# Package 'inbreedR'

May 8, 2026

**Title** Analysing Inbreeding Based on Genetic Markers

**Version** 0.3.3

**Description** A framework for analysing inbreeding and heterozygosity-fitness correlations (HFCs) based on microsatellite and SNP markers.

**Depends** R (>= 3.2.1)

**License** GPL-2

**LazyData** true

**Imports** data.table (>= 1.9.6), parallel, stats, graphics

**Suggests** testthat (>= 0.10.0), knitr, rmarkdown, covr

**VignetteBuilder** knitr

**RoxygenNote** 6.0.1

**NeedsCompilation** no

**Author** Martin A. Stoffel [aut, cre],
Mareike Esser [aut],
Joseph Hoffman [aut],
Marty Kardos [aut]

**Maintainer** Martin A. Stoffel <martin.adam.stoffel@gmail.com>

**Repository** CRAN

**Date/Publication** 2022-02-02 11:00:07 UTC

## Contents

---

`bodyweight` *Oldfield mouse bodyweight data*

---

## Description

Bodyweight data for 36 oldfield mice.

## Format

A vector with 36 elements.

## References

Hoffman, J.I., Simpson, F., David, P., Rijks, J.M., Kuiken, T., Thorne, M.A.S., Lacey, R.C. & Dasmahapatra, K.K. (2014) High-throughput sequencing reveals inbreeding depression in a natural population. Proceedings of the National Academy of Sciences of the United States of America, 111: 3775-3780. Doi: 10.1073/pnas.1318945111

---

`check_data` *Checks the data for consistency with the inbreedR working format.*

---

## Description

The `inbreedR` working format is an i * l genotype matrix, whereby each individual is a row and each column is a locus. Heterozygosity at a given locus should be coded as `1`, homozygosity as `0` and missing values should be coded as `NA`.

## Usage

```
check_data(genotypes, num_ind = NULL, num_loci = NULL)
```

## Arguments

| | |
|---|---|
| `genotypes` | `data.frame` (or `matrix`) with individuals in rows and loci in columns, containing genotypes coded as `0` (homozygote), `1` (heterozygote) and `NA` (missing) |
| `num_ind` | Number of individuals |
| `num_loci` | Number of loci / markers |

### Details

Checks that (1) the genotype data just contains 3 elements, which is 0 for homozygote, 1 for heterozygote and NA for missing data, (2) the number of individuals corresponds to the number of rows and the number of loci corresponds to the number of columns, (3) the data type is numeric. .

### Value

TRUE if the data format is correct, error message if any test failed

### Author(s)

Martin A. Stoffel (martin.adam.stoffel@gmail.com)

### Examples

```
data(mouse_msats)
# tranform raw genotypes into 0/1 format
genotypes <- convert_raw(mouse_msats)
# check data
check_data(genotypes, num_ind = 36, num_loci = 12)
```

---

## convert_raw *Genotype format converter*

---

### Description

Turns raw genotype data into 0 (homozygote), 1 (heterozygote) and NA (missing), which is the working format for the inbreedR functions. A raw genotype matrix has individuals in rows and each locus in two adjacent columns. Individual ID's can be rownames. Type data(mouse_msats) for an example raw genotype data frame.

### Usage

```
convert_raw(genotypes)
```

### Arguments

| | |
|---|---|
| genotypes | Raw genotype data.frame or matrix. Rows represent individuals and each locus has two adjacent columns. Alleles within loci can be coded as numbers (e.g. microsatellite length) or characters (e.g. "A", "T") See data(mouse_msat) for an example. Missing values should be coded as NA. |

### Value

data.frame object with 0 (homozygote), 1 (heterozygote) and NA (missing data). Each locus is a column and each individual is a row.

