Package 'rehh'

February 13, 2019

Maintainer Mathieu Gautier <mathieu.gautier@inra.fr>

Author Mathieu Gautier, Alexander Klassmann and Renaud Vitalis

Version 2.0.4

License GPL (>= 2)

Title Searching for Footprints of Selection using ``Extended Haplotype Homozygosity'' Based Tests

Description Population genetic data in form of ``Single Nucleotide Polymorphisms'' (SNPs) is often used to identify genomic regions that have been under recent natural or artificial selection and might provide clues about the molecular mechanisms of adaptation. The concept of an ``Extended Haplotype Homozygosity'' (EHH), introduced by (Sabeti 2002) <doi:10.1038/nature01140>, has given rise to several derived statistics designed for whole genome scans. The package provides functions to compute three of these, namely: ``iHS'' (Voight 2006) <doi:10.1371/journal.pbio.0040072> for detecting selection within a single population as well as ``Rsb'' (Tang 2007) <doi:10.1371/journal.pbio.0050171> and ``XP-EHH'' (Sabeti 2007) <doi:10.1038/nature06250> to detect (differential) selection between two populations. Various plotting functions are also included to facilitate visualization and interpretation of these statistics.

Depends R (>= 2.10), rehh.data, gplots, methods

Suggests knitr, rmarkdown, bookdown

VignetteBuilder knitr

NeedsCompilation yes

Repository CRAN

Date/Publication 2019-02-13 16:20:11 UTC

R topics documented:

`rehh-package` *Searching for footprints of selection using "Extended Haplotype Homozygosity" based statistics*

Description

Population genetic data in form of "Single Nucleotide Polymorphisms" (SNPs) is often used to identify genomic regions that have been under recent natural or artificial selection and might provide clues about the molecular mechanisms of adaptation. The concept of an "Extended Haplotype Homozygosity" (EHH), introduced by (Sabeti 2002), has given rise to several derived statistics designed for whole genome scans. The package provides functions to compute three of these, namely: "iHS" (Voight 2006) for detecting selection within a single population as well as "Rsb" (Tang 2007) and "XP-EHH" (Sabeti 2007) to detect (differential) selection between two populations. Various plotting functions are also included to facilitate visualization and interpretation of these statistics.

Details

Package:	rehh
Version:	2.0.4
License:	GPL(>=2)
Depends:	gplots , methods

Index:

`bifurcation.diagram`	`Plot of an haplotype bifurcation diagram`
`calc_ehh`	`EHH and iHH computations at a given core SNP`
`calc_ehhs`	`EHHS and iES computations at a`

	given core SNP
data2haplohh	Converting data into an object of class haplohh
distribplot	Distribution of standardized iHS, Rsb or XP-EHH values
haplohh_cgu_bta12	Example of an 'haplohh' object
ies2rsb	Compute Rsb (standardized ratio of iES from two populations)
ies2xpehh	Compute XP-EHH (standardized ratio of iES from two populations)
ihh2ihs	Compute iHS (standardized iHH)
ihsplot	Plot iHS over a genome
make.example.files	Generate example input files
rsbplot	Plot Rsb over a genome
xpehhplot	Plot XP-EHH over a genome
scan_hh	Computing EHH based statistics over a whole chromosome

References

Gautier M., Klassmann A., and Vitalis R. (2017). rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources*, **17**, 78–90.

Gautier M. and Vitalis R. (2012). rehh: An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*, **28**(8), 1176–1177.

Gautier M. and Naves M. (2011). Footprints of selection in the ancestral admixture of a New World Cattle breed. *Molecular Ecology*, (20), 3128–3143.

Sabeti P. C., Reich D. E., Higgins J. M., Levine H. Z. P., Richter D. J., others, (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832-837.

Sabeti P. C., Varilly P., Fry B., Lohmueller J., Hostetter E., others, (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913-918.

Tang K., Thornton K. R., Stoneking M., (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol* **5**, e171.

Voight B. F., Kudaravalli S., Wen X., Pritchard J. K., (2006). A map of recent positive selection in the human genome. *PLoS Biol* **4**, e72.

`bifurcation.diagram` *plot of an haplotype bifurcation diagram*

Description

A haplotype bifurcation diagram visualizes the decay of EHH of a "core" allele of a focal SNP at increasing distances.

