Package 'DSPRqtl'

June 7, 2013

Maintainer Elizabeth King <egking@uci.edu>

License GPL-2

Title Analysis of DSPR phenotypes

LazyData yes

Type Package

LazyLoad yes

Version 2.0-1

Author Elizabeth King

Description Package to analyze phenotypes measured on the Drosophila Synthetic Population Resource

Depends R (>= 2.13)

Suggests DSPRqtlDataA, DSPRqtlDataB

Date 2013-06-06

Collate
'DSPRgenos.R' 'DSPRpeaks.R' 'DSPRscan.R' 'entropy.pos.R''findCI.R' 'findQTL.R' 'founderNs.R' 'geno.means.R' package.R' 'DSPRperm.R' 'LocalInt.R''DSPRplot.R'

R topics documented:

`DSPRqtl-package` *DSPRqtl*

Description

Analysis of DSPR phenotypes

Details

Package:	DSPRqtl
Type:	Package
Version:	2.0-1
Date:	2013-06-06
License:	GPL-2
LazyLoad:	yes
LazyData:	yes

Author(s)

Elizabeth King

`ADHdata` *ADH data*

Description

The ADH activity phenotype data.

Format

A data frame with 3 variables.

`patRIL` numeric vector

`matRIL` numeric vector

`adh` numeric vector

ADHscan *ADH scan*

Description

The ADH genome scan results.

Format

A list with 4 variables.

`LODscores` data.frame containing positions and LODscores

`model` formula

`design` character vector

`phenotype` data.frame containing RIL ids and phenotypes

DSPRgenos *DSPR genotype probabilities*

Description

Function to generate genotype probabilities across the genome for a given DSPR dataset. This function is most useful for those wishing to fit their own models.

Usage

```
DSPRgenos(design, phenotype.dat, id.col, output = "list")
```

Arguments

`design`	a character string. For inbred RIL designs: 'inbredA', 'inbredB'. For cross designs: AAcross, BBcross, or 'ABcross'. A and B refer to the pA and pB set of DSPR RILs.
`phenotype.dat`	`data.frame` containing phenotype data. For inbred designs, there must be a column of numeric RIL ids (must be named patRIL). For cross designs, there must be both a patRIL and matRIL column specifying the maternal and paternal RIL ids. Cross designs also require a sex column for correct specification of the genotypes on the X chromosome.
`id.col`	a character string identifying the name of the column containing unique ids for the samples. e.g. for an inbred design, the patRIL column can be used as the id.
`output`	a character string identifying the preferred format for the output of genolist. Options are 'list' or 'array'. Default is 'list'.

Value

A list containing:

genolist	a list containing the matrix of additive genotype probabilities at each position in the genome. The list is in the same order as the list of positions described below. Column names are the different DSPR haplotypes and row names are the unique ids provided in id.col. Alternatively, if output='array', the output is a 3 dimensional array [samples,haplotypes,positions]. Haplotypes and samples are named while positions are in the order of the list of positions described below.
positions	a data.frame containing regularly spaced positions (every 10KB) in the genome where the genotype probabilities are calculated. Columns are: chr = chromosome, Ppos = physical position (zero offset), Gpos = genetic position, Gaxis = cummulative genetic position.
phenotype	the phenotype.dat data.frame ordered by the specified id column and in the same order as the genotype information at each position in the genome in the geno list described above

Author(s)

Elizabeth King (<egking@uci.edu>)

DSPRpeaks *Find and Summarize QTL*

Description

DSPRpeaks takes output from DSPRscan. Locates and summarizes QTL peaks.