## Author(s)

Martin Stoffel (martin.adam.stoffel@gmail.com)

## Examples

```
# Mouse microsatellite data with missing values coded as NA
data(mouse_msats)
genotypes <- convert_raw(mouse_msats)
head(genotypes)
```

---

`g2_microsats` *Estimating g2 from microsatellite data*

---

## Description

Estimating g2 from microsatellite data

## Usage

```
g2_microsats(genotypes, nperm = 0, nboot = 0, boot_over = "inds",
  CI = 0.95, verbose = TRUE)
```

## Arguments

| | |
|---|---|
| `genotypes` | `data.frame` with individuals in rows and loci in columns, containing genotypes coded as 0 (homozygote), 1 (heterozygote) and `NA` (missing) |
| `nperm` | Number of permutations for testing the hypothesis that the empirical g2-value is higher than the g2 for random associations between individuals and genotypes. |
| `nboot` | Number of bootstraps for estimating a confidence interval |
| `boot_over` | Bootstrap over individuals by specifying "inds" and over loci with "loci". Defaults to "ind". |
| `CI` | Confidence interval (default to 0.95) |
| `verbose` | If FALSE, nothing will be printed to show the status of bootstraps and permutations. |

## Details

Calculates g2 from smaller datasets. The underlying formula is compationally expensive due to double summations over all paits of loci (see David et al. 2007). Use convert_raw to convert raw genotypes (with 2 columns per locus) into the required format.

## Value

g2_microsats returns an object of class "inbreed". The functions ‘print‘ and ‘plot‘ are used to print a summary and to plot the distribution of bootstrapped g2 values and CI.

An ‘inbreed‘ object from `g2_microsats` is a list containing the following components:

| | |
|---|---|
| `call` | function call. |
| `g2` | g2 value |
| `p_val` | p value from permutation test |
| `g2_permut` | g2 values from permuted genotypes |
| `g2_boot` | g2 values from bootstrap samples |
| `CI_boot` | confidence interval from bootstraps |
| `se_boot` | standard error of g2 from bootstraps |
| `nobs` | number of observations |
| `nloc` | number of markers |

## Author(s)

Martin A. Stoffel (martin.adam.stoffel@gmail.com) & Mareike Esser (messer@techfak.uni-bielefeld.de)

## References

David, P., Pujol, B., Viard, F., Castella, V. and Goudet, J. (2007), Reliable selfing rate estimates from imperfect population genetic data. Molecular Ecology, 16: 2474

## Examples

```
data(mouse_msats)
# tranform raw genotypes into 0/1 format
genotypes <- convert_raw(mouse_msats)
(g2_mouse <- g2_microsats(genotypes, nperm = 1000, nboot = 100, boot_over = "inds", CI = 0.95))
```

---

# g2_snps *Estimating g2 from larger datasets, such as SNPs*

---

## Description

Estimating g2 from larger datasets, such as SNPs

## Usage

```
g2_snps(genotypes, nperm = 0, nboot = 0, boot_over = "inds", CI = 0.95,
  parallel = FALSE, ncores = NULL, verbose = TRUE)
```

## Arguments

| | |
|---|---|
| genotypes | data.frame with individuals in rows and loci in columns, containing genotypes coded as 0 (homozygote), 1 (heterozygote) and NA (missing) |
| nperm | number or permutations for to estimate a p-value |
| nboot | number of bootstraps to estimate a confidence interval |
| boot_over | Bootstrap over individuals by specifying "inds" and over loci with "loci". Defaults to "ind". |
| CI | confidence interval (default to 0.95) |
| parallel | Default is FALSE. If TRUE, bootstrapping and permutation tests are parallelized |
| ncores | Specify number of cores to use for parallelization. By default, all available cores are used. |
| verbose | If FALSE, nothing will be printed to show the status of bootstraps and permutations. |

## Details

Calculates g2 from SNP datasets. Use convert_raw to convert raw genotypes (with 2 columns per locus) into the required format

## Value

g2_snps returns an object of class "inbreed". The functions ‘print‘ and ‘plot‘ are used to print a summary and to plot the distribution of bootstrapped g2 values and CI.

An ‘inbreed‘ object from g2_snps is a list containing the following components:

| | |
|---|---|
| call | function call. |
| g2 | g2 value |
| p_val | p value from permutation test |
| g2_permut | g2 values from permuted genotypes |
| g2_boot | g2 values from bootstrap samples |
| CI_boot | confidence interval from bootstrap distribution |
| se_boot | standard error of g2 from bootstraps |
| nobs | number of observations |
| nloc | number of markers |

## Author(s)

Martin A. Stoffel (martin.adam.stoffel@gmail.com) & Mareike Esser (messer@techfak.uni-bielefeld.de)

## References

Hoffman, J.I., Simpson, F., David, P., Rijks, J.M., Kuiken, T., Thorne, M.A.S., Lacey, R.C. & Dasmahapatra, K.K. (2014) High-throughput sequencing reveals inbreeding depression in a natural population. Proceedings of the National Academy of Sciences of the United States of America, 111: 3775-3780. Doi: 10.1073/pnas.1318945111