Usage

```
bifurcation.diagram(haplohh, mrk_foc, all_foc=1, nmrk_l=10, nmrk_r=10,
                limhapcount = 10, refsize = 0.1, linecol = "blue",
                main_leg = NA, xlab_leg = "Position")
```

Arguments

haplohh	An object of class haplohh (see data2haplohh).
mrk_foc	Either the number of the marker in the haplohh object (as integer) or its name (as string) to specify the focal marker
all_foc	either 1 or 2 depending on the chosen core allele (resp. ancestral or derived)
nmrk_l	Number of markers to be considered upstream of the focal SNP
nmrk_r	Number of markers to be considered downstream of the focal SNP
limhapcount	Minimal number of haplotypes containing the core allele at the focal SNP
refsize	Controls the relative width of the diagram lines on the plot
linecol	Color of the lines on the diagram
main_leg	Main legend of the diagram. By default, the name of the SNP together with the allele considered
xlab_leg	Legend on the xaxis of the diagram

Details

The function 'bifurcation.diagram()' draws haplotype bifurcation diagrams [@Sabeti2002] that visualize the decay of *EHH* around a focal SNP. A stark contrast of ancestral and derived bifurcation diagrams should correspond to outlier values of *ihs*. In the plot the root (focal SNP) is identified by a vertical dashed line. The diagram is bi-directional, portraying decay along both sides of the focal SNP. Moving in one direction, each marker is an opportunity for a bifurcation further differentiating (extended) haplotypes. The thickness of the lines corresponds to the number of chromosomes with the same extended haplotype.

Value

The function returns a plot.

References

Sabeti, P.C. and Reich, D.E. and Higgins, J.M. and Levine, H.Z.P and Richter, D.J. and Schaffner, S.F. and Gabriel, S.B. and Platko, J.V. and Patterson, N.J. and McDonald, G.J. and Ackerman, H.C. and Campbell, S.J. and Altshuler, D. and Cooper, R. and Kwiatkowski, D. and Ward, R. and Lander, E.S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**, 832–837.

Examples

```
#example haplohh object (280 haplotypes, 1424 SNPs)
#see ?haplohh_cgu_bta12 for details
data(haplohh_cgu_bta12)
#plotting bifurcation diagram for both ancestral and derived allele
#from the focal SNP at position 456
#which display a strong signal of selection
layout(matrix(1:2,2,1))
#ancestral allele
bifurcation.diagram(haplohh_cgu_bta12,mrk_foc=456,all_foc=1,
nmrk_l=20,nmrk_r=20)
#derived allele
bifurcation.diagram(haplohh_cgu_bta12,mrk_foc=456,all_foc=2,
nmrk_l=20,nmrk_r=20)
##
dev.off()
```

---

calc_ehh *EHH and iHH computations at a given core SNP*

Description

Compute Extended Haplotype Homozygosity (EHH) and integrated EHH (iHH) for a given focal SNP.

Usage

```
calc_ehh(haplohh, mrk, limhaplo = 2, limehh = 0.05,
    scalegap = NA, maxgap = NA,
    discard_integration_at_border = TRUE, plotehh = TRUE,
    lty = 1, lwd = 1.5, col = c("blue", "red"), xlab = "Position",
    ylab = expression(Extended ~ haplotype ~ homozygosity ~ (italic(EHH))),
    cex.lab = 1.25, main = NA, cex.main = 1.5)
```

Arguments

haplohh	An object of class haplohh (see data2haplohh).
mrk	Either the number of the marker in the haplohh object (as integer) or its name (as string) to specify the focal marker
limhaplo	Minimal number of haplotypes to continue computing EHH away from the core SNP. Useless, if no missing data. However, when some data are missing, haplotypes with missing data are removed from the computation. Hence as we compute EHH further from the core SNP, less haplotypes are expected
limehh	Limit at which EHH stops to be evaluated
scalegap	Scales gaps larger than the specified size to the specified size (default=NA, i.e. no scaling)

maxgap	Maximum allowed gap in bp between two SNPs below which EHH stops to be evaluated (default=NA, i.e., no limitation)
discard_integration_at_border	If TRUE and if first or last marker or a gap (larger than maxgap) is reached and EHH is greater than limehh, then iHH is set to NA
plotehh	If TRUE, EHH values for both ancestral and derived allele are plotted for each position
lty	Line type for the ancestral and derived allele iHH (respectively) curves
lwd	Line width for the ancestral and derived allele iHH (respectively) curves
col	Color for the ancestral and derived allele iHH (respectively) curves
xlab	Legend for the x–axis
ylab	Legend for the y–axis
cex.lab	Size of the axis legend
main	Main legend of the EHHS plot
cex.main	Size of the main legend

