Standard `haplin` run

HAPLIN RUN

Load data:

```
pres.data <- genDataLoad(filename = "data_preprocessed",
 dir.in = "data")
```

Standard run, first SNP in file:

```
 haplin(data = pres.data, markers = 1)
```

Haplin output: Missing genotypes

```
## HAPLIN, VERSION 7.2.3 ##
opening ff C:/Users/hkgjess/AppData/Local/Temp/Rtmpk9XS4l/Users/hkg
The following 84 data lines were dropped due to missing data:
 2 6 7 15 42 48 49 50 52 56 58 62 63 70 81 84 90 94 110 111 132 143
```

- There's a lot of missing here...
- This may be due to mother-child dyads, without genotyped father.
- Use the option **use.missing = T** to impute.
- Imputation is done without artificial improvement of precision.

HAPLIN OUTPUT

```
None of the retained lines contained Mendelian inconsistencies
```

OK, good.

```
Running EM for preliminary estimates of haplotype frequencies
  ...  Done
Removing unused haplotypes...  Done
```

Haplin retains only alleles/haplotypes with frequencies above the threshold, default is 0.01.

```
NOTE: ONLY SINGLE REFERENCE CATEGORY METHOD ALLOWED FOR
  TWO HAPLOTYPES/ALLELES!
 (reference has been set to 2 )
```

No problem, this is standard when only two alleles

HAPLIN OUTPUT: CONVERGENCE

```
Using EM to estimate model with no effect:
EM iter: 1    |GLM deviance: 0            |Coefficients: 2.2352e-18
EM iter: 2    |GLM deviance: 5.63961      |Coefficients: -1.67805
EM iter: 3    |GLM deviance: 5.63961      |Coefficients: -1.67805

Using EM to estimate full model:
EM iter: 1    |GLM deviance: 0            |Coefficients: 2.46795e-1
EM iter: 2    |GLM deviance: 5.38964      |Coefficients: -1.73422
EM iter: 3    |GLM deviance: 5.38964      |Coefficients: -1.73422

Estimation finished, preparing output...  Done
```

- Perhaps more than you needed to know.... `verbose = F` turns it off
- Shows convergence of parameters.
- **But** lots of EM iterations may signify problems, such as haplotypes bridging a location with low LD.

HAPLIN OUTPUT: ARGUMENT SETTINGS

```
----Arguments supplied to haplin in this run:----

filespecs: markers = 1

model: design = "triad", use.missing = FALSE, xchrom = FALSE,
  comb.sex = "double", maternal = FALSE, poo = FALSE,
  test.maternal = FALSE, scoretest = "no"

variables: ccvar = NULL, strata = NULL, sex = NULL

haplos: reference = "reciprocal", response = "free",
 threshold = 0.01, max.haplos = NULL, haplo.file = NULL

control: resampling = "no", max.EM.iter = 50,
 data.out = "no", verbose = TRUE, printout = TRUE
```

This is mostly for reference, to show what input parameters were fed to `haplin`.

HAPLIN OUTPUT: TRIO ACCOUNTING

```
----Data summary:----

Number of triads in original file: 559

Accounting for possible loss of triads:

 Cause of loss     Triads removed Triads remaining
 Missing data                 84              475
 Mendelian incons.             0              475
 Unused haplotypes             0              475

Triads remaining for analysis: 475
```

OK. As mentioned above, `use.missing = T` would probably be better.

HAPLIN OUTPUT: TRIO ACCOUNTING

Cause of loss	Triads removed	Triads remaining
Missing data	84	475
Mendelian incons.	0	475
Unused haplotypes	0	475

- Removing unused haplotypes may sometimes cause trios to disappear, since some families may only have genotypes compatible with the rare haplotypes.
- Here: Only two “haplotypes” (only one SNP)
  Minor Allele Frequency large enough that both are retained.

Haplin output: Marker summary info

```
NOTE: In the following, the most frequent allele
 is printed as upper-case, all others are lower-case

Marker rs1:
Missing alleles:  0
 Allele Frequency Percent
 c       116        4.1
 G      2734       95.9
 total  2850      100.0
Chi-squared test for HWE, p-value:  0.8067
```

- Important to check HWE test
- The allele frequency is just a raw count, not estimated

Haplin output: Allele/haplotype frequencies

```
Haplotypes removed because of low frequencies:
None

Haplotypes used in the analysis, with coding:
c G
1 2

----Estimation results:----

Number of haplotypes: 2

Haplotype frequencies with 95% confidence intervals:
 Haplotype Frequency(%) lower upper
 c          3.77         2.73  5.21
 G         96.23        94.79 97.27
```