Usage

```
DSPRpeaks(qtldat, method, threshold, LODdrop, BCIprob)
```

Arguments

qtldat	An object of class gscan. Output from DSPRscan.
threshold	numeric vector of length one consisting of the signficance threshold. Default is 6.8 for inbred designs and 10.1 for the ABcross. Use DSPRperm to get a threshold specific to a given dataset.
method	a character string specifying the method to use for the confidence interval. Options are: 'LODdrop' calculates a LOD drop interval for the drop amount specified, 'BCI' calculates the Bayesian credible interval for the fraction specified, and 'both' calculates and returns both.
LODdrop	numeric vector of length one consisting of the LOD drop to be used when using method 'LODdrop'. Default value is 2 which approximates a 95% confidence interval for the inbred designs. Users of the ABcross design should consider using a larger LOD drop.
BCIprob	numeric vector of length one consisting of the nominal Bayes fraction. Default value is 0.95.

Value

A list of class peaks containing a list for each significant peak, each containing:

threshold	the specified signficance threshold
peak	A single row data.frame with the chromosome, physical position (bp), genetic position (cM) and LOD score for the peak
LODdrop	the specified LOD drop for the support interval
CI	the upper and lower bounds of the confidence interval
founderNs	the number of RILs with each founder genotype at the peak
geno.means	the estimated means and standard errors for each founder genotype
perct.var	the percent variation explained by the QTL
entropy	the proporiton of missing information (entropy)

Author(s)

Elizabeth King (<egking@uci.edu>)

DSPRperm *DSPR permutation test*

Description

DSPRperm performs a permutation test for a DSPR dataset.

Usage

```
DSPRperm(model, design, phenotype.dat, id.col,
  batch = 1000, niter = 1000, alpha = 0.05, sex)
```

Arguments

model	an object of class formula: a symbolic description of the null model to be fitted at each position (e.g., phenotype ~ 1). The genotype effects to be fitted will be added based on design.
design	a character string. One of either 'inbredA', 'inbredB', or 'ABcross' corresponding to the pA and pB set of inbred RILs or the pA-pB cross design. For round robin designs or other cross designs, use the more flexible DSPRgenos and standard model fitting functions in R.
phenotype.dat	data.frame containing phenotype data. For inbred designs, there must be a column of numeric RIL ids (must be named patRIL). For the ABcross design, there must be both a patRIL and matRIL column specifying the pA and pB RIL ids.
id.col	a character string identifying the name of the column containing unique ids for the samples. e.g. for an inbred design, the patRIL column can be used as the id.
batch	A numeric vector of length one specifying the number of positions to be examined at a time. A larger number will use more memory but can be faster. Default is 1000.

niter	A numeric vector of length one specifying the number of permutations to run. Default is 1000.
alpha	The alpha level for the genome-wide signficance threshold. Default is 0.05. Raw maximum LOD scores are also returned and quantile can be used to test other alphas.
sex	a character string (either 'm' or 'f') specifying the sex of the measured individuals. This argument must be supplied for a cross design for correct specification of the genotypes on the X chromosome.

Details

The permutation test can take a very long time to run (over 24 hrs).

Value

A list of class pt containing:

maxLODs	A vector containing the maximum LOD score obtained for each permutation of the data.
alpha	the specified alpha level
threshold	the signficance threshold at the specified alpha level

Author(s)

Elizabeth King (<egking@uci.edu>)

DSPRplot *Genome Scan Plot*

Description

Function to plot genome scan results for the DSPR RILs

Usage

```
DSPRplot(qtldata, threshold, legNames = NULL)
```

Arguments

qtldata	a list of output from DSPRscan with each list element corrsponding to one DSPRscan result. e.g. to plot multple genome scans on a single plot: qtldata <- list(scan1,scan
threshold	numeric vector of length one consisting of the signficance threshold. Default is 6.8 for inbred designs and 10.1 for the ABcross. Use DSPRperm to get a threshold specific to a given dataset. dataset.
legNames	a character vector with names for each DSPRscan result to be plotted. Defaults to the phenotype column names used in DSPRscan.

Author(s)

Elizabeth King (<egking@uci.edu>)

DSPRscan *DSPR Genome Scan*

Description

Function to perform a genome scan for data generated with the DSPR RILs. This function is capable of fitting standard models. Users requiring more flexibility should use the DSPRgenos function along with standard model fitting functions in R.