Details

EHH are computed at each position upstream and downstream of the focal SNP for both derived and ancestral allele. This allows in turn the computation of the integrated EHH relative to map distances (iHH).

Value

The returned value is a list containing the following elements:

ehh	A matrix of two rows and nsnps columns containing EHH estimates at each chromosome position relative to the focal SNP for the ancestral (first row) and derived (second row) alleles.
nhaplo_eval	A matrix of two rows and nsnps columns containing the number of evaluated haplotypes at each chromosome position relative to the focal SNP for the ancestral (first row) and derived (second row) alleles.
freq_al1	the frequency of the ancestral allele of the focal SNP.
ihh	A vector of two elements corresponding respectively to the iHH (integrated EHH) for the ancestral and derived allele.

References

Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, **20**, 3128–3143.

Sabeti, P.C. et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**, 832–837.

Sabeti, P.C. et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, **449**, 913–918.

Tang, K. and Thornton, K.R. and Stoneking, M. (2007). A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *Plos Biology*, **7**, e171.

Voight, B.F. and Kudaravalli, S. and Wen, X. and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. *Plos Biology*, **4**, e72.

See Also

calc_ehhs,data2haplohh,scan_hh

Examples

```
#example haplohh object (280 haplotypes, 1424 SNPs)
#see ?haplohh_cgu_bta12 for details
data(haplohh_cgu_bta12)

#computing EHH statisitics for the focal SNP at position 456
# which displays a strong signal of selection
res.ehh<-calc_ehh(haplohh_cgu_bta12,mrk=456)
```

`calc_ehhs` *EHHS and iES computations at a given core SNP*

Description

Compute site-specific Extended Haplotype Homozygosity (EHHS) and integrated EHHS (iES) for a given focal SNP.

Usage

```
calc_ehhs(haplohh, mrk, limhaplo = 2, limehhs = 0.05,
    scalegap = NA, maxgap = NA,
    discard_integration_at_border = TRUE, plotehhs = TRUE,
    lty = 1, lwd = 1.5, col = c("blue", "red"), xlab = "Position",
    ylab = expression(Site ~ specific ~ italic(EHH) ~ (italic(EHHS))),
    cex.lab = 1.25, main = NA, cex.main = 1.5)
```

Arguments

`haplohh`	An object of class `haplohh` (see `data2haplohh`).
`mrk`	Either the number of the marker in the haplohh object (as integer) or its name (as string) to specify the focal marker
`limhaplo`	Minimal number of haplotypes to continue computing EHHS away from the core SNP. Useless, if no missing data. However, when some data are missing, haplotypes with missing data are removed from the computation. Hence as we compute EHH further from the core SNP, less haplotypes are expected
`limehhs`	Limit at which EHHS stops to be evaluated

scalegap	Scales gaps larger than the specified size to the specified size (default=NA, i.e. no scaling)
maxgap	Maximum allowed gap in bp between two SNPs below which EHHS stops to be evaluated (default=NA, i.e., no limitation)
discard_integration_at_border	If TRUE and if first or last marker or a gap (larger than maxgap) is reached and EHHS is greater than limehh, then IES is set to NA
plotehhs	If TRUE, EHHS estimates are plotted for each position
lty	Line type for the EHHS_Sabeti_et_al_2007 and EHHS_Tang_et_al_2007 (respectively) curves
lwd	Line width for the EHHS_Sabeti_et_al_2007 and EHHS_Tang_et_al_2007 (respectively) curves
col	Color for the EHHS_Sabeti_et_al_2007 and EHHS_Tang_et_al_2007 (respectively) curves
xlab	Legend for the x–axis
ylab	Legend for the y–axis
cex.lab	Size of the axis legend
main	Main legend of the EHHS plot
cex.main	Size of the main legend

Details

EHHS are computed at each position upstream and downstream of the focal SNP. This allows in turn the computation of the integrated EHHS relative to map distances (iES).