NOTE: These are frequencies estimated from full model

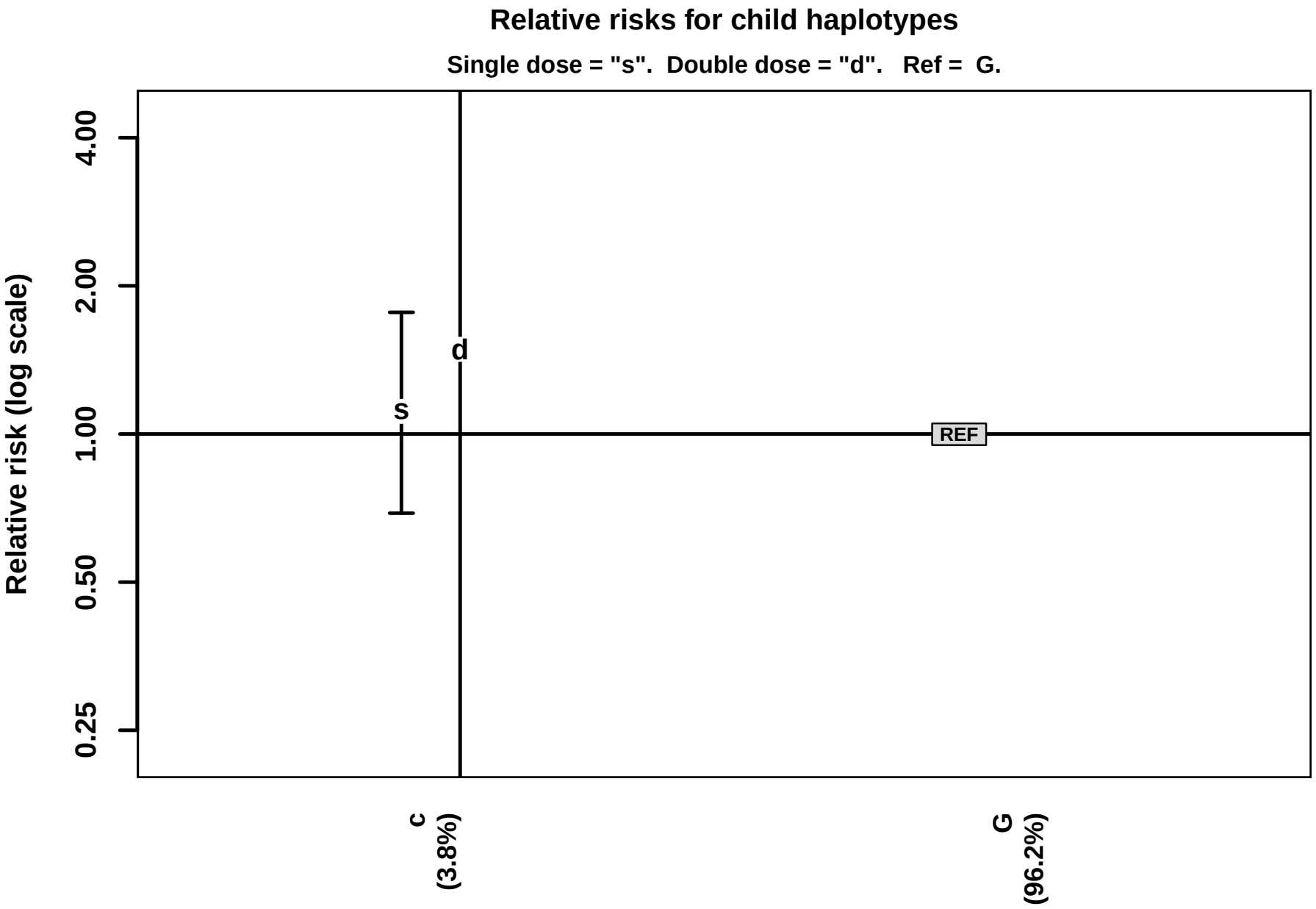
Haplin output: Effect estimates

```
Single- and double dose effects (Relative Risk) with 95% confidence
Reference method: ref.cat
Reference category: 2 (Haplotype G)
Response model: free

----Child haplotypes----
 Haplotype Dose     Relative Risk Lower CI   Upper CI   P-value
 c         Single   1.11          0.69       1.77       0.668
 c         Double   1.49          0.189      12.1       0.709

 G         Single   REF
 G         Double   REF
```

- A single dose of “c” is harmless.
- A double dose of “c” increases risk by 50%,

  ... but CIs are WIDE because “c” is rare, and thus “cc” is even rarer.

Haplin output: Effect estimates



Haplin output: Likelihood ratio test

```
Overall test for difference between null model (no effects)
and full model:
------------
LIKELIHOOD RATIO TEST:
 Loglike null model:        -319.0939
 Loglike full model:        -318.9689
 df:                           2.0000
 Likelihood ratio p-value:     0.8825

(NOTE: The test may be sensitive to rare haplotypes)
```

Likelihood ratio p-value tests an *overall* difference between the

- full model (all relative risks included)

and the

- null model (only haplotype frequencies.)