## Examples

```
# load SNP genotypes in 0 (homozygous), 1 (heterozygous), NA (missing) format.
# low number of bootstraps and permutations for computational reasons.
data(mouse_snps)
(g2_mouse <- g2_snps(mouse_snps, nperm = 10, nboot = 10, CI = 0.95, boot_over = "loci"))

# parallelized version for more bootstraps or permutations
## Not run:
(g2_mouse <- g2_snps(mouse_snps, nperm = 1000, nboot = 1000,
                     CI = 0.95, parallel = TRUE, ncores = 4))

## End(Not run)
```

---

# HHC *Calculates heterzygosity-heterozygosity correlations with standardized multilocus heterozygosities (sMLH)*

---

## Description

Loci are randomly devided into two equal groups and the correlation coefficient between the resulting sMLH values is calculated.

## Usage

```
HHC(genotypes, reps = 100, CI = 0.95)
```

## Arguments

| | |
|---|---|
| genotypes | data.frame with individuals in rows and loci in columns, containing genotypes coded as 0 (homozygote), 1 (heterozygote) and NA (missing) |
| reps | number of repetitions, i.e. splittings of the dataset |
| CI | size of the confidence interval around the mean het-het correlation (default is 0.95) |

## Value

| | |
|---|---|
| call | function call. |
| HHC_vals | vector of HHC's obtained by randomly splitting the dataset |
| summary_exp_r2 | r2 mean and sd for each number of subsetted loci |
| nobs | number of observations |
| nloc | number of markers |

## Author(s)

Martin A. Stoffel (martin.adam.stoffel@gmail.com)

## References

Balloux, F., Amos, W., & Coulson, T. (2004). Does heterozygosity estimate inbreeding in real populations?. Molecular Ecology, 13(10), 3021-3031.

## Examples

```
data(mouse_msats)
genotypes <- convert_raw(mouse_msats)
(out <- HHC(genotypes, reps = 100, CI = 0.95))
```

---

`inbreedR` *inbreedR: Workflows for analysing variance in inbreeding and HFCs based on SNP or microsatellite markers.*

---

## Description

`inbreedR` contains the following functions:

g2_microsats g2_snps convert_raw check_data r2_hf r2_Wf HHC sMLH MLH simulate_g2 simulate_r2_hf plot.inbreed print.inbreed

## Details

A correlation between heterozygosity (h) and fitness (W) requires a simultaneous effect of inbreeding level (f) on both of them. A heterozygosity-fitness correlation (HFC) thus is the product of two correlations, which can be summarized in the following equation:

$$r(W,h) = r(W,f)r(h,f)$$

Estimating these parameters and their sensitivity towards the number and type of genetic markers used is the central framework of the inbreedR package. At the heart of measuring inbreeding based on genetic markers is the g2 statistic, which estimates the correlation of heterozygosity across markers, called identity disequilibrium (ID). ID is a proxy for inbreeding.

The package has three main goals:

- Assessing identity disequilibria and the potential to detect heterozygosity-fitness correlations
- Providing insights on the sensitivity of these measures based on the number/type of molecular markers used
- Implementing computationally efficient functions in a flexible environment for analysing inbreeding and HFC's with both small and large datasets.

For a short introduction to inbreedR start with the vignette: `browseVignettes(package = "inbreedR")`

## Author(s)

Martin Stoffel (martin.adam.stoffel@gmail.com), Mareike Esser (messer@uni-bielefeld.de)

## References

Slate, J., David, P., Dodds, K. G., Veenvliet, B. A., Glass, B. C., Broad, T. E., & McEwan, J. C. (2004). Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Heredity, 93(3), 255-265.

Szulkin, M., Bierne, N., & David, P. (2010). HETEROZYGOSITY-FITNESS CORRELATIONS: A TIME FOR REAPPRAISAL. Evolution, 64(5), 1202-1217.

David, P., Pujol, B., Viard, F., Castella, V. and Goudet, J. (2007), Reliable selfing rate estimates from imperfect population genetic data. Molecular Ecology, 16: 2474

Hoffman, J.I., Simpson, F., David, P., Rijks, J.M., Kuiken, T., Thorne, M.A.S., Lacey, R.C. & Dasmahapatra, K.K. (2014) High-throughput sequencing reveals inbreeding depression in a natural population. Proceedings of the National Academy of Sciences of the United States of America, 111: 3775-3780.

