

Package ‘Haplin’

May 27, 2009

Type Package

Title Analyzing case-parent triad and/or case-control data with SNP haplotypes

Version 3.0.1

Date 2009-05-25

Author Håkon K. Gjessing

Maintainer Håkon K. Gjessing <hakon.gjessing@fhi.no>

Depends MASS, mgcv

Description Haplin performs a genetic association analysis of case-parent triad (trio) data with multiple markers. It can also incorporate complete or incomplete control triads, for instance independent control children. Estimation is based on haplotypes, for instance SNP haplotypes, even though phase is not known from the genetic data. Haplin estimates relative risk (RR + conf.int.) and p-value associated with each haplotype. It uses maximum likelihood estimation to make optimal use of data from triads with missing genotypic data, for instance if some SNPs has not been typed for some individuals. Haplin also allows estimation of effects of maternal haplotypes, particularly appropriate in perinatal epidemiology.

License GPL (>= 2)

URL <http://www.uib.no/smis/gjessing/genetics/software/haplin/>

Encoding latin1

Repository CRAN

Date/Publication 2009-05-27 08:58:57

R topics documented:

haplin	2
pedToHaplin	6
plot.haplin	8
print.haplin	9
print.summary.haplin	10
summary.haplin	12

haplin	<i>Fitting log-linear models to case-parent triad and/or case-control data</i>
--------	--------------------------------------------------------------------------------

Description

Produces an object of class `haplin`, which is the result of fitting the log-linear models to the data

Usage

```
haplin(filename,
markers = "ALL", n.vars = 0, sep = " ", allele.sep = ";",
na.strings = "NA", design = "triad", use.missing = FALSE,
xchrom = FALSE, maternal = FALSE, scoretest = "no",
ccvar = NULL, covar = NULL, sex = NULL, reference = "reciprocal",
response = "free", threshold = 0.01, max.haplos = NULL,
haplo.file = NULL, resampling = FALSE, max.EM.iter = 50,
data.out = FALSE, verbose = TRUE, printout = TRUE)
```

Arguments

Of the following arguments, only `filename` is required. Use of the remaining arguments will depend on the type of analysis.

<code>filename</code>	A character string giving the name and path of the ASCII data file to be read.
<code>markers</code>	Default is "ALL", which means HAPLIN uses all available markers in the data set in the analysis. For the current version of HAPLIN the number of markers used at a single run should probably not exceed 4 or 5 due to the computational burden. The <code>markers</code> argument can be used to select appropriate markers from the file without creating a new file for the selected markers. For instance, if <code>markers</code> is set to <code>c(2,4)</code> , HAPLIN will only use the second and fourth markers supplied in the data set. When running HAPLIN, it may be a good idea to start exploring a few markers at a time, using this argument.
<code>n.vars</code>	Numeric. The number of variables (columns) in the data file before (to the left) of the genetic data.
<code>sep</code>	The character separator used in the data file to separate between "columns", where each column contains the two alleles of a single individual at a single marker.
<code>allele.sep</code>	The character separator used in the data file to separate the two alleles for a single individual in a single marker. The recommended (default) separator is ";", but for SNPs an empty "" is also common.
<code>na.strings</code>	The character string indicating missing data in the data file. Default is to use "NA" in place of, for instance, C;T for a SNP that hasn't been typed in that individual.

design	The value "triad" is used for the standard case triad design, without independent controls. The value "cc.triad" means a combination of case triads and control triads. This requires the argument <code>ccvar</code> to point to the data column containing the case-control variable. The value "cc" means a simple case-control design, where the parents have not been genotyped (there are no data columns for parental genes)
use.missing	A logical value used to determine whether triads with missing data should be included in the analysis. When set to TRUE, Haplin uses the EM algorithm to obtain risk estimates, also taking into account triads with missing data. The standard errors and p-values are adjusted to correct for this. The default, however, is FALSE. When FALSE, all triads having any sort of missing data are excluded before the analysis is run. Note that Haplin only looks at markers actually used in the analysis, so that if the markers argument (see below) is used to select a collection of markers for analysis, Haplin only excludes triads with missing data on the included markers.
xchrom	Not yet implemented
maternal	If TRUE, maternal effects are estimated as well as the standard fetal effects.
scoretest	Special interest only. If "no", no score test is computed. If "yes", an overall score p-value is included in the output, and the individual score values are returned in the haplin object. If "only", haplin is only run under the null hypothesis, and a simple score object is returned instead of the full haplin object. Useful if only score testing is needed.
ccvar	Numeric. Should give the column number for the column containing the case-control indicator in the data file. Needed for the "cc" and "cc.triad" designs. The column should contain two numeric values, of which the largest one is always used to denote cases.
covar	Not yet implemented
sex	Not yet implemented
reference	Decides how HAPLIN chooses its reference category for the effect estimates. Default value is "reciprocal". With the reciprocal reference the effect of a single or double dose of each haplotype is measured relative to the remaining haplotypes. This means that a new reference category is used for each single haplotype. Other possible values are "population" (which is similar to reciprocal, but where the reference category is always the total population), and "ref.cat", where a single haplotype is used as reference for all the rest. For ref.cat, the default is to choose the most frequent haplotype as the reference haplotype. The reference haplotype can be set explicitly by giving a numeric value for the reference argument. Note that the numeric value refers to the haplotype's position among the haplotypes selected for analysis by HAPLIN. This means that one should run HAPLIN once first to see what haplotypes are used before giving a numeric value to reference.
response	The default value "free" means that both single- and double dose effects are estimated. Choosing "mult" instead specifies a multiplicative dose-response model.
threshold	Sets the (approximate) lower limit for the haplotype frequencies of those haplotypes that should be retained in the analysis. Haplotypes that are less frequent are removed, and information about this is given in the output.

