

# Package ‘ldblock’

October 18, 2017

**Title** data structures for linkage disequilibrium measures in populations

**Version** 1.6.0

**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

**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

**NeedsCompilation** no

## R topics documented:

ldblock-package . . . . .	2
downloadPopByChr . . . . .	3
expandSnpSet . . . . .	3
hml . . . . .	4
ldByGene . . . . .	5
ldstruct-class . . . . .	6
s3_1kg . . . . .	6
<b>Index</b>	<b>8</b>

---

 ldblock-package

*data structures for linkage disequilibrium measures in populations*


---

## Description

Define data structures for linkage disequilibrium measures in populations.

## Details

The DESCRIPTION file:

```

Package:      ldblock
Title:       data structures for linkage disequilibrium measures in populations
Version:     1.6.0
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
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
  
```

Index of help topics:

downloadPopByChr	download hapmap resource with LD estimates
expandSnpSet	Given a set of SNP identifiers, use LD to expand the set to include linked loci
hmlD	import hapmap LD data and create a structure for its management
ldByGene	obtain LD statistics in region specified by a gene model
ldblock-package	data structures for linkage disequilibrium measures in populations
ldstruct-class	Class "'ldstruct"'
s3_1kg	Create a URL referencing 1000 genomes content in AWS S3

## Author(s)

VJ Carey <stvjc@channing.harvard.edu>

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

## Examples

```
# see vignette
```

---

downloadPopByChr	<i>download hapmap resource with LD estimates</i>
------------------	---

---

**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	<i>Given a set of SNP identifiers, use LD to expand the set to include linked loci</i>
--------------	--

---

**Description**

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

**Usage**

```
expandSnpSet(rs1, lb = 0.8, ldstruct, chrn = "chr17", popn = "CEU",
  txtgzfn = dir(system.file("hapmap", package = "ldblock"), full.names = TRUE))
```

**Arguments**

rs1	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 <code>ldstruct-class</code>
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 <a href="http://hapmap.ncbi.nlm.nih.gov/downloads/ld_data/2009-02_phaseIII_r2/">hapmap.ncbi.nlm.nih.gov/downloads/ld_data/2009-02_phaseIII_r2/</a> . It will be processed by <code>read.delim</code> .
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 <a href="#">genemodel</a>
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 <a href="#">readVcf</a>
stats	passed to <a href="#">ld</a>
depth	passed to <a href="#">ld</a>

**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	<i>Class "ldstruct"</i>
----------------	-------------------------

---

### 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	<i>Create a URL referencing 1000 genomes content in AWS S3</i>
--------	--

---

### Description

Create a URL referencing 1000 genomes content in AWS S3.

### Usage

```
s3_1kg(chrnum, tag = "20130502", wrap = function(x) TabixFile(x), tmp1 = NULL, dropchr = TRUE)
stack1kg(chrs=as.character(1:22))
```

**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 tmp1 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
tmp1	alternate template for full URL, useful if versions prior to 2010 are of interest
dropchr	if TRUE chrnum will have 'chr' removed if present

**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, 6

## \*Topic **models**

downloadPopByChr, 3

expandSnpSet, 3

hml, 4

ldByGene, 5

s3\_1kg, 6

## \*Topic **package**

ldblock-package, 2

downloadPopByChr, 3

expandSnpSet, 3

genemodel, 5

hml, 4

ld, 5

ldblock (ldblock-package), 2

ldblock-package, 2

ldByGene, 5

ldmat (ldstruct-class), 6

ldmat, ldstruct-method (ldstruct-class),  
6

ldstruct-class, 6

read.delim, 4

readVcf, 5

s3\_1kg, 6

stack1kg (s3\_1kg), 6

TabixFile, 7