---

`MLH` *Calculate multilocus heterozygosity (MLH)*

---

## Description

MLH is defined as the total number of heterozygous loci in an individual divided by the number of loci typed in the focal individual. An MLH of 0.5 thus means that 50 percent of an indiviudals loci are heterozygous.

## Usage

```
MLH(genotypes)
```

## Arguments

| | |
|---|---|
| `genotypes` | data.frame with individuals in rows and loci in columns, containing genotypes coded as 0 (homozygote), 1 (heterozygote) and NA (missing) |

## Value

Vector of individual multilocus heterozygosities

## Author(s)

Martin A. Stoffel (martin.adam.stoffel@gmail.com)

## References

Coltman, D. W., Pilkington, J. G., Smith, J. A., & Pemberton, J. M. (1999). Parasite-mediated selection against inbred Soay sheep in a free-living, island population. Evolution, 1259-1267.

## Examples

```
data(mouse_msats)
genotypes <- convert_raw(mouse_msats)
het <- MLH(genotypes)
```

---

`mouse_msats` *Oldfield mouse microsatellite data*

---

## Description

Dataset with each microsatellite locus in two adjecent columns (one per allel). Missing values are coded as `NA`.

## Format

A data frame with 36 observations at 13198 loci.

## References

Hoffman, J.I., Simpson, F., David, P., Rijks, J.M., Kuiken, T., Thorne, M.A.S., Lacey, R.C. & Dasmahapatra, K.K. (2014) High-throughput sequencing reveals inbreeding depression in a natural population. Proceedings of the National Academy of Sciences of the United States of America, 111: 3775-3780. Doi: 10.1073/pnas.1318945111

Dasmahapatra KK, Lacy RC, Amos W (2007) Estimating levels of inbreeding using AFLP markers. Heredity 100:286-295.

---

`mouse_snps` *Oldfield mouse SNP data*

---

## Description

Mouse snp data in 0 (homozygous), 1(heterzygous) and `NA` (missing) format. Each row represents an individual and each column is a locus.

## Format

A `data.frame` with 36 observations at 13198 loci.

## References

Hoffman, J.I., Simpson, F., David, P., Rijks, J.M., Kuiken, T., Thorne, M.A.S., Lacey, R.C. & Dasmahapatra, K.K. (2014) High-throughput sequencing reveals inbreeding depression in a natural population. Proceedings of the National Academy of Sciences of the United States of America, 111: 3775-3780. Doi: 10.1073/pnas.1318945111

Dasmahapatra KK, Lacy RC, Amos W (2007) Estimating levels of inbreeding using AFLP markers. Heredity 100:286-295.

## plot.inbreed *Plot an inbreed object*

### Description

Plot an inbreed object

### Usage

```
## S3 method for class 'inbreed'
plot(x, true_g2 = FALSE, plottype = c("boxplot",
  "histogram"), ...)
```

### Arguments

| | |
|---|---|
| x | An inbreed object. |
| true_g2 | For plotting a simulate_g2 output. If TRUE, plots the real g2 (based on realized f) as a reference line. |
| plottype | deprecated. "boxplot" or "histogram" to plot the output of r2_hf() and to show either the boxplots through resampling of loci or the histogram from the bootstrapping of r2 over individuals. |
| ... | Additional arguments to the hist() function for the g2 and HHC functions. Additional arguments to the boxplot() function for plotting the result of the r2_hf() function. |

### Author(s)

Martin Stoffel (martin.adam.stoffel@gmail.com)

### See Also

g2_snps, g2_microsats

## print.inbreed *Print an inbreed object*

### Description

Displays the results a inbreed object.

