

Package: hmmibdr (via r-universe)

June 9, 2024

Type Package

Title HMM Identity by Descent

Version 0.2.0

Author OJ Watson

Maintainer Oliver Watson <oj.watson@hotmail.co.uk>

Description Wrapper for HMMIBD in Rcpp.

License MIT + file LICENCE

LazyData true

Depends R (>= 3.2.1)

Imports Rcpp (>= 1.0.0)

Suggests testthat

BugReports <https://github.com/OJWatson/hmmibdr/issues>

LinkingTo Rcpp

Encoding UTF-8

RoxygenNote 7.1.1

SystemRequirements C++11

Repository <https://plasmogenepi.r-universe.dev>

RemoteUrl <https://github.com/OJWatson/hmmibdr>

RemoteRef HEAD

RemoteSha 933879606470f921ebc629710b749b71baa4154f

Contents

hmmibdr	2
hmmibd_c	2
hmm_ibd	3

Index	5
--------------	----------

hmmibdr	<i>Wrapper for HMMIBD in Rcpp</i>
---------	-----------------------------------

Description

Wrapper for HMMIBD in Rcpp

Details

Rcpp implementation of THMMIBD

References

<https://github.com/glipsnort/hmmIBD>

hmmibd_c	<i>hmmIBD</i>
----------	---------------

Description

hmmIBD implementation from <https://github.com/glipsnort/hmmIBD>

Usage

```
hmmibd_c(param_list)
```

Arguments

`param_list` A list of parameters created with `hmm_ibd`

Details

`hmmibd_c` implements hidden Markov model for detecting segments of shared ancestry (identity by descent) in genetic sequence data.

Description

This function triggers the c code for the categorical method

Usage

```
hmm_ibd(  
  input_file,  
  output_file,  
  allele_freqs = NULL,  
  genotypes_sec_pop = NULL,  
  allele_freqs_sec_pop = NULL,  
  max_fit_iterations = NULL,  
  exclude_ids = NULL,  
  analysis_ids = NULL,  
  num_gens = NULL,  
  overwrite = FALSE,  
  fract_only = FALSE,  
  eps = 0.001,  
  min_inform = 10,  
  min_discord = 0,  
  max_discord = 1,  
  nchrom = 14,  
  min_snp_sep = 5,  
  rec_rate = 7.4e-07,  
  cache = TRUE  
)
```

Arguments

<code>input_file</code>	File of genotype data. See below for format.
<code>output_file</code>	Output file name. Two output files will be produced, with ".hmm.txt" and ".hmm_fract.txt" appended to the supplied name.
<code>allele_freqs</code>	File of allele frequencies for the sample population. Format: tab-delimited, no header, one variant per row. Line format: <chromosome (int)> <position (bp, int)> <allele 1 freq> <all 2 freq> ... The genotype and frequency files must contain exactly the same variants, in the same order. If no file is supplied, allele frequencies are calculated from the input data file.
<code>genotypes_sec_pop</code>	File of genotype data from a second population; same format as for -i. (added in 2.0.0)
<code>allele_freqs_sec_pop</code>	File of allele frequencies for the second population; same format as for -f. (added in 2.0.0)

max_fit_iterations	Maximum number of fit iterations (defaults to 5).
exclude_ids	File of sample ids to exclude from all analysis. Format: no header, one id (string) per row. (Note: b stands for "bad samples".)
analysis_ids	File of sample pairs to analyze; all others are not processed by the HMM (but are still used to calculate allele frequencies). Format: no header, tab-delimited, two sample ids (strings) per row. (Note: "g" stands for "good pairs".)
num_gens	Cap on the number of generations (floating point). Sets the maximum value for that parameter in the fit. This is useful if you are interested in recent IBD and are working with a population with substantial linkage disequilibrium. Specifying a small value will force the program to assume little recombination and thus a low transition rate; otherwise it will identify the small blocks of LD as ancient IBD, and will force the number of generations to be large.
overwrite	Boolean detailing if output files already exist, should they be overwritten. Default = FALSE
fract_only	Boolean detailing whether to return just the fract. Default = FALSE
eps	Numeric for error rate in genotype calls. Default = .001
min_inform	Minimum number of informative sites in a pairwise. Default = 10
min_discord	Minimum discordance in comparison. Default = 0. Set > 0 to skip identical pairs
max_discord	Maximum discordance in comparison. Default = 1. Set < 1 to skip unrelated pairs
nchrom	Number of chromosomes. Default = 14 for falciparum
min_snp_sep	Minimum snp distance, i.e. skip next snp(s) if too close to last one. Default = 5 (bp)
rec_rate	Recombination rate. Default = $7.4e-7$. ($7.4e-5$ cM/bp or 13.5 kb/cM Miles et al, Genome Res 26:1288-1299 (2016))
cache	Should files created by hmm_ibd be cached (i.e. not deleted). Default = TRUE, i.e. keep the files

Value

return list of summary data frames of hmmIBD output

Index

hmm_ibd, 3
hmmibd_c, 2
hmmibdr, 2