<code>max.haplos</code>	Not yet implemented
<code>haplo.file</code>	Not yet implemented
<code>resampling</code>	Default is FALSE. When FALSE, the individual haplotypes reconstructed by the EM algorithm as assumed known when computing CIs and p-values. If set to "jackknife" a jackknife-based resampling procedure is used when computing confidence intervals and p-values for effect estimates. This takes more time, but corrects the CIs and p-values for the uncertainty contained in unphased data. Note: in this version of Haplin, the resampling is no longer needed since the confidence intervals and p-values are already corrected in the standard computation.
<code>max.EM.iter</code>	The maximum number of iterations used by the EM algorithm. This value can be increased if necessary, which sometimes is the case with e.g. case-control data which a substantial amount of missing. However, for triad data with little missing information there is usually no need for many iterations.
<code>data.out</code>	Not yet implemented
<code>verbose</code>	Default is T (=TRUE). During the EM algorithm, HAPLIN prints the estimated parameters and deviance for each step. To avoid the output, set this argument to F (=FALSE).
<code>printout</code>	Logical. If TRUE (default), haplin prints a full summary of the results after finishing the estimation. If FALSE, no such printout is given, but the <code>summary</code> function can later be applied to a saved result to get the same summary.

Details

The output can be examined by `print`, `summary` and `plot`.

Value

An object of class `haplin` is returned

Warning

Typically, some of the included haplotypes will be relatively rare, such as a frequency of 1% - 5%. For those haplotypes there may be too little data to estimate the double doses properly, so the estimates may be unreliable. This is seen from the extremely wide confidence intervals. The rare double dose estimates should be disregarded, but the remaining single and double dose estimates are valid. To avoid the problem one can also reduce the model to a purely multiplicative model by setting `response = "mult"`.

Note

Further information is found on the web page

Author(s)

Håkon K. Gjessing
 Professor of Biostatistics
 Division of Epidemiology

Norwegian Institute of Public Health
(hakon.gjessing@fhi.no)

References

Gjessing HK and Lie RT. Case-parent triads: Estimating single- and double-dose effects of fetal and maternal disease gene haplotypes. *Annals of Human Genetics* (2006) 70, pp. 382-396.

Web Site: <http://www.uib.no/smis/gjessing/genetics/software/haplin/>

See Also

[plot.haplin](#), [pedToHaplin](#)

Examples

```
## Not run:

# Standard run:
haplin("data.dat")

# Specify path, estimate maternal effects:
haplin("C:/work/data.dat", maternal = T)

# Specify path, use haplotype no. 2 as reference:
haplin("C:/work/data.dat", reference = 2)

# Remove more haplotypes from estimation by increasing the threshold
# to 5%:
haplin("C:/work/data.dat", threshold = 0.05)

# Estimate maternal effects, using the most frequent haplotype as reference.
# Use all data, including triads with missing data. Select
# markers 3, 4 and 8 from the supplied data.
haplin("C:/work/data.dat", use.missing = T, maternal = T,
reference = "ref.cat", markers = c(3,4,8))
# Note: in this version of Haplin, the jackknife is
# no longer necessary since the standard errors are already corrected.

# Some examples showing how to save the Haplin result and later
# recall plot and summary results:

# Same analysis as above, saving the result in the object "result.1":
result.1 <- haplin("C:/work/data.dat", use.missing = T, maternal = T,
reference = "ref.cat", markers = c(3,4,8))

# Replot the saved result (fetal effects):
plot(result.1)

# Replot the saved result (maternal effects):
plot(result.1, plot.maternal = T)
```

```

# Print a very short summary of saved result:
result.1

# A full summary of saved result, with confidence intervals and
# p-values (the same as haplin prints when running):
summary(result.1)

# Some examples when the data file contains two covariates,
# the second is the case-control variable:

# The following standard triad run is INCORRECT since it disregards
# case status:
haplin("data.dat", use.missing = T, n.vars = 2, design = "triad")

# Combined run on "hybrid" design, correctly using both case-parent
# triads and control-parent triads:
haplin("data.dat", use.missing = T, n.vars = 2, ccvar = 2,
design = "cc.triad")

# If parent columns are not in the file, a plain case-control
# run can be used:
haplin("data.dat", use.missing = T, n.vars = 2, ccvar = 2,
design = "cc", response = "mult", reference = "ref.cat")

## End(Not run)