Usage

```
DSPRscan(model, design, phenotype.dat, id.col,
  batch = 1000, sex)
```

Arguments

model	an object of class formula: a symbolic description of the null model to be fitted at each position (e.g., phenotype ~ 1). The genotype effects to be fitted will be added based on design.
design	a character string. One of either 'inbredA', 'inbredB', or 'ABcross' corresponding to the pA and pB set of inbred RILs or the pA-pB cross design. For round robin designs or other cross designs, use the more flexible DSPRgenos and standard model fitting functions in R.
phenotype.dat	data.frame containing phenotype data. For inbred designs, there must be a column of numeric RIL ids (must be named patRIL). For the ABcross design, there must be both a patRIL and matRIL column specifying the pA and pB RIL ids.
id.col	a character string identifying the name of the column containing unique ids for the samples. e.g. for an inbred design, the patRIL column can be used as the id.
batch	A numeric vector of length one specifying the number of positions to be examined at a time. A larger number will use more memory but can be faster. Default is 1000.
sex	a character string (either 'm' or 'f') specifying the sex of the measured individuals. This argument must be supplied for a cross design for correct specification of the genotypes on the X chromosome.

Value

A list of class gscan containing:

LODscores	A data.frame consisting of the chromosome, physical position (bp), genetic position (cM) and LOD score for each position.
model	the model specification
design	the design specification
phenotype	the phenotype data.frame specified

Author(s)

Elizabeth King (<egking@uci.edu>)

`entropy.pos` *Entropy at a position*

Description

`entropy.pos` calculates the entropy (proportion of missing information) at a given position.

Usage

```
entropy.pos(peakChr, peakPos, design, phenotype.dat,
  id.col, sex)
```

Arguments

`peakChr`	character vector of length one. Must be one of the major chromosome arms in the *Drosophila* genome ('X','2L','2R','3L',or '3R').
`peakPos`	numeric vector of length one. A position in base pairs in the DSPR position list (every 10kb).
`design`	a character string. For inbred RIL designs: 'inbredA', 'inbredB'. For cross designs: AAcross, BBcross, or 'ABcross'. A and B refer to the pA and pB set of DSPR RILs.
`phenotype.dat`	`data.frame` containing phenotype data. For inbred designs, there must be a column of numeric RIL ids (must be named patRIL). For the ABcross design, there must be both a patRIL and matRIL column specifying the pA and pB RIL ids. Cross designs also require a sex column for correct specification of the genotypes on the X chromosome.
`id.col`	a character string identifying the name of the column containing unique ids for the samples. e.g. for an inbred design, the patRIL column can be used as the id.
`sex`	a character string (either 'm' or 'f') specifying the sex of the measured individuals. This argument must be supplied for the AB cross design for correct specification of the genotypes on the X chromosome.

Value

A numeric vector: the entropy at the given position for the set of RILs in the phenotype.dat `data.frame`. In the case of the ABcross, the A and B entropy are calculated separately and both are returned.

Author(s)

Elizabeth King (<egking@uci.edu>)

References

Shannon, C.E. 1984. A mathematical theory of communication. *Bell System Technical Journal* 27(3): 379-423. http://en.wikipedia.org/wiki/Information_theory#Entropy

findCI *Calculate LOD support interval*

Description

findCI calculates the LOD support interval for a given LOD peak.

Usage

```
findCI(peakChr, peakPos, qtldat, method, LODdrop,
  BCIprob)
```

Arguments

peakChr	character vector of length one. Must be one of the major chromosome arms in the *Drosophila* genome ('X','2L','2R','3L',or '3R').
peakPos	numeric vector of length one. A position in base pairs in the DSPR position list (every 10kb).
qtldat	data.frame of chromosome, position, and LOD scores (column names chr,Ppos,Gpos,LOD). List element LODscores from DSPRscan.
method	a character string specifying the method to use for the confidence interval. Options are: 'LODdrop' calculates a LOD drop interval for the drop amount specified, 'BCI' calculates the Bayesian credible interval for the fraction specified, and 'both' calculates and returns both.
LODdrop	numeric vector of length one consisting of the LOD drop to be used when using method 'LODdrop'. Default value is 2 which approximates a 95% confidence interval for the inbred designs. Users of the ABcross design should consider using a larger LOD drop.
BCIprob	numeric vector of length one consisting of the nominal Bayes fraction. Default value is 0.95.