### Usage

```
## S3 method for class 'inbreed'
print(x, ...)
```

### Arguments

| | |
|---|---|
| x | An inbreed object from one of the inbreedR functions. |
| ... | Additional arguments; none are used in this method. |

### Author(s)

Martin Stoffel (martin.adam.stoffel@gmail.com)

### See Also

g2_snps, g2_microsats, plot

---

## r2_hf — *Expected r2 between standardized multilocus heterozygosity (h) and inbreeding level (f)*

---

### Description

Expected r2 between standardized multilocus heterozygosity (h) and inbreeding level (f)

### Usage

```
r2_hf(genotypes, type = c("msats", "snps"), nboot = NULL,
  parallel = FALSE, ncores = NULL, CI = 0.95)
```

### Arguments

| | |
|---|---|
| genotypes | data.frame with individuals in rows and loci in columns, containing genotypes coded as 0 (homozygote), 1 (heterozygote) and NA (missing) |
| type | specifies g2 formula to take. Type "snps" for large datasets and "msats" for smaller datasets. |
| nboot | number of bootstraps over individuals to estimate a confidence interval around r2(h, f) |
| parallel | Default is FALSE. If TRUE, bootstrapping and permutation tests are parallelized |
| ncores | Specify number of cores to use for parallelization. By default, all available cores but one are used. |
| CI | confidence interval (default to 0.95) |

### Value

| | |
|---|---|
| call | function call. |
| r2_hf_full | expected r2 between inbreeding and sMLH for the full dataset |
| r2_hf_boot | expected r2 values from bootstrapping over individuals |
| CI_boot | confidence interval around the expected r2 |
| nobs | number of observations |
| nloc | number of markers |

## Author(s)

Martin A. Stoffel (martin.adam.stoffel@gmail.com)

## References

Slate, J., David, P., Dodds, K. G., Veenvliet, B. A., Glass, B. C., Broad, T. E., & McEwan, J. C. (2004). Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Heredity, 93(3), 255-265.

Szulkin, M., Bierne, N., & David, P. (2010). HETEROZYGOSITY-FITNESS CORRELATIONS: A TIME FOR REAPPRAISAL. Evolution, 64(5), 1202-1217.

## Examples

```
data(mouse_msats)
genotypes <- convert_raw(mouse_msats)
(out <- r2_hf(genotypes, nboot = 100, type = "msats", parallel = FALSE))
plot(out)
```

---

# r2_Wf *Expected r2 between inbreeding level (f) and fitness (W)*

---

## Description

Expected r2 between inbreeding level (f) and fitness (W)

## Usage

```
r2_Wf(genotypes, trait, family = "gaussian", type = c("msats", "snps"),
  nboot = NULL, parallel = FALSE, ncores = NULL, CI = 0.95)
```

## Arguments

| | |
|---|---|
| genotypes | A data.frame with individuals in rows and loci in columns, containing genotypes coded as 0 (homozygote), 1 (heterozygote) and NA (missing). |
| trait | vector of any type which can be specified in R's glm() function. Sequence of individuals has to match sequence of individuals in the rows of the genotypes data.frame. |
| family | distribution of the trait. Default is gaussian. For other distributions, just naming the distribution (e.g. binomial) will use the default link function (see ?family). Specifying another link function can be done in the same way as in the glm() function. A binomial distribution with probit instead of logit link would be specified with family = binomial(link = "probit") |
| type | specifies g2 formula to take. Type "snps" for large datasets and "msats" for smaller datasets. |
| nboot | number of bootstraps over individuals to estimate a confidence interval around r2(W, f). |

| | |
|---|---|
| parallel | Default is FALSE. If TRUE, bootstrapping and permutation tests are parallelized. |
| ncores | Specify number of cores to use for parallelization. By default, all available cores but one are used. |
| CI | confidence interval (default to 0.95) |

## Value

| | |
|---|---|
| call | function call. |
| exp_r2_full | expected r2 between inbreeding and sMLH for the full dataset |
| r2_Wf_boot | expected r2 values from bootstrapping over individuals |
| CI_boot | confidence interval around the expected r2 |
| nobs | number of observations |
| nloc | number of markers |

## Author(s)

Martin A. Stoffel (martin.adam.stoffel@gmail.com)

## References

Slate, J., David, P., Dodds, K. G., Veenvliet, B. A., Glass, B. C., Broad, T. E., & McEwan, J. C. (2004). Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Heredity, 93(3), 255-265.

Szulkin, M., Bierne, N., & David, P. (2010). HETEROZYGOSITY-FITNESS CORRELATIONS: A TIME FOR REAPPRAISAL. Evolution, 64(5), 1202-1217.

