Package 'scrm'

January 8, 2024

Type Package

Title Simulating the Evolution of Biological Sequences

Version 1.7.5

Author Paul Staab [aut, cph],
Zhu Sha [aut, cph],
Dirk Metzler [aut, cre, cph, ths],
Gerton Lunter [aut, cph, ths]

Maintainer Dirk Metzler <metzler@bio.lmu.de>

Description A coalescent simulator that allows the rapid simulation of biological sequences under neutral models of evolution, see Staab et al. (2015) <doi:10.1093/bioinformatics/btu861>. Different to other coalescent based simulations, it has an optional approximation parameter that allows for high accuracy while maintaining a linear run time cost for long sequences. It is optimized for simulating massive data sets as produced by Next-Generation Sequencing technologies for up to several thousand sequences.

URL https://github.com/scrm/scrm-r

BugReports https://github.com/scrm/scrm-r/issues

License GPL (>= 3)

Depends R (>= 3.1.0)

Imports Rcpp (>= 0.11.2)

Suggests ape, knitr, rmarkdown, testthat (>= 0.9.0)

VignetteBuilder knitr

LinkingTo Rcpp

RoxygenNote 6.1.0

NeedsCompilation yes

Repository CRAN

Date/Publication 2024-01-08 13:30:02 UTC

R topics documented:

`scrm-package` *Simulating the evolution of biological sequences*

Description

The Sequential Coalescent with Recombination Model (SCRM) is an approximation of the Ancestral Recombination Graph. It can be used to simulate the neutral evolution of chromosomes/biological sequences subject to possibly complicated population structure. The program *scrm* is an implementation of this model that is designed to act as an drop-in replacement for the widely adopted coalescent simulator *ms*. This package contains *scrm* along with an R interface.

Author(s)

Paul Staab, Zhu Sha, Dirk Metzler & Gerton Lunter

Maintainer: Paul Staab <develop@paulstaab.de>

See Also

- `scrm` for details on how to use *scrm*,
- `vignette('scrm-Arguments')` for an overview of command line arguments and
- `vignette('scrm-TreesForApe')` for an example on using genealogies simulated with *scrm* with package 'ape'.

`scrm` *Simulate the evolution of biological sequences*

Description

This function provides an interface for calling *scrm* from R. The command line options are passed via the `args` argument. The vignette 'scrm-Arguments' contains details about the available options. Summary statistics are converted into an R format. Additionally, there is an option to write the original command line output into a file.

Usage

```
scrm(args, file = "")
```

Arguments

`args`	A string containing the command-line arguments for scrm. Look at scrms vignette for a description of available arguments.
`file`	If provided, scrm will additionally write it is output into a file with the given file, using an ms-like text output.

Value

A named list of summary statistics. Most summary statistics are again a list, where each entry contains the value for one locus. For the site frequency spectrum, the summary statistic is a matrix, where each row contains the spectrum for one locus.

Seeding

The R version of scrm uses random number from R's random generator. Therefore, the '-seed' argument of the command-line version will be ignored, and no seed is given in the output. Use the R function `set.seed` prior to calling this function to ensure reproducibility of results.

See Also

- `vignette('scrm-Arguments')` for an overview of command line arguments and
- `vignette('scrm-TreesForApe')` for an example on using genealogies simulated with *scrm* with package 'ape'.

Examples

```
set.seed(789)
# 5 Chromosomes with 100 bases each with recombination and mutation
sum_stats <- scrm('5 1 -r 3.1 100 -t 1.5 -T -L')
str(sum_stats)

# Simulate the site frequency spectrum at 3 loci. For each locus
# 10 Chromosomes of 1Mb length are sampled from two populations with
# migration inbetween.
scrm('10 3 -r 400 1000000 -l 100000 -I 2 4 6 0.5 -t 300 -oSFS')$sfs
```

Index