Package 'RAREsim'

November 12, 2026

Version 1.16.0

Date 2020-08-01

Title Simulation of Rare Variant Genetic Data

Description Haplotype simulations of rare variant genetic data that emulates real data can be performed with RAREsim. RAREsim uses the expected number of variants in MAC bins - either as provided by default parameters or estimated from target data - and an abundance of rare variants as simulated HAPGEN2 to probabilistically prune variants. RAREsim produces haplotypes that emulate real sequencing data with respect to the total number of variants, allele frequency spectrum, haplotype structure, and variant annotation.

URL https://github.com/meganmichelle/RAREsim

BugReports https://github.com/meganmichelle/RAREsim/issues

License GPL-3

LazyData FALSE

biocViews Genetics, Software, VariantAnnotation, Sequencing

RoxygenNote 7.1.2

Encoding UTF-8

VignetteBuilder knitr

Depends R (>= 4.1.0)

Imports nloptr

Suggests markdown, ggplot2, BiocStyle, rmarkdown, knitr, testthat (>= 3.0.0)

Config/testthat/edition 3

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

git_branch RELEASE_3_23

git_last_commit a37e532

git_last_commit_date 2026-04-28

Repository Bioconductor 3.23

Date/Publication 2026-11-11

Author Megan Null [aut],
Ryan Barnard [cre]

Maintainer Ryan Barnard <rbarnard1107@gmail.com>

Contents

afs 2
afs_afr 3
expected_variants 3
fit_afs 4
fit_nvariant 5
nvariant 5
nvariant_afr 6

Index **7**

`afs` *afs function - Calculates proportion of variants per rare MAC bins*

Description

The afs function calculates the proportion of variants in minor allele count (MAC) bins given parameters, as described in RAREsim.

Usage

```
afs(alpha = NULL, beta = NULL, b = NULL, mac_bins, pop = NULL)
```

Arguments

`alpha`	AFS function parameter alpha, does not need to be specified if default parameters are used
`beta`	AFS function parameter beta, does not need to be specified if default parameters are used
`b`	AFS function parameter b, does not need to be specified if default parameters are used
`mac_bins`	The rare MAC bins to use, with Lower and Upper boundaries defined
`pop`	The population: AFR, EAS, NFE or SAS - specified when using default parameters

Details

The default parameters will be used if an ancestrial population is specified: pop = 'AFR', 'EAS', 'NFE', or 'SAS'. Otherwise, the parameters alpha, beta, and b need to be provided. Alpha, beta, and b can be estimated from target data using the *Fit_afs* function. The MAC bins should be exhaustive, non-overlapping bins of rare allele counts with column names Lower and Upper.

Value

data frame with the MAC bins provided and proportion of variants in each bin

Examples

```
data('afs_afr')
mac <- afs_afr[,c(1:2)]
afs(mac_bins = mac, pop = 'AFR')
afs(alpha = 1.594622, beta =  -0.2846474, b  = 0.297495, mac_bins = mac)
```

---

afs_afr *African/African American target data from gnomAD v2.1 (Karczewski, 2020)*

Description

African/African American target data from gnomAD v2.1 (Karczewski, 2020)

Usage

```
data(afs_afr)
```

Format

A data frame for the Non-Finnish European AFS target data. The first two columns define the MAC Bin boundaries. The third column is the proportion of variants in that bin.

Details

Used to fit the afs function with *fit_afs*

---

expected_variants *Combines Number of Variants and AFS functions*

Description

This function combines the Number of Variants and AFS functions to produces the expected number of variants per Kb in each MAC bin

Usage

```
expected_variants(Total_num_var, mac_bin_prop)
```

Arguments

Total_num_var	estimated total number of variants in the region of interest
mac_bin_prop	The MAC bins to use, with three columns: Lower, Upper, and Prop. Lower and Upper define the MAC bins boundaries and Prop is the proportion of variants in each respective bin. Only define for rare variants

Value

data frame with the MAC bins and expected variants

Examples

```
data('afs_afr')
mac <- afs_afr[,c(1:2)]
expected_variants(Total_num_var = 19.029*nvariant(pop='AFR', N = 8128),
mac_bin_prop = afs(mac_bins = mac, pop = 'AFR'))
```

---

fit_afs *Given target data, fit the AFS function*

Description

This function takes AFS target data and estimates parameters for the AFS function A dataframe specifying the rare MAC bins and the observed proportion of variants is used to fit the data The proportion of rare variants (p_rv) is by default the sum of the rare allele count bins. The proportion be manually specified if desired

Usage

```
fit_afs(Observed_bin_props, p_rv = NULL)
```

Arguments

Observed_bin_props	data frame with 3 columns, Lower, Upper (of MAC bins) and proportion of variants in that MAC bin
p_rv	proportion of rare variants - default is the sum of the rare MAC bin proportions

Value

list of parameters - alpha, beta, and b as well as fitted proportions

Examples

```
data("afs_afr")
fit_afs(Observed_bin_props = afs_afr)
```

fit_nvariant *Given target data, fit the Number of Variants function*

Description

This function takes Number of Variants target data and estimates parameters for the Number of Variants function. A dataframe specifying the number of variants per Kb at various sample sizes is required to fit the data

Usage

```
fit_nvariant(Observed_variants_per_kb)
```

Arguments

Observed_variants_per_kb	A data frame with the first column sample size and the second variants per Kb, both numeric

Value

Vector of parameters - phi and omega

Examples

```
data("nvariant_afr")
fit_nvariant(nvariant_afr)
```

nvariant *Number of Variants function*

Description

The Number of Variants (nvariant) function calculates the number of variants per kilobase, as described in RAREsim. N is the number of individuals. The Number of Variants function changes with N. The default parameters will be used if an ancestrial population is specified: (AFR, EAS, NFE, or SAS) is specified. Otherwise, the parameters phi and omega need to be provided. Phi and omega can be estimated from target data using the Fit_nvariant function.

Usage

```
nvariant(phi = NULL, omega = NULL, N, pop = NULL)
```

Arguments

phi	parameter phi
omega	parameter omega
N	sample size in number of individuals
pop	population - only needs to be specified if using default parameters

Value

the number of variants per kb

Examples

```
nvariant(N = 8128, pop = 'AFR')
nvariant(phi = 0.1638108, omega = 0.6248848, N = 8128)
```

---

nvariant_afr *African target data for the Number of Variants function from gnomADv2.1 (Karczewski, 2020)*

Description

African target data for the Number of Variants function from gnomADv2.1 (Karczewski, 2020)

Usage

```
data(nvariant_afr)
```

Format

Number of Variants target data. First column is sample size and second column is number of variants per Kb

Index

∗ **RAREsim**
    nvariant, 5
∗ **datasets**
    afs_afr, 3
    nvariant_afr, 6

afs, 2
afs_afr, 3

expected_variants, 3

fit_afs, 4
fit_nvariant, 5

nvariant, 5
nvariant_afr, 6