## Examples

```
data(mouse_msats)
data(bodyweight)
genotypes <- convert_raw(mouse_msats)

(out <- r2_Wf(genotypes = genotypes, trait = bodyweight, family = "gaussian", type = "msats",
              nboot = 100, parallel = FALSE, ncores = NULL, CI = 0.95))
```

---

# simulate_g2 *Simulate $g2$*

---

## Description

This function can be used to simulate genotype data, draw subsets of loci and calculate the respective $g2$ values. Every subset of markers is drawn independently to give insights into the variation and precision of $g2$ calculated from a given number of markers and individuals.

## Usage

```
simulate_g2(n_ind = NULL, H_nonInb = 0.5, meanF = 0.2, varF = 0.03,
  subsets = NULL, reps = 100, type = c("msats", "snps"), CI = 0.95)
```

## Arguments

| | |
|---|---|
| n_ind | number of individuals to sample from the population |
| H_nonInb | true genome-wide heteorzygosity of a non-inbred individual |
| meanF | mean realized inbreeding f |
| varF | variance in realized inbreeding f |
| subsets | a vector specifying the sizes of marker-subsets to draw. Specifying subsets = c(2, 5, 10, 15, 20) would draw marker sets of 2 to 20 markers. The minimum number of markers to calculate g2 is 2. |
| reps | number of resampling repetitions |
| type | specifies g2 formula. Type "snps" for large datasets and "msats" for smaller datasets. |
| CI | Confidence intervals to calculate (default to 0.95) |

## Details

The simulate_g2 function simulates genotypes from which subsets of loci can be sampled independently. These simulations can be used to evaluate the effects of the number of individuals and loci on the precision and magnitude of $g2$. The user specifies the number of simulated individuals (n_ind), the subsets of loci (subsets) to be drawn, the heterozygosity of non-inbred individuals (H_nonInb) and the distribution of $f$ among the simulated individuals. The $f$ values of the simulated individuals are sampled randomly from a beta distribution with mean (meanF) and variance (varF) specified by the user (e.g. as in wang2011). This enables the simulation to mimic populations with known inbreeding characteristics, or to simulate hypothetical scenarios of interest. For computational simplicity, allele frequencies are assumed to be constant across all loci and the simulated loci are unlinked. Genotypes (i.e. the heterozygosity/homozygosity status at each locus) are assigned stochastically based on the $f$ values of the simulated individuals. Specifically, the probability of an individual being heterozygous at any given locus ($H$) is expressed as $H = H0(1 - f)$ , where $H0$ is the user-specified heterozygosity of a non-inbred individual and $f$ is an individual's inbreeding coefficient drawn from the beta distribution.

## Value

simulate_g2 returns an object of class "inbreed". The functions 'print' and 'plot' are used to print a summary and to plot the g2 values with means and confidence intervals

An 'inbreed' object from simulate_g2 is a list containing the following components:

| | |
|---|---|
| call | function call. |
| estMat | matrix with all r2(h,f) estimates. Each row contains the values for a given subset of markers |
| true_g2 | "true" g2 value based on the assigned realized inbreeding values |
| n_ind | specified number of individuals |

| | |
|---|---|
| subsets | vector specifying the marker sets |
| reps | repetitions per subset |
| H_nonInb | true genome-wide heteorzygosity of a non-inbred individual |
| meanF | mean realized inbreeding f |
| varF | variance in realized inbreeding f |
| min_val | minimum g2 value |
| max_val | maximum g2 value |
| all_CI | confidence intervals for all subsets |
| all_sd | standard deviations for all subsets |

## Author(s)

Marty Kardos (marty.kardos@ebc.uu.se) & Martin A. Stoffel (martin.adam.stoffel@gmail.com)

## Examples

```
data(mouse_msats)
genotypes <- convert_raw(mouse_msats)
sim_g2 <- simulate_g2(n_ind = 10, H_nonInb = 0.5, meanF = 0.2, varF = 0.03,
                      subsets = c(4,6,8,10), reps = 100,
                      type = "msats")
plot(sim_g2)
```

---

# simulate_r2_hf *Calculates the expected squared correlation between heteorzygosity and inbreeding for simulated marker sets*

---

## Description

This function can be used to simulate genotype data, draw random subsamples and calculate the expected squared correlations between heterozygosity and fitness ($r2(h, f)$). Every subset of markers is drawn independently to give insights into the variation and precision of $r2(h, f)$ calculated from a given number of markers and individuals.

