Package 'MapRtools'

February 19, 2021

Title Tools for genetic mapping

Version 0.19

Author Jeffrey B. Endelman

Maintainer Jeffrey Endelman <endelman@wisc.edu>

Description Tools for genetic mapping

Depends R (>= 3.5.0)

License GPL-3

LazyData true

RoxygenNote 7.1.1

Encoding UTF-8

Imports ggplot2, scam, seriation, CVXR, Matrix

Suggests knitr, rmarkdown

VignetteBuilder knitr

R topics documented:

genetic_map *Multi-point estimation of a genetic map*

Description

Multi-point estimation of a genetic map

Usage

```
genetic_map(x, LOD, n.point = 5)
```

Arguments

x	matrix of pairwise map distances (cM) between the marker-bins for one chromosome
LOD	matrix of LOD scores between marker-bins
n.point	Number of points used for estimation

Details

Uses LOD-score weighted least-squares regression method described by Stam (1993). Markers must be binned (e.g., using LDbin) for this function to work properly. Argument n.point controls how many pairwise distances are used in the linear regression. n.point=2 means only adjacent bins; n.point=3 means adjacent bins and bins with one intervening marker, etc. Marker names taken from the rownames attribute of x.

Value

data frame with columns marker,position (in cM)

inverse_map_fn *Inverse map function*

Description

Computes recombination frequency from map distance

Usage

```
inverse_map_fn(x, model)
```

Arguments

x	map distance (cM)
model	Either "Haldane" or "Kosambi"

Value

recombination frequency

LDbin *Create marker bins based on LD*

Description

Create marker bins based on LD

Usage

```
LDbin(geno, r2.thresh = 0.99)
```

Arguments

geno	matrix of haplotype dosages (markers x indiv)
r2.thresh	threshold for binning

Details

Bins are created based on hierarchical clustering with `hclust` and `method='single'`, using $1 - r^2$ as the dissimilarity metric. The argument `r2.thresh` controls the height for cutting the dendrogram to create the bins. The marker with the least missing data for each bin is chosen to represent it.

Value

List containing

bins data frame with two columns: marker,bin

geno genotype matrix for the bins

r2 r2 matrix for the bins

LG *Make linkage groups based on clustering*

Description

Make linkage groups based on clustering

Usage

```
LG(LODmat, thresh = seq(2, 20, by = 2))
```

Arguments

LODmat	matrix of LOD scores for the marker bins
thresh	numeric vector of thresholds for clusterings

Details

If `thresh` is a numeric vector with multiple LOD thresholds, the function returns a plot showing the number of markers per LG. If `thresh` is a single value, the function returns a data frame with the LG assignment for each marker. LGs are numbered from the largest to smallest group.

Value

Either a ggplot2 object or data frame of linkage groups (see Details)

`LGtrim` *Trim a linkage group based on genotype frequencies*

Description

Trim a linkage group based on genotype frequencies

Usage

```
LGtrim(geno, LODmat, thresh)
```

Arguments

`geno`	matrix of haplotype dosages (markers x samples)
`LODmat`	matrix of LOD scores for the markers
`thresh`	numeric vector of thresholds for clusterings

Details

This function should only be run on a single linkage group (to form the linkage groups, use `LG`. If `thresh` is a numeric vector with multiple LOD thresholds, the function returns a plot showing the impact of the threshold on genotype frequencies. If `thresh` is a single value, the function returns a vector of the marker names that are retained. The rownames of `geno` and `LODmat` must match.

Value

Either a ggplot2 object or a vector of marker names (see Details)

`LL` *Log-likelihood for inbred line-derived mapping populations*

Description

Log-likelihood for inbred line-derived mapping populations

Usage

```
LL(r, counts, pop.type)
```

Arguments

`r`	recombination frequency
`counts`	3x3 contingency table for haplotype dosages 0,1,2
`pop.type`	One of the following: "DH","BC","F2","RIL.self","RIL.sib"

Details

The argument `counts` can be constructed using the `table` function for two markers. Genotype coding must represent dosage of a founder haplotype. For BC populations, possible allele dosages are 0,1. For DH and RIL pops, it is 0,2. For F2 pops, it is 0,1,2.