Value

The returned value is a list containing the following elements:

EHHS_Tang_et_al_2007	A vector of nsnps columns containing EHHS estimates at each chromosome position relative to the focal SNP computed as described in the Tang et al. (2007).
EHHS_Sabeti_et_al_2007	A vector of nsnps columns containing EHHS estimates at each chromosome position relative to the focal SNP computed as described in the Sabeti et al. (2007).
nhaplo_eval	A matrix of two rows and nsnps columns containing the number of evaluated haplotypes at each chromosome position relative to the focal SNP for the ancestral (first row) and derived (second row) alleles.
IES_Sabeti_et_al_2007	Integrated EHHS (computed using the estimator by Sabeti et al. (2007)) over the chromosome.
IES_Tang_et_al_2007	Integrated EHHS (computed using the estimator by Tang et al. (2007)) over the chromosome.

References

Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, **20**, 3128–3143.

Sabeti, P.C. et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, **449**, 913–918.

Tang, K. and Thornton, K.R. and Stoneking, M. (2007). A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *Plos Biology*, **7**, e171.

See Also

calc_ehh,data2haplohh,scan_hh

Examples

```
#example haplohh object (280 haplotypes, 1424 SNPs)
#see ?haplohh_cgu_bta12 for details
data(haplohh_cgu_bta12)
#computing EHH statistics for the focal SNP at position 456
#which displays a strong signal of selection
res.ehhs<-calc_ehhs(haplohh_cgu_bta12,mrk=456)
```

---

data2haplohh *Converting data into an object of class haplohh*

Description

Converts input data files to an object of class haplohh.

Usage

```
data2haplohh(hap_file, map_file, min_maf = 0, min_perc_geno.hap = 100,
                min_perc_geno.snp = 100, chr.name = NA, popsel = NA,
                recode.allele = FALSE, haplotype.in.columns = FALSE)
```

Arguments

`hap_file`	Path to the file containing haplotype data (see details section below for information about input file format)
`map_file`	Path to the file containing map information (see details section below for information about input file format
`min_maf`	Threshold on Minor Allele Frequency (SNPs displaying a MAF lower than `min_maf` are discarded)
`min_perc_geno.hap`	Threshold on percentage of missing data for haplotypes (Haplotypes with less than `min_perc_geno.hap` percent SNPs genotyped are discarded). By default, `min_perc_geno.hap`=100, hence only fully genotyped haplotypes are retained

`min_perc_geno.snp`
: Threshold on percentage of missing data for SNPs (SNPs genotyped on less than `min_perc_geno.snp` percent haplotypes are discarded). By default, `min_perc_geno.snp`=100, hence only fully genotyped SNPs are retained

`chr.name`
: Name of the chromosome considered (relevant if several chromosomes are represented in the map file)

`popsel`
: Code of the population considered in the fastPHASE output haplotype file (relevant if `hap_file` is a fastPHASE output and haplotypes originate from different population)

`recode.allele`
: If TRUE, allele in the haplotypes are recoded according to the map file information. If FALSE a rough verification is performed to check only 0 (code for missing data), 1 (code for ancestral allele) or 2 (code for derived allele) are present in the haplotype file

`haplotype.in.columns`
: If TRUE, phased input haplotypes are assumed to be in columns (as produced by the SHAPEIT2 program (O'Connell et al., 2014).

Details

Three haplotype input formats are supported:

- a standard format with haplotypes in rows and snps in columns (with no header, but a haplotype id)
- a (transposed) format similar to the one produced by the phasing program SHAPEIT2 program (O'Connell et al., 2014) in which haplotypes are in columns and snps in rows (with no header and no snp id)
- output files from fastPHASE program (Sheet and Stephens, 2006). If the input haplotypes are not in transposed format (i.e., `haplotype.in.columns` is FALSE, as by default), the function automatically checks if the file is in fastPHASE output format. In this latter case, if haplotypes from several different population were phased simultaneously (-u fastPHASE option was used), the function ask interactively which population should be considered (a list of population number are proposed) unless specified with the `popsel` argument.