## Usage

```
simulate_r2_hf(n_ind = NULL, H_nonInb = 0.5, meanF = 0.2, varF = 0.03,
  subsets = NULL, reps = 100, type = c("msats", "snps"), CI = 0.95)
```

## Arguments

| | |
|---|---|
| n_ind | number of individuals to sample from the population |
| H_nonInb | true genome-wide heteorzygosity of a non-inbred individual |
| meanF | mean realized inbreeding f |
| varF | variance in realized inbreeding f |
| subsets | a vector specifying the sizes of marker-subsets to draw. Specifying subsets = c(2, 5, 10, 15, 20) would draw marker sets of 2 to 20 markers. The minimum number of markers is 2. |
| reps | number of resampling repetitions |
| type | specifies g2 formula. Type "snps" for large datasets and "msats" for smaller datasets. |
| CI | Confidence intervals to calculate (default to 0.95) |

## Details

The simulate_r2_hf function simulates genotypes from which subsets of loci can be sampled independently. These simulations can be used to evaluate the effects of the number of individuals and loci on the precision and magnitude of the expected squared correlation between heterozygosity and inbreeding ($r2(h, f)$). The user specifies the number of simulated individuals (n_ind), the subsets of loci (subsets) to be drawn, the heterozygosity of non-inbred individuals (H_nonInb) and the distribution of $f$ among the simulated individuals. The $f$ values of the simulated individuals are sampled randomly from a beta distribution with mean (meanF) and variance (varF) specified by the user (e.g. as in wang2011). This enables the simulation to mimic populations with known inbreeding characteristics, or to simulate hypothetical scenarios of interest. For computational simplicity, allele frequencies are assumed to be constant across all loci and the simulated loci are unlinked. Genotypes (i.e. the heterozygosity/homozygosity status at each locus) are assigned stochastically based on the $f$ values of the simulated individuals. Specifically, the probability of an individual being heterozygous at any given locus ($H$) is expressed as $H = H0(1 - f)$ , where $H0$ is the user-specified heterozygosity of a non-inbred individual and $f$ is an individual's inbreeding coefficient drawn from the beta distribution.

## Value

simulate_r2_hf returns an object of class "inbreed". The functions 'print' and 'plot' are used to print a summary and to plot the r2(h, f) values with means and confidence intervals

An 'inbreed' object from simulate_g2 is a list containing the following components:

| | |
|---|---|
| call | function call. |
| estMat | matrix with all r2(h,f) estimates. Each row contains the values for a given subset of markers |
| n_ind | specified number of individuals |
| subsets | vector specifying the marker sets |
| reps | repetitions per subset |
| H_nonInb | true genome-wide heteorzygosity of a non-inbred individual |
| meanF | mean realized inbreeding f |

| | |
|---|---|
| varF | variance in realized inbreeding f |
| min_val | minimum g2 value |
| max_val | maximum g2 value |
| all_CI | confidence intervals for all subsets |
| all_sd | standard deviations for all subsets |

## Author(s)

Marty Kardos (marty.kardos@ebc.uu.se) & Martin A. Stoffel (martin.adam.stoffel@gmail.com)

## Examples

```
data(mouse_msats)
genotypes <- convert_raw(mouse_msats)
sim_r2 <- simulate_r2_hf(n_ind = 10, H_nonInb = 0.5, meanF = 0.2, varF = 0.03,
                      subsets = c(4,6,8,10), reps = 100,
                      type = "msats")
plot(sim_r2)
```

---

# sMLH *Calculate multilocus heterozygosity (MLH)*

---

## Description

sMLH is defined as the total number of heterozygous loci in an individual divided by the sum of average observed heterozygosities in the population over the subset of loci successfully typed in the focal individual.

## Usage

```
sMLH(genotypes)
```

## Arguments

| | |
|---|---|
| genotypes | data.frame with individuals in rows and loci in columns, containing genotypes coded as 0 (homozygote), 1 (heterozygote) and NA (missing) |

## Value

Vector of individual standardized multilocus heterozygosities

## Author(s)

Martin A. Stoffel (martin.adam.stoffel@gmail.com)

## References

Coltman, D. W., Pilkington, J. G., Smith, J. A., & Pemberton, J. M. (1999). Parasite-mediated selection against inbred Soay sheep in a free-living, island population. Evolution, 1259-1267.

## Examples

```
data(mouse_msats)
genotypes <- convert_raw(mouse_msats)
het <- sMLH(genotypes)
```

# Index