Package: proxysnps (via r-universe)

May 28, 2026

Title Get proxy SNPs for a SNP in the 1000 Genomes Project

Version 0.0.1

Description This package implements functions to query remote VCF files. You can use it to find proxy SNPs in linkage disequilibrium with SNPs of interest or to calculate allele frequencies in different populations.

Depends R (>= 3.2.1)

License MIT + file LICENSE

LazyData true

URL https://github.com/slowkow/proxysnps

BugReports https://github.com/slowkow/proxysnps/issues

Imports RCurl

Suggests myvariant

RoxygenNote 5.0.1

Config/pak/sysreqs make

Repository https://mrcieu.r-universe.dev

Date/Publication 2018-05-16 13:31:03 UTC

RemoteUrl https://github.com/slowkow/proxysnps

RemoteRef HEAD

RemoteSha 9a9a3158d5895d2f94ab9d01213e92f8c50bf716

Contents

compute_ld *Compute two commonly used linkage disequilibrium statistics.*

Description

Compute R.squared and D.prime for two binary numeric vectors.

Usage

```
compute_ld(x, y)
```

Arguments

x	a numeric vector of ones and zeros
y	a numeric vector of ones and zeros

Details

Find more details here: https://en.wikipedia.org/wiki/Linkage_disequilibrium

Value

A list with two items:

R.squared Squared Pearson correlation coefficient.

D.prime Coefficient of linkage disequilibrium D divided by the theoretical maximum.

Examples

```
compute_ld(c(0,0,0,1,1,1), c(1,1,1,1,0,0))
```

get_proxies *Get proxy SNPs for a SNP at a given genomic position.*

Description

Returns a dataframe with proxy SNPs.

Usage

```
get_proxies(chrom = NA, pos = NA, query = NA, window_size = 1e+05,
  pop = NA)
```

Arguments

chrom	a chromosome name (1-22,X) without "chr"
pos	a positive integer indicating the position of a SNP
window_size	a positive integer indicating the size of the window
pop	the name of a 1000 Genomes population (AMR,AFR,ASN,EUR,...). Set this to NA to use all populations.

Details

Currently, this is hard-coded to access 1000 Genomes phase3 data hosted by Brian Browning (author of BEAGLE):

http://bochet.gcc.biostat.washington.edu/beagle/1000_Genomes_phase3_v5a/

This implementation discards multi-allelic markers that have a "," in the ALT column.

The pop can be any of: ACB, ASW, BEB, CDX, CEU, CHB, CHS, CLM, ESN, FIN, GBR, GIH, GWD, IBS, ITU, JPT, KHV, LWK, MSL, MXL, PEL, PJL, PUR, STU, TSI, YRI. It can also be any super-population: AFR, AMR, EAS, EUR, SAS.

Find more details here: http://www.1000genomes.org/faq/which-populations-are-part-your-study

Value

A dataframe with the following columns:

CHROM Chromosome name, e.g. "1"

POS Position, e.g. 583090

ID Identifier, e.g. "rs11063140"

REF Reference allele, e.g. "A"

ALT Alternative allele, e.g. "G"

MAF Minor allele frequency, e.g. 0.1

R.squared Squared Pearson correlation coefficient, e.g. 1.0

D.prime D prime value, e.g. 1.0

CHOSEN Binary indicator set to TRUE for the SNP of interest

Examples

```
d <- get_proxies(chrom = "12", pos = 583090, window_size = 1e5, pop = "AFR")
head(d)
```

get_vcf *Get data for a genomic region from a remote VCF file.*

Description

Returns a list with three dataframes for individuals, SNPs, and genotypes.

Usage

```
get_vcf(chrom, start, end, pop = NA)
```

Arguments

chrom	a chromosome name (1-22,X) without "chr"
start	a positive integer indicating the start of a genomic region
end	a positive integer indicating the end of a genomic region
pop	the name of a 1000 Genomes population (AMR,AFR,ASN,EUR,...)

Details

Currently, this is hard-coded to access 1000 Genomes phase3 data hosted by Brian Browning (author of BEAGLE):

http://bochet.gcc.biostat.washington.edu/beagle/1000_Genomes_phase3_v5a/

This implementation discards multi-allelic markers that have a "," in the ALT column.

The pop can be any of: ACB, ASW, BEB, CDX, CEU, CHB, CHS, CLM, ESN, FIN, GBR, GIH, GWD, IBS, ITU, JPT, KHV, LWK, MSL, MXL, PEL, PJL, PUR, STU, TSI, YRI. It can also be any super-population: AFR, AMR, EAS, EUR, SAS.

Find more details here: http://www.1000genomes.org/faq/which-populations-are-part-your-study

Value

A list with three dataframes:

ind A dataframe with information about individuals: Family.ID, Individual.ID, Paternal.ID, Maternal.ID, Gender, Population, Relationship, Siblings, Second.Order, Third.Order, Other.Comments, SuperPopulation

meta First 8 columns of the VCF file: CHROM, POS, ID, REF, ALT, QUAL, FILTER, INFO

geno Columns 10 onward of the VCF file. All genotypes are converted to 0s and 1s representing REF and ALT alleles. This dataframe has two columns for each individual.

Examples

```
vcf <- get_vcf(chrom = "12", start = 533090, end = 623090, pop = "AFR")
names(vcf)
```

Index