Value

log-likelihood

`map_fn` *Map functions*

Description

Computes cM map distance from recombination frequency

Usage

```
map_fn(r, model)
```

Arguments

`r`	recombination frequency
`model`	Either "Haldane" or "Kosambi"

Value

Map distance in cM

`MLEL` *Max Likelihood Estimation of Linkage*

Description

Max Likelihood Estimation of Linkage

Usage

```
MLEL(geno, pop.type, LOD, n.core = 1)
```

Arguments

`geno`	Matrix of haplotype dosages (markers x indiv)
`pop.type`	One of the following: "DH","BC","F2"
`LOD`	Logical, whether to return LOD (TRUE) or recomb freq (FALSE)
`n.core`	For parallel execution on multiple cores

Details

Can be used to estimate either the LOD score or recombination frequency, depending on the value of `LOD`. Genotype coding must represent dosage of a founder haplotype. For BC populations, possible allele dosages are 0,1. For DH and RIL pops, it is 0,2. For F2 pops, it is 0,1,2.

Value

Matrix with RF or LOD

`order_markers` *Order markers by solving the TSP*

Description

Order markers by solving the TSP

Usage

```
order_markers(x)
```

Arguments

`x`	distance matrix

Details

Uses R package seriation to minimize the distance between adjacent markers. For example, `x` could be a matrix of recombination frequencies or monotone decreasing transformation of LOD scores.

Value

a list containing

path optimized order as a vector of integers

distance sum of adjacent distances

`plot_coverage` *Plot marker coverage of the genome*

Description

Plot marker coverage of the genome

Usage

```
plot_coverage(map, limits = NULL)
```

Arguments

map	data frame with columns chrom & position
limits	optional data frame with columns chrom & position, with the maximum length for each chromosome

Details

If limits not provided, then the maximum values in map are used.

Value

ggplot2 variable

---

plot_genofreq *Plot and filter markers based on genotype frequency vs position*

Description

Plot and filter markers based on genotype frequency vs position

Usage

```
plot_genofreq(geno, thresh = 0.1, span = 0.3)
```

Arguments

geno	haplotype dosage matrix (markers x indiv)
thresh	threshold for removing markers (see Details)
span	parameter to control degree of smoothing for spline (higher = less smooth)

Details

Genotypes should be coded 0,1,2. Markers are removed if their residual to the fitted spline exceeds thresh. Markers are assumed to be ordered. Function designed to be used for one chromosome.

Value

List containing

outliers character vector of marker names

plot ggplot2 variable

plot_haplo *Visualize haplotype dosage*

Description

Visualize haplotype dosage in diploid biparental population from two inbreds

Usage

```
plot_haplo(geno, map)
```

Arguments

geno	matrix of haplotype dosages (markers x indiv)
map	data frame with 3 columns (marker, chrom, position)

Details

Input matrix geno should have rownames attribute that matches marker names in the first column of map.

Value

ggplot object

plot_LD *Plot LD vs distance*

Description

Plot LD vs distance

Usage

```
plot_LD(r2, map, max.pair = 10000, dof = 8)
```

Arguments

r2	squared correlation matrix
map	data frame with 3 columns (marker, chrom, position)
max.pair	maximum number of r2 pairs for the spline
dof	degrees of freedom for the spline

Details

A monotone decreasing, convex spline is fit using R package scam. The input matrix r2 should have rownames attribute that matches marker names in the first column of map.

Value

List containing

plot ggplot object

spline data frame with fitted values for the spline

`plot_square` *Plot square (dis)similarity matrix*

Description

Plot square (dis)similarity matrix

Usage

```
plot_square(data, lims = NULL)
```

Arguments

`data`	squared correlation matrix
`lims`	numeric 3-vector with the low,mid,high points for the colors

Details

Can be used to plot squared correlation, recomb frequency, LOD and more. By default, `lims` equals (0,median,max)

Value

ggplot2 variable

Index