```

pedToHaplin

Convert from ped format data to Haplin format

Description

Converts an ASCII file from a standard ped format to the Haplin format

Usage

```
pedToHaplin(indata, outdata, merge = F, na.strings = "0", sep,
colnames.out = F)
```

Arguments

indata	A character string giving the name and path of the ASCII data file to be converted.
outdata	A character string giving the name and path for saving the converted file.
merge	If the alleles of each genotype are in two separate columns in the <code>indata</code> file, they must be merged (with the ";" separator) in the <code>outdata</code> file. This is done by setting <code>merge = TRUE</code> . Otherwise, it must be set to <code>FALSE</code> .

<code>na.strings</code>	The symbol used to denote missing data in <code>indata</code> . It is passed directly to R's <code>read.table</code>
<code>sep</code>	Column separator in <code>indata</code> . If unspecified, any white space will be used, as in <code>read.table</code> .
<code>colnames.out</code>	Provided just for the purpose of checking data. If <code>TRUE</code> , adds <code>colnames</code> to the <code>outdata</code> file to make it more readable. NOTE: Haplin does currently not use <code>colnames</code> , so this should be set to <code>FALSE</code> when producing the file to run on.

Details

Important: The first 6 columns should always be family id, individual id, father's id, mother's id, sex and casetype, in that order, then followed by the genetic data columns. If the genetic data columns are separated into two individual alleles, one should use the option `merge = TRUE` to merge them in the output file. If they are already in single columns, for instance as CT or C;T, `merge` should be set to `FALSE` (default).

Additional covariates can be included in the input file. If so, they should be placed after the 6 leading columns but before the genetic data. In this case, one should make sure the genetic data columns are already merged, and that `merge = FALSE`. (The `merge = TRUE` option when covariates are present will hopefully be implemented at some point...)

Value

The `outdata` file is written to disk. `pedToHaplin` returns (invisibly) the converted data file.

Warning

`pedToHaplin` has not been extensively tested, so you should always check the output before using it.

Note

Further information is found on the web page

Author(s)

Håkon K. Gjessing
Professor of Biostatistics
Division of Epidemiology
Norwegian Institute of Public Health
(hakon.gjessing@fhi.no)

References

Gjessing HK and Lie RT. Case-parent triads: Estimating single- and double-dose effects of fetal and maternal disease gene haplotypes. *Annals of Human Genetics* (2006) 70, pp. 382-396.

Web Site: <http://www.uib.no/smis/gjessing/genetics/software/haplin/>

See Also[haplin](#)**Examples**

```
## Not run:

# Standard run on supplied test file:
pedToHaplin("test_pedToHaplin.ped", outdata = "test_pedToHaplin_result.txt",
colnames.out = F, merge = T)

## End(Not run)
```

plot.haplin

Plot a haplin object

Description

Plot a haplin object and (optionally) produce picture files

Usage

```
## S3 method for class 'haplin':
plot(x, reference, separate.plots = F, filename,
filetype = "png", use.dd, ...)
```

Arguments

Of the following arguments, only `x` is required.

<code>x</code>	A haplin object, i.e. the result of running haplin.
<code>reference</code>	Same as <code>reference</code> argument in haplin. Note that when plotting, you can only choose "reciprocal", "population" or "ref.cat". You cannot use a numeric value to change the reference category, to do that haplin must be run over again. (See the <code>reference</code> argument of haplin.)
<code>separate.plots</code>	Logical. If you estimate effects of both fetal and maternal genes you can decide whether or not to plot them in the same plot. The default is the same plot (TRUE), the alternative (FALSE) means in separate plots. If you choose separate plots you may have to set the graphics window to "recording" to make sure you can scroll back to the first plot.
<code>filename</code>	If you want a file containing the plot to be produced, give a character string for the filename.
<code>filetype</code>	The default filetype is "png", alternatively you can choose "jpeg".

use.dd Numeric vector indicating which double dose estimates should be plotted. For instance, if set to c(1,3) only the first and third haplotypes will be drawn with double dose estimates. This is useful if some haplotypes are rare and you want to exclude the uncertain estimates from the plot.