The map file contains SNP information in five columns:

- SNP name/id
- chromosome
- position (physical or genetic)
- ancestral allele encoding
- derived allele encoding

The SNPs must be in the same order as in the haplotype for the chromosome considered. If several chromosomes are represented in the map file, one can provide the name of the chromosome of interest (corresponding to the haplotype under study) with the `chr.name` argument. Haplotypes are recoded (if the `recode.allele` option is activated) according to the ancestral and derived allele definition available in the map file (fourth and fifth columns) as :0=missing data, 1=ancestral allele, 2=derived allele. If the latter encoding is detected in the haplotype data, no recoding is

performed. Note that the cross populations statistics such as Rsb and XP-EHH do not need information about ancestral and derived allele status. Finally, the arguments `min_perc_geno.hap`, `min_perc_geno.snp` and `min_maf` are evaluated in this order.

Value

The returned value is an object of class haplohh

References

Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet*, **78**, 629-644.

O'Connell J, Gurdasani D, Delaneau O, et al (2014) A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet*, **10**, e1004234.

See Also

calc_ehh,calc_ehhs,scan_hh,make.example.files

Examples

```
#Copy example files in the current working directory.
make.example.files()
#using the fastPHASE output haplotype example file
hap<-data2haplohh(hap_file="bta12_hapguess_switch.out",map_file="map.inp",
min_maf=0.05,popsel=7,chr.name=12,recode.allele=TRUE)
#using the standard output haplotype example file
hap<-data2haplohh(hap_file="bta12_cgu.hap",map_file="map.inp",
min_maf=0.05,chr.name=12,recode.allele=TRUE)
```

---

distribplot *Distribution of standardized iHS, Rsb or XP-EHH values*

Description

Plot the observed distribution of standardized iHS, Rsb or XP-EHH values together with the expected standard Gaussian distribution

Usage

```
distribplot(data, lty = 1, lwd = 1.5,col = c("blue", "red"),
            main = "Genome-wide distribution", xlab = "", cex.main = 1.5,
            cex.lab = 1.25, qqplot = TRUE)
```

Arguments

data	A vector of iHS, Rsb or XPEHH values.
col	A vector describing color of the Observed and expected Gaussian distribution
main	Character string for the plot legend
xlab	Character string for the X-axis legend
cex.lab	Size of axis legends
cex.main	Size of the main legend
lty	Line Type
lwd	Line Width
qqplot	If TRUE a qq-plot is drawn

Value

The function returns a plot.

See Also

scan_hh,ihh2ihs,ies2rsb,ihsplot,rsbplot,ies2xpehh,xpehhplot

Examples

```
data(wgscan.cgu)
## results from a genome scan (44,057 SNPs) see ?wgscan.eut for details
val.ihs<-ihh2ihs(wgscan.cgu)$iHS[,3]
##standardize
distribplot(val.ihs,main="iHS (CGU population)")
dev.off()
```

---

haplohh-class *Class* "haplohh"

Description

An object of class haplohh contains all relevant haplotype information (see below).

Objects from the Class

Objects can be created by calls of the form `new("haplohh", ...)`.

Slots

`haplo` Object of class `"matrix"`: haplotypes with alleles coded as 0 (missing data), 1 (ancestral allele) or 2 (derived allele)

`position` Object of class `"numeric"`:position of the SNPs in the chromosome

`snp.name` Object of class `"character"`:names of the SNP

`chr.name` Object of class `"character"`:name of the chromosome SNPs are mapping to

`nhap` Object of class `"numeric"`:number of haplotypes

`nsnp` Object of class `"numeric"`:number of SNPs in the haplotypes

See Also

data2haplohh

Examples

```
showClass("haplohh")
```

---

haplohh_cgu_bta12 *Example of an* haplohh *object*

Description

The object contains haplotype data for 140 cattle individuals (280 haplotypes) belonging to the Creole breed from Guadeloupe (CGU) and 1424 SNPs (mapping to chromosome BTA12).

Usage

```
data(haplohh_cgu_bta12)
```

References

Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, **20**, 3128–3143.

See Also

data2haplohh

ies2rsb *Compute Rsb (standardized ratio of iES between two populations)*

Description

Compute Rsb (standardized ratio of iES between two populations).