Value

A data.frame with two rows containing the chromosome, physical position (bp), genetic position (cM) and LOD scores corresponding to the lower and upper bound. A list of two data.frames are returned when the method used is 'both'.

Author(s)

Elizabeth King (<egking@uci.edu>)

References

Manichaikul, A., J. Dupuis, S. Sen, and K.W. Broman. 2006. Performance of bootstrap confidence intervals for the location of a quantitative trait locus. *Genetics* 174: 481-489.

findQTL *Find QTL*

Description

findQTL finds peak LOD scores that exceed the given threshold (identifies signficant QTL).

Usage

```
findQTL(qtldat, threshold, design)
```

Arguments

qtldat	data.frame of chromosome, position, and LOD scores (column names chr,Ppos,Gpos,LOD). List element LODscores from DSPRscan.
threshold	numeric vector of length one consisting of the signficance threshold. Default is 6.8 for inbred designs and 10.1 for the ABcross. Use DSPRperm to get a threshold specific to a given dataset.
design	a character string. One of either 'inbredA', 'inbredB', or 'ABcross' corresponding to the pA and pB set of inbred RILs or the pA-pB cross design.

Value

A data.frame consisting of the chromosome, physical position (bp), genetic position (cM) and LOD score for each significant peak.

Author(s)

Elizabeth King (<egking@uci.edu>)

founderNs *Numbers of founder genotypes*

Description

founderNs calculates the number of founder genotypes represented in the DSPR RILs at a given position.

Usage

```
founderNs(peakChr, peakPos, design, phenotype.dat,
  id.col)
```

Arguments

peakChr	character vector of length one. Must be one of the major chromosome arms in the *Drosophila* genome ('X','2L','2R','3L',or '3R').
peakPos	numeric vector of length one. A position in base pairs in the DSPR position list (every 10kb).
design	a character string. For inbred RIL designs: 'inbredA', 'inbredB'. For cross designs: AAcross, BBcross, or 'ABcross'. A and B refer to the pA and pB set of DSPR RILs.
phenotype.dat	data.frame containing phenotype data. For inbred designs, there must be a column of numeric RIL ids (must be named patRIL). For the ABcross design, there must be both a patRIL and matRIL column specifying the pA and pB RIL ids. Cross designs also require a sex column for correct specification of the genotypes on the X chromosome.
id.col	a character string identifying the name of the column containing unique ids for the samples. e.g. for an inbred design, the patRIL column can be used as the id.

Value

A named numeric vector consisting of the numbers of RILs in the phenotype.dat data.frame with each founder genotype at the given position. A RIl is assigned a founder genotype if the HMM probability is over 0.95. The number of RILs with a heterozygous genotype and an uncertain genotype are also returned.

Author(s)

Elizabeth King (<egking@uci.edu>)

---

geno.means *Founder genotype means and standard errors*

Description

geno.means estimates the founder genotype means and standard errors at a given position.

Usage

```
geno.means(peakChr, peakPos, model, design,
  phenotype.dat, id.col, sex)
```

Arguments

peakChr	character vector of length one. Must be one of the major chromosome arms in the *Drosophila* genome ('X','2L','2R','3L',or '3R').
peakPos	numeric vector of length one. A position in base pairs in the DSPR position list (every 10kb).
model	an object of class formula: a symbolic description of the null model to be fitted at each position (e.g., phenotype ~ 1). The genotype effects to be fitted will be added based on design.

design	a character string. One of either 'inbredA', 'inbredB', or 'ABcross' corresponding to the pA and pB set of inbred RILs or the pA-pB cross design. For round robin designs or other cross designs, use the more flexible DSPRgenos and standard model fitting functions in R.
phenotype.dat	data.frame containing phenotype data. For inbred designs, there must be a column of numeric RIL ids (must be named patRIL). For the ABcross design, there must be both a patRIL and matRIL column specifying the pA and pB RIL ids.
id.col	a character string identifying the name of the column containing unique ids for the samples. e.g. for an inbred design, the patRIL column can be used as the id.
sex	a character string (either 'm' or 'f') specifying the sex of the measured individuals. This argument must be supplied for a cross design for correct specification of the genotypes on the X chromosome.

