Package 'ldblock'

October 16, 2018

Title data structures for linkage disequilibrium measures in populations

Version 1.10.0

Author VJ Carey <stvjc@channing.harvard.edu>

Description Define data structures for linkage disequilibrium measures in populations.

Suggests RUnit, BiocGenerics, knitr

Imports Matrix, snpStats, erma, VariantAnnotation, GenomeInfoDb, Rsamtools, GO.db, GenomicFiles (>= 1.13.6), BiocGenerics (>= 0.25.1)

Depends R (>= 3.1), methods, Homo.sapiens

Maintainer VJ Carey <stvjc@channing.harvard.edu>

License Artistic-2.0

LazyLoad yes

BiocViews genetics, SNP, GWAS, LinkageDisequilibrium

VignetteBuilder knitr

git_url https://git.bioconductor.org/packages/ldblock

git_branch RELEASE_3_7

git_last_commit a143371

git_last_commit_date 2018-04-30

Date/Publication 2018-10-15

R topics documented:

ldblock-package 2
downloadPopByChr 2
expandSnpSet 3
hmld 4
ldByGene 4
ldstruct-class 5
s3_1kg 6

Index **7**

ldblock-package *data structures for linkage disequilibrium measures in populations*

Description

Define data structures for linkage disequilibrium measures in populations.

Details

The DESCRIPTION file: This package was not yet installed at build time.

Index: This package was not yet installed at build time.

Author(s)

VJ Carey <stvjc@channing.harvard.edu>

Maintainer: VJ Carey <stvjc@channing.harvard.edu>

Examples

```
# see vignette
```

downloadPopByChr *download hapmap resource with LD estimates*

Description

download hapmap resource with LD estimates

Usage

```
downloadPopByChr(chrname = "chr1",
 popname = "CEU",
 urlTemplate = "http://hapmap.ncbi.nlm.nih.gov/downloads/ld_data/2009-02_phaseIII_r2/ld_%%CHRN%%_
 targfolder = Sys.getenv("LDBLOCK_TXTGZ_DIR"))
```

Arguments

chrname	UCSC format tag for chromosome
popname	hapmap three letter code for population, e.g. 'CEU'
urlTemplate	pattern for creating URL given chr and pop
targfolder	destination

Details

delivers HapMap LD data to ‘targfolder‘

Value

just run for side effect of download.file

Examples

```
## Not run:
 downloadPopByChr()

## End(Not run)
```

`expandSnpSet` *Given a set of SNP identifiers, use LD to expand the set to include linked loci*

Description

Given a set of SNP identifiers, use LD to expand the set to include linked loci

Usage

```
expandSnpSet(rsl, lb = 0.8, ldstruct, chrn = "chr17", popn = "CEU",
   txtgzfn = dir(system.file("hapmap", package = "ldblock"), full.names = TRUE))
```

Arguments

`rsl`	input list – SNPs not found in the LD structure are simply returned along with those found, and the expansion list, all combined in a vector
`lb`	lower bound on statistic used to retrieve loci in LD
`ldstruct`	instance of `ldstruct-class`
`chrn`	chromosome identifier
`popn`	population identifier (one of 'CEU', 'MEX', ...)
`txtgzfn`	path to gzipped hapmap file with LD information

Details

direct use of elementwise arithmetic comparison

Value

character vector

Note

As of 2015, it appears that locus names are more informative than addresses for determining SNP identity across resources.

Examples

```
og = Sys.getenv("LDBLOCK_TXTGZ_DIR")
on.exit( Sys.setenv("LDBLOCK_TXTGZ_DIR" = og ) )
Sys.setenv("LDBLOCK_TXTGZ_DIR"=system.file("hapmap", package="ldblock"))
ld17 = hmld(chr="chr17", pop="CEU")
ee = expandSnpSet( ld17@allrs[1:10], ldstruct = ld17 )
```

hmld *import hapmap LD data and create a structure for its management*

Description

import hapmap LD data and create a structure for its management

Usage

```
hmld(hmgztxt, poptag, chrom, genome = "hg19", stat = "Dprime")
```

Arguments

hmgztxt	name of gzipped text file as distributed at hapmap.ncbi.nlm.nih.gov/downloads/ld_data/2009-02_phaseIII_r2/. It will be processed by read.delim.
poptag	heuristic tag identifying population
chrom	heuristic tag for chromosome name
genome	genome tag
stat	statistic to use, "Dprime", "R2", and "LOD" are options

Details

generates a sparse matrix representation of pairwise LD statistics and binds metadata on variant name and position

Value

instance of ldstruct class

Examples

```
getClass("ldstruct")
# see vignette
```

ldByGene *obtain LD statistics in region specified by a gene model*

Description

Obtain LD statistics in region specified by a gene model.