Usage

```
ies2rsb(hh_pop1, hh_pop2, popname1 = NA, popname2 = NA,
            method = "bilateral")
```

Arguments

hh_pop1	A matrix with nsnps rows and six columns (Chromosome name, position of the SNP, Frequency of the ancestral allele, iHH for the ancestral allele, iHH for the derived allele and iES) obtained after performing a scan on the first population.
hh_pop2	A matrix with nsnps rows and six columns (Chromosome name, position of the SNP, Frequency of the ancestral allele, iHH for the ancestral allele, iHH for the derived allele and iES) obtained after performing a scan on the second population.
popname1	Name of the first population compared (character string).
popname2	Name of the second population compared (character string).
method	Either "bilateral" or "unilateral". If bilateral (resp. unilateral), the pvalue (assuming Rsb follows a standard Gaussian distribution under neutrality) corresponds to a bilateral (resp. unilateral) tests

Details

Ratio of iES (population 1 over population 2) computed and standardized as described in Tang et al. (2007)

Value

The returned value is a matrix with nsnps rows and four columns (Chromosome name, position of the SNP, Rsb and Pvalue)

References

Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, **20**, 3128–3143.

Tang, K. and Thornton, K.R. and Stoneking, M. (2007). A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *Plos Biology*, **7**, e171.

See Also

calc_ehhs,scan_hh,distribplot,rsbplot

Examples

```
data(wgscan.cgu) ; data(wgscan.eut)
## results from a genome scan (44,057 SNPs)
##see ?wgscan.eut and ?wgscan.cgu for details
res.rsb<-ies2rsb(wgscan.cgu,wgscan.eut,"CGU","EUT")
```

`ies2xpehh` *Compute XP-EHH (standardized ratio of iES from two populations) as described in Sabeti et al. (2007)*

Description

Compute XP-EHH (standardized ratio of iES from two populations) as described in Sabeti et al. (2007).

Usage

```
ies2xpehh(hh_pop1, hh_pop2, popname1 = NA, popname2 = NA,
            method = "bilateral")
```

Arguments

`hh_pop1`	A matrix with nsnps rows and six columns (Chromosome name, position of the SNP, Frequency of the ancestral allele, iHH for the ancestral allele, iHH for the derived allele and iES) obtained after performing a scan on the first population.
`hh_pop2`	A matrix with nsnps rows and six columns (Chromosome name, position of the SNP, Frequency of the ancestral allele, iHH for the ancestral allele, iHH for the derived allele and iES) obtained after performing a scan on the second population.
`popname1`	Name of the first population compared (character string).
`popname2`	Name of the second population compared (character string).
`method`	Either "bilateral" or "unilateral". If bilateral (resp. unilateral), the pvalue (assuming XP-EHH follows a standard Gaussian distribution under neutrality) corresponds to a bilateral (resp. unilateral) tests

Details

Ratio of iES (population 1 over population 2) computed and standardized as described in Sabeti et al. (2007)

Value

The returned value is a matrix with nsnps rows and four columns (Chromosome name, position of the SNP, XP-EHH and Pvalue)

References

Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, **20**, 3128–3143.

Sabeti, P.C. et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, **449**, 913–918.

See Also

calc_ehhs,scan_hh,distribplot,rsbplot

Examples

```
data(wgscan.cgu) ; data(wgscan.eut)
## results from a genome scan (44,057 SNPs)
##see ?wgscan.eut and ?wgscan.cgu for details
xpehh<-ies2xpehh(wgscan.cgu,wgscan.eut,"CGU","EUT")
```

---

ihh2ihs *Compute iHS (standardized iHH)*

Description

Compute iHS (standardized iHH).

Usage

```
ihh2ihs(res_ihh, freqbin = 0.025, minmaf = 0.05)
```

Arguments

`res_ihh`	A dataframe with nsnps rows and seven columns as obtained from the `scan_hh` function applied to the population of interest.
`freqbin`	Size of the bin to standardize log(iHHA/iHHD) according to the underlying "core" allele frequency at the focal SNP. Allele frequency bins are built from `minmaf` to 1-`minmaf` in steps of size `freqbin`. If `freqbin` is set to 0, standardization is performed considering each observed frequency as a discrete frequency class (useful in case of a large number of SNPs and few different haplotypes). If `freqbin` is set to an integer of 1 or greater, a corresponding number of equally sized bins are used.
`minmaf`	SNPs with a MAF (Minor Allele Frequency) lower than `minmaf` are discarded from the analysis

Details

iHS is calculated as described in Voight et al. (2006)

Value

The returned value is a list containing two elements

res.ihs	a dataframe with nsnps rows and four columns (Chromosome name, position of the SNP, iHS and Pvalue in a log10 scale)
summary.class	matrix with nclasses rows and three columns (Number of SNPs belonging to this class,position of the SNP, mean iHH in this class, standard deviation of iHH in this class)

References

Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, **20**, 3128–3143.

