estimated from linkage disequilibrium and allele frequencies of SNPs. Genome research, 21(6), 821-829

See Also

'Ne_Med', 'Ne_CI','NeLd'

NeUpdate   *Function to prepare PLINK input file for LD computation*

Description

This function relies on the snp.list file created by the previous function (Nemap) to update your PLINK data file with the correct genetic map information

Usage

```
NeUpdate(plink.file, snp.list, outfile)
```

Arguments

`plink.file`   prefix of the .bed .bim .fam file to update and format for Ne estimation

`snp.list`   list of markers informations obtained from the function NeMap

`outfile`   prefix of your PLINK output files formatted for LD computation and Ne estimation

Details

we suggest to not make the same prefix of the previous plink file for the outfile. The PLINK executable has to be in the same folder of the data files

References

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics, 81(3), 559-575

See Also

'Nemap'

Ne_CI   *Confidence intervals for the long-term Ne*

Description

This function estimate the long-term Ne and its confidence intervals (default 5th ,50th and 95th percentile, but they can be changed when calling the function)

Usage

```
Ne_CI(Nestimate.output, ci = c(0.05, 0.5, 0.95))
```

Arguments

Nestimate.output

> a dataframe obtained from the Nestimate function

ci

> a vector,it contains which quantiles must be estimated for the long term Ne, default (5th, 50th and 95th percentile)

Details

The long term Ne is calculated as the harmonic mean of the effective population sizes along the generations in the past i.e. in each recombination distance category. The confidence interval of the long term Ne is calculated using each chromosome as a replicate

Value

a dataframe which contains the quantiles of the long term Ne

See Also

'Nestimate'

Ne_Med   *Estimation of the effective population size over time*

Description

This function calculates the demographic function(effective population size over time) of a population along with its confidence interval, for each bin.

Usage

```
Ne_Med(Nestimate.output, method = 'MG', ci = FALSE,
ci.int = c(0.05, 0.5, 0.95))
```

Arguments

Nestimate.output

> a dataframe obtained from the Nestimate function

method

> two possible methods used for estimate the Ne in Nestimate: 'Mcevoy' or 'MG', (default 'MG')

ci

> if confidence intervals must be calculated, (default FALSE)

ci.int

> if ci=TRUE which quantiles of the distribution must be calculated, (default 5th, 50th and 95th percentile)

Details

Ne_Med takes the output of Nestimate and calculates the median of effective population (calculated from all autosomes) estimated for each temporal point, depending of the method used to estimate Ne: 'McEvoy' or 'MG' you could have 50 or 250 temporal points, you must specify the method used. In addition you could estimate different quantiles of the distribution of Ne for each temporal point.

Value

This function returns a dataframe with the first three columns indicating the quantiles of the distribution of the effective population size for each bin over all chromosomes (the default values are the 90 percent confidence interval and the median value) and the last column with the moment in time to which the effective population size is referred (1/2c generation ago).

References

McEvoy, B. P., Powell, J. E., Goddard, M. E., & Visscher, P. M. (2011). Human population dispersal ,'Out of Africa' estimated from linkage disequilibrium and allele frequencies of SNPs. Genome research, 21(6), 821-829

See Also

'Nestimate'

Tdverg   *Time of divergence between populations*

Description

This function returns a matrix of the time of divergence between populations in generation

Usage

```
Tdverg(Fst, All_H)
```

Arguments

Fst	a distance matrix with Fst values estimates using Reynolds
All_H	a text file with a list of the long-term Ne for each population, with an header that match the population labels reported in the FST matrix

Details

This function applies the following formula for each pair of populations: $T = \ln(1-FST)/\ln(1-1/2Ne)$, where Ne is the mean of the long-term Ne for the considered pair of populations. It is possible to obtain the confidence intervals of the divergence time using the 5th and 95th percentile of the long-term Ne (calculated from Ne_CI) as input. The order of the populations into the Fst matrix and into the vector of Ne could be different but the names of the populations must be the same in each item.

Value

a matrix with the time of divergence between population measured as generations

References

Hayes, B. J., Visscher, P. M., McPartlan, H. C., & Goddard, M. E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Research, 13(4), 635-643.

Tenesa, A., Navarro, P., Hayes, B. J., Duffy, D. L., Clarke, G. M., Goddard, M. E., & Visscher, P. M. (2007). Recent human effective population size estimated from linkage disequilibrium. Genome research, 17(4), 520-526.

McEvoy, B. P., Powell, J. E., Goddard, M. E., & Visscher, P. M. (2011). Human population dispersal 'Out of Africa' estimated from linkage disequilibrium and allele frequencies of SNPs. Genome research, 21(6), 821-829

See Also

'Ne_CI'

Index