Usage

```
ldByGene(sym = "MMP24",
  vcf = system.file("vcf/c20exch.vcf.gz", package = "gQTLstats"),
  flank = 1000,
  vcfSLS = "NCBI",
  genomeSLS = "hg19",
  stats = "D.prime", depth = 10)
```

Arguments

sym	A standard gene symbol for use with genemodel
vcf	Path to a tabix-indexed VCF file
flank	number of basepairs to flank gene model for search
vcfSLS	seqlevelsStyle (SLS) token for VCF; will be imposed on gene model
genomeSLS	character tag for genome, to be used with readVcf
stats	passed to ld
depth	passed to ld

Value

sparse matrix representation of selected LD statistic, as returned by ld

Examples

```
ld1 = ldByGene(depth=150)
image(ld1[1:200,1:200], col.reg=heat.colors(120), colorkey=TRUE,
 main="SNPs in MMP24 (chr20)")
```

---

ldstruct-class *Class* "ldstruct"

Description

Manage information about LD statistics as reported by HapMap.

Objects from the Class

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

Slots

ldmat: Object of class "dsCMatrix" sparse representation of statistics

chrom: Object of class "character" chromosome tag in UCSC format

genome: Object of class "character" genome tag

allpos: Object of class "numeric" coordinates

poptag: Object of class "character" hapmap founder population tag, 'CEU', 'MEX' etc.

statInUse: Object of class "character" code for statistic retrieved, one of 'Dprime', 'LOD', 'R2'

allrs: Object of class "character" all SNP identifiers, sometimes in affy format

Methods

ldmat signature(x = "ldstruct"): extract sparse matrix

Examples

```
showClass("ldstruct")
```

s3_1kg *Create a URL referencing 1000 genomes content in AWS S3*

Description

Create a URL referencing 1000 genomes content in AWS S3.

Usage

```
s3_1kg(chrnum, tag = "20130502", wrap = function(x) TabixFile(x), tmpl = NULL, dropchr = TRUE)
stack1kg(chrs=as.character(1:22), index = FALSE)
```

Arguments

chrnum	a character string denoting a chromosome, such as '22'
chrs	a vector of chromosome names for extraction from 1000 genomes VCF collection
tag	a character string identifying the version, ignored if tmpl is non-null; valid tag values are the default or "20101123"
wrap	The URL is returned after evaluating wrap on it; default is useful when Tabix indexing is to be used
tmpl	alternate template for full URL, useful if versions prior to 2010 are of interest
dropchr	if TRUE chrnum will have 'chr' removed if present
index	a logical indicating if the vcf index files should be created (for stack1kg)

Details

stack1kg produces a VcfStack instance with references to VCF for 1000 genomes autosomal chrs. S3-resident VCF files with version "v5a.20130502" are used.

Value

by default, a TabixFile instance

Examples

```
s3_1kg("22")
## Not run:
 require(VariantAnnotation)
 scanVcfHeader(s3_1kg("22"))

## End(Not run)
```

Index

∗Topic **classes**
ldstruct-class, 5
∗Topic **models**
downloadPopByChr, 2
expandSnpSet, 3
hmld, 4
ldByGene, 4
s3_1kg, 6
∗Topic **package**
ldblock-package, 2

downloadPopByChr, 2

expandSnpSet, 3

genemodel, *5*

hmld, 4

ld, *5*
ldblock *(*ldblock-package*)*, 2
ldblock-package, 2
ldByGene, 4
ldmat *(*ldstruct-class*)*, 5
ldmat,ldstruct-method *(*ldstruct-class*)*, 5
ldstruct-class, 5

read.delim, *4*
readVcf, *5*

s3_1kg, 6
stack1kg *(*s3_1kg*)*, 6

TabixFile, *6*