Voight, B.F. and Kudaravalli, S. and Wen, X. and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. *Plos Biology*, **4**, e72.

See Also

calc_ehh,scan_hh,distribplot,ihsplot

Examples

```
data(wgscan.cgu)
## results from a genome scan (44,057 SNPs)
##see ?wgscan.eut and ?wgscan.cgu for details
res.ihs<-ihh2ihs(wgscan.cgu)
```

---

ihsplot *Plot iHS over a genome*

Description

Plot iHS over a genome.

Usage

```
ihsplot(ihsdata, plot.pval = TRUE, ylim.scan = 2, pch = 16, cex = 0.5,
            cex.lab = 1.25, main = NA, cex.main = 1.5, cex.axis=1.)
```

Arguments

ihsdata	A list obtained with the ihh2ihs function.
plot.pval	Either TRUE or FALSE if Pvalue should not be plotted
ylim.scan	An horizontal line is added at the corresponding coordinate, for instance to represent a significance threshold
pch	Type of the points representing SNPs in the plot(s)

cex	Size of the points representing SNPs in the plot(s)
cex.lab	Size of axis legends
main	Main Legend of the plot
cex.main	Size of the main legend
cex.axis	Size of the axis annotations

Value

The function returns a plot

References

Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, **20**, 3128–3143.

Voight, B.F. and Kudaravalli, S. and Wen, X. and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. *Plos Biology*, **4**, e72.

See Also

ihh2ihs

Examples

```
data(wgscan.cgu)
## results from a genome scan (44,057 SNPs)
##see ?wgscan.eut and ?wgscan.cgu for details
res.ihs<-ihh2ihs(wgscan.cgu)
ihsplot(res.ihs)
```

---

make.example.files *Creating example input files*

Description

This function copies the following example files to the working directory:

- bta12_cgu.hap an haplotype input file in standard format
- bta12_cgu.thap an haplotype input file in transposed format
- bta12_hapguess_switch.out an haplotype input file in fastphase output format
- map.inp a SNP information input file

These files contain data for 280 haplotypes (originating from 140 individuals belonging to the Creole cattle breed from Guadeloupe) of 1,424 SNPs mapping to bovine chromosome 12 (BTA12) (see reference below).

Usage

```
make.example.files()
```

References

Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, **20**, 3128–3143.

See Also

data2haplohh

Examples

```
make.example.files()
```

---

rsbplot *Plot Rsb over a genome*

Description

Plot Rsb over a genome.

Usage

```
rsbplot(data,plot.pval = TRUE, ylim.scan = 2, pch = 16, cex = 0.5,
            cex.lab = 1.25, main = NA, cex.main = 1.5, cex.axis=1.)
```

Arguments

data	A dataframe obtained using ies2rsb function.
plot.pval	Either TRUE or FALSE if Pvalue should not be plotted
ylim.scan	An horizontal line is added at the corresponding coordinate, for instance to represent a significance threshold
pch	Type of the points representing SNPs in the plot(s)
cex	Size of the points representing SNPs in the plot(s)
cex.lab	Size of axis legends
main	Main Legend of the plot
cex.main	Size of the main legend
cex.axis	Size of the axis annotations

Value

The function returns a plot

References

Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, **20**, 3128–3143.

Tang, K. and Thornton, K.R. and Stoneking, M. (2007). A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *Plos Biology*, **7**, e171.

See Also

ies2rsb

Examples

```
data(wgscan.cgu) ; data(wgscan.eut)
## results from a genome scan (44,057 SNPs)
#see ?wgscan.eut and ?wgscan.cgu for details
res.rsb<-ies2rsb(wgscan.cgu,wgscan.eut,"CGU","EUT")
rsbplot(res.rsb)
```

---

scan_hh *Computing EHH based statistics over a whole chromosome*

Description

Compute Extended Haplotype Homozygosity (EHH), site-specific EHH (EHHS), integrated EHH (iHH) and integrated EHHS (iES) for all SNPs of a chromosome (or linkage group).