... Further arguments to be passed on to the plot function

Note

Further information is found on the web page

Author(s)

Håkon K. Gjessing
Professor of Biostatistics
Division of Epidemiology
Norwegian Institute of Public Health
(hakon.gjessing@fhi.no)

References

Gjessing HK and Lie RT. Case-parent triads: Estimating single- and double-dose effects of fetal and maternal disease gene haplotypes. *Annals of Human Genetics* (2006) 70, pp. 382-396.

Web Site: <http://www.uib.no/smis/gjessing/genetics/software/haplin/>

See Also

[haplin](#)

Examples

```
## Not run:  
  
# Produce separate plots for child and mother, dump plots to files:  
res <- haplin("data.dat", use.missing = T, maternal = T)  
plot(res, separate.plots = T, filename = "Haplinres.png")  
  
## End(Not run)
```

print.haplin

Print a haplin object

Description

Print basic information about a haplin object

Usage

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

Arguments

x A haplin object, i.e. the result of running haplin.
... Other arguments, passed on to print.

Note

Further information is found on the web page

Author(s)

Håkon K. Gjessing
Professor of Biostatistics
Division of Epidemiology
Norwegian Institute of Public Health
<hakon.gjessing@fhi.no>

References

Gjessing HK and Lie RT. Case-parent triads: Estimating single- and double-dose effects of fetal and maternal disease gene haplotypes. *Annals of Human Genetics* (2006) 70, pp. 382-396.

Web Site: <http://www.uib.no/smis/gjessing/genetics/software/haplin/>

See Also

[haplin](#)

```
print.summary.haplin
```

Print the summary of a haplin object

Description

Print the result of applying `summary` to a haplin object

Usage

```
## S3 method for class 'summary.haplin':  
print(x, digits, ...)
```

Arguments

`x` A `haplin` object, i.e. the result of running `haplin`.
`digits` The number of digits to be used in the printout. Defaults to 3.
`...` Other arguments (ignored).

Note

Further information is found on the web page

Author(s)

Håkon K. Gjessing
Professor of Biostatistics
Division of Epidemiology
Norwegian Institute of Public Health
<hakon.gjessing@fhi.no>

References

Gjessing HK and Lie RT. Case-parent triads: Estimating single- and double-dose effects of fetal and maternal disease gene haplotypes. *Annals of Human Genetics* (2006) 70, pp. 382-396.

Web Site: <http://www.uib.no/smis/gjessing/genetics/software/haplin/>

See Also

[haplin](#)

Examples

```
## Not run:  
  
# Standard summary:  
res <- haplin("data.dat", use.missing = T, maternal = T)  
summary(res)  
  
# Increase number of digits in printout  
print(summary(res), digits = 8)  
  
## End(Not run)
```

summary.haplin *Summary of a haplin object*

Description

Provides detailed information about estimation results from a haplin object.

Usage

```
## S3 method for class 'haplin':  
summary(object, reference, ...)
```

Arguments

object	A haplin object, i.e. the result of running haplin.
reference	Same as reference argument in haplin. Note that when producing the summary, you can only choose "reciprocal", "population" or "ref.cat". You cannot use a numeric value to change the reference category, to do that haplin must be run over again. (See the reference argument of haplin.)
...	Other arguments (ignored).

Note

Further information is found on the web page

Author(s)

Håkon K. Gjessing
Professor of Biostatistics
Division of Epidemiology
Norwegian Institute of Public Health
{hakon.gjessing@fhi.no}

References

Gjessing HK and Lie RT. Case-parent triads: Estimating single- and double-dose effects of fetal and maternal disease gene haplotypes. *Annals of Human Genetics* (2006) 70, pp. 382-396.

Web Site: <http://www.uib.no/smis/gjessing/genetics/software/haplin/>

See Also

[haplin](#)

Examples

```
## Not run:  
  
# Produce separate plots for child and mother, dump plots to files:  
res <- haplin("data.dat", use.missing = T, maternal = T)  
summary(res)  
  
## End(Not run)
```

Index

`haplin`, [2](#), [8–12](#)

`pedToHaplin`, [5](#), [6](#)

`plot.haplin`, [5](#), [8](#)

`print.haplin`, [9](#)

`print.summary.haplin`, [10](#)

`summary.haplin`, [12](#)