Value

A data.frame of the estimated mean and standard error for each founder genotype. If a covariate is included in the model statement, the estimate will be the founder genotype mean after correcting for the covariate. If using the ABcross design, a list of data.frames for pA and pB are returned.

Author(s)

Elizabeth King (<egking@uci.edu>)

LocalInt *Local Interval Mapping*

Description

LocalInt performs standard interval mapping around a given peak for a range of positions specified by the user.

Usage

```
LocalInt(peakChr, peakPos, range = 100, phenotype.dat,
  pheno.name, design)
```

Arguments

peakChr	character vector of length one. Must be one of the major chromosome arms in the *Drosophila* genome ('X','2L','2R','3L',or '3R').
peakPos	numeric vector of length one. A position in base pairs in the DSPR position list (every 10kb).
range	numeric vector of length one specifying the number of positions to test on either side of the peak (positions are every 10kb). Default is 100.
phenotype.dat	data.frame containing a column of ril ids (must be named patRIL) and phenotypes.
pheno.name	a character string specifying the column name of the phenotype data
design	a character string. One of either 'inbredA' or 'inbredB' corresponding to the pA and pB set of inbred RILs. Other crossing designs may be supported in the future.

Value

A `data.frame` consisting of the chromosome, physical position (bp), genetic position (cM) and LOD score for each position.

Author(s)

Elizabeth King (<egking@uci.edu>)

`perct.var` *Effect size*

Description

`perct.var` calculates the percent of the variance explained by the QTL.

Usage

```
perct.var(peakChr, peakPos, model, design, phenotype.dat,
  id.col, sex)
```

Arguments

`peakChr`	character vector of length one. Must be one of the major chromosome arms in the *Drosophila* genome ('X','2L','2R','3L',or '3R').
`peakPos`	numeric vector of length one. A position in base pairs in the DSPR position list (every 10kb).
`model`	an object of class formula: a symbolic description of the null model to be fitted at each position (e.g., `phenotype ~ 1`). The genotype effects to be fitted will be added based on `design`.
`design`	a character string. One of either 'inbredA', 'inbredB', or 'ABcross' corresponding to the pA and pB set of inbred RILs or the pA-pB cross design. For round robin designs or other cross designs, use the more flexible DSPRgenos and standard model fitting functions in R.
`phenotype.dat`	`data.frame` containing phenotype data. For inbred designs, there must be a column of numeric RIL ids (must be named patRIL). For the ABcross design, there must be both a patRIL and matRIL column specifying the pA and pB RIL ids.
`id.col`	a character string identifying the name of the column containing unique ids for the samples. e.g. for an inbred design, the patRIL column can be used as the id.
`sex`	a character string (either 'm' or 'f') specifying the sex of the measured individuals. This argument must be supplied for a cross design for correct specification of the genotypes on the X chromosome.

Value

A numeric vector of length one: the percent variance explained by the QTL.

Author(s)

Elizabeth King (<egking@uci.edu>)

`positionlist_wgenetic` *Positionlist*

Description

List of regularly spaced positions every 10kb used for data analysis of DSPR data. chr is the chromosome arm, Ppos is the position in base pairs, Gpos is the position in centiMorgans, and Gaxis is used for plotting the entire genome on a single axis.

Format

A data frame with 4 variables.

`chr` character vector of chromosome

`Ppos` numeric vector

`Gpos` numeric vector

`Gaxis` numeric vector

Index