Usage

```
scan_hh(haplohh, limhaplo = 2, limehh = 0.05, limehhs = 0.05,
        scalegap = NA, maxgap = NA,
        discard_integration_at_border = TRUE, threads = 1)
```

Arguments

haplohh	An object of class haplohh (see data2haplohh).
limhaplo	Minimal number of haplotypes to continue computing EHH away from the core SNP. Useless, if no missing data. However, when some data are missing, haplotypes with missing data are removed from the computation. Hence as we compute EHH further from the core SNP, less haplotypes are expected
limehh	Limit at which EHH stops to be evaluated
limehhs	Limit at which EHHS stops to be evaluated
scalegap	Scales gaps larger than the specified size to the specified size (default=NA, i.e. no scaling)
maxgap	Maximum allowed gap in bp between two SNPs below which EHH and EHHS stop to be evaluated (default=NA, i.e., no limitation)

`discard_integration_at_border`
: If TRUE and if first or last marker or a gap (larger than `maxgap`) is reached and EHH(S) is greater than `limehh(s)`, then iHH/IES is set to `NA`

`threads`
: Number of threads to parallelize compuation

Details

Extended Haplotype Homozygosity (EHH), site-specific EHH (EHHS), integrated EHH (iHH) and integrated EHHS (iES) are computed for all SNPs of the chromosome (or linkage group). This function is several times faster as a procedure calling in turn `calc_ehh` and `calc_ehhs` for all SNPs. To perform a whole genome-scan this function needs to be called for each chromosome and the results concatenated.

Value

The returned value is a dataframe with `haplohh@nsnps` rows and seven columns (Chromosome name, position of the SNP, Frequency of the ancestral allele, iHH for the ancestral allele, iHH for the derived allele, iES using the estimator by Sabeti et al. (2007) estimator and iES using the estimator by Tang et al. (2007))

References

Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, **20**, 3128–3143.

Sabeti, P.C. et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**, 832–837.

Sabeti, P.C. et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, **449**, 913–918.

Tang, K. and Thornton, K.R. and Stoneking, M. (2007). A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *Plos Biology*, **7**, e171.

Voight, B.F. and Kudaravalli, S. and Wen, X. and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. *Plos Biology*, **4**, e72.

See Also

calc_ehh,calc_ehhs,data2haplohh,ihh2ihs,ies2rsb

Examples

```
#example haplohh object (280 haplotypes, 1424 SNPs)
#see ?haplohh_cgu_bta12 for details
data(haplohh_cgu_bta12)
res.scan<-scan_hh(haplohh_cgu_bta12)
```

`xpehhplot` *Plot XP-EHH over a genome*

Description

Plot XP-EHH over a genome.

Usage

```
xpehhplot(data, plot.pval = TRUE, ylim.scan = 2, pch = 16,
            cex = 0.5, cex.lab = 1.25, main = NA, cex.main = 1.5,
            cex.axis=1.)
```

Arguments

`data`	A dataframe obtained using `ies2xpehh` function.
`plot.pval`	Either TRUE or FALSE if Pvalue should not be plotted
`ylim.scan`	An horizontal line is added at the corresponding coordinate, for instance to represent a significance threshold
`pch`	Type of the points representing SNPs in the plot(s)
`cex`	Size of the points representing SNPs in the plot(s)
`cex.lab`	Size of axis legends
`main`	Main Legend of the plot
`cex.main`	Size of the main legend
`cex.axis`	Size of the axis annotations

Value

The function returns a plot

References

Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, **20**, 3128–3143.

Sabeti, P.C. et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, **449**, 913–918.

See Also

ies2xpehh

Examples

```
data(wgscan.cgu) ; data(wgscan.eut)
## results from a genome scan (44,057 SNPs)
#see ?wgscan.eut and ?wgscan.cgu for details
res.xpehh<-ies2xpehh(wgscan.cgu,wgscan.eut,"CGU","EUT")
xpehhplot(res.